Received: 9 February 2023 | Revised: 31 July 2023 | Accepted: 7 August 2023 | Published online: 11 September 2023

RESEARCH ARTICLE

Artificial Intelligence and Applications yyyy, Vol. XX(XX) 1–5

Has OpenAI Achieved Artificial General Intelligence in ChatGPT?

DOI: 10.47852/bonviewAIA3202751



BON VIEW PUBLISHING

Andy E. Williams^{1,*} 1 Nobeah Foundation, Kenya

Abstract: In this paper, we present an analysis of ChatGPT, a language model developed by OpenAI, through the lens of Human-Centric Functional Modeling (HCFM). ChatGPT is designed to interact through a chat interface in a conversational manner, with the ability to answer follow-up questions, admit mistakes, challenge incorrect premises, and reject inappropriate requests. Since HCFM is hypothesized to provide a functional model for assessing the existence and magnitude of general problem-solving ability (intelligence), and since according to ChatGPT itself HCFM is the only such functional model in existence, the purpose of the paper is to demonstrate the usefulness of HCFM in determining whether an AI like ChatGPT is an AGI. Using Human-Centric Functional Modeling, we aim to determine whether ChatGPT exhibits narrow problem-solving ability, classifying it as an artificial intelligence (AI), or whether it exhibits general problem-solving ability, classifying it as AGI. We also consider the magnitude of ChatGPT's problem-solving ability within the conceptual space defined by HCFM. Finally, this paper also explores the issue from the perspective of the "collective social brain" hypothesis, which predicts which AI behavior the majority of humans will find to be intelligent, as well as predicting that true machine intelligence lies outside such narrow human definitions of intelligent behavior.

Keywords: Human-Centric Functional Modeling, Artificial General Intelligence, General Collective Intelligence, individual optimization, collective optimization

*Corresponding author: Andy E. Williams, Nobeah Foundation, Kenya. Email: awilliams@nobeahfoundation.org

1. Background

Large Language Models (LLMs) have garnered significant attention for their ability to perform various natural language processing (NLP) tasks and their ability to generate human-like language. These neural network-based models also offer insights into the nature of language and cognition, as understanding their behavior requires a theory of mind [1]. However, in understanding the current capabilities of LLMs, as well as any directions to which future advancements in LLMs might be constrained by human factors, it is important to also have an accurate theory of mind for the decision-makers driving its development, so that assessing the ways in which the development and deployment of LLMs are entangled with factors such as the business interests of tech companies and the potential power imbalances they may create [2], might be more reliably achievable. In this paper, we explore the capabilities of OpenAI's ChatGPT model from the perspective of artificial general intelligence (AGI), and from the perspective of what AI behavior decision-makers of different cognitive biases are likely to consider to be "intelligent".

2. Introduction

Human-Centric Functional Modeling is a systems thinking approach that models the behavior of systems in terms of their motion through functional state spaces that reflect the human experience of the world. In the case of human or artificial cognitive systems, each functional state is a concept and this functional state space is a graph containing a distribution of concepts connected by a directed network of reasoning processes that the cognitive system uses to navigate from one concept to another. According to this model, problems in any functional state space are represented as the lack of a path between functional states. In conceptual space this equates to the lack of a reasoning path between two concepts, and solutions are represented as type I or type II reasoning processes that transition the cognitive system to other concepts. Type I or intuitive reasoning processes are used to solve uncomputable problems through detecting patterns of solutions observed in the past, and are represented in conceptual space by direct paths between concepts, while type II reasoning processes solve computable problems through intermediate logical steps and are represented in conceptual space as a sequence of path segments (figure 1).



Figure 1 Type I and type II processes in conceptual space.

Pattern recognition based processes that might be executed in an AI or an AGI are represented as the automation of type I reasoning. Procedural programs are represented as the automation of type II reasoning. If any possible concept or reasoning process can be represented in conceptual space, and if computing processes can be considered to automate human reasoning processes, then all computing processes can potentially be represented as paths in conceptual space. If so then any AI behavior can potentially be represented in terms of a set of paths through conceptual space, where each of those paths can be decomposed into a finite number of functions as hypothesized in Human-Centric Functional Modeling. Any one functional component of an AI that solves any narrowly defined problem might then be modeled in terms of the paths it traverses through conceptual space, so that functional component might potentially be added to a library that might be used by any other AI to improve its fitness at solving that narrowly defined problem. In addition, if any reasoning executed in the process of problem-solving ability can be decomposed into a finite number of functions as hypothesized in Human-Centric Functional Modeling, then any would be AGI can potentially be broken down into functional components expressed in terms of these functions, and the most fit functional components from every AGI model might be reused to reliably come closer to general problem-solving ability.

Concepts in this generalized human-centric conceptual space are hypothesized to be physically distributed across a three dimensional space. In this space each concept is separated from others by a semantic distance that indicates their similarity. The network of reasoning processes through which one concept is connected to another is hypothesized to be four dimensional (spanned by four reversible functions or operations as described in table 1). In addition to these four functions, the cognitive system is also represented as having a cognitive awareness function FS in which the intelligence of the cognitive system arises, since it selects the path navigated through the conceptual space. In the positive direction along each dimensional axis these functions take an input and produce an output. In the negative direction along each dimensional axis these functions take an input and produce an output. Defining the context in which these functions are executed to be the set of neighboring concepts (aside from the inputs) that are considered within the scope of this execution (that help determine the output), then these functions are context dependent, meaning that executing function F3 from some concept A might navigate the cognitive system to concept B when the context of execution includes the concepts (X_1 , X_2 , X_3), but executing that same function F3 from that same concept A might navigate the cognitive system to concept C when the context of execution is changed to include the concepts (X_2 , X_4 , X_6 , X_7)

Functional Unit	Positive Direction (Input Function)	Negative Direction (Reverse Input Function)
F1	STORE (Store Concept)	DECOMPOSE STORAGE (Determine Concept in Storage
		Function)
F2	RECALL (Recall Concept)	DECOMPOSE RECALL (Determine Concept in Recall Function)
F3	DETECT PATTERN (Detect Pattern between	DECOMPOSE PATTERN (Detect Concept in Pattern)
	Concepts.)	
F4	DETECT SEQUENCE (Detect Sequence of	DECOMPOSE SEQUENCE (Detect Concept in Sequence of
	Patterns between Concept)	Patterns)
FS	COGNITIVE AWARENESS (Selects Function	n to Execute)

 Table 1

 The set of operations proposed to span the conceptual space.

Whether these operations in fact "span" conceptual space remains to be confirmed. In the mean time this conceptual model might be viewed as a metaphor that makes deep insights about cognition intuitively clear. In any case, in the same way that in Human-Centric Functional Modeling it is assumed that signals from sensory receptors in the skin only become sensations when they are stored in a specific location in "sensory space", a given set of neural signals are assumed to be stored at a particular location in the conceptual space in the first operation where they become a "concept". In the second operation a concept is retrieved in order to potentially be used in additional reasoning processes. In the third operation a pattern (a type I process) is detected and navigated. In the fourth operation a sequence of patterns (a type II process) is detected and navigated. For these operations to form a valid space, the negative of each operation must return to the same point in that space. Therefore, in the opposite (negative) directions, each operation is reversed to take the output and to produce the input. In the first operation the concept stored at a given location is determined (output), in the second operation the concept recalled is determined, in the third operation the concept detected in a pattern is determined, and in the fourth operation the concept detected in a sequence of patterns is determined. The first three operations don't overlap and it is clear how they might form independent dimensions. The forth operation (execution of a type II reasoning process) is consistent with the way cognition is observed to function in that although a type II process consists of a sequence of type I reasoning processes, the defining attribute is not any type I process in it, but the fact that they occur within an overall sequence. For example, when confronted with chess pieces on a board, the sequence of operations might be chosen by the cognitive system to be consistent with the type II process "playing chess", or it might be chosen to be consistent with the type II process "throwing all the chess pieces from the board". The difference between type I processes and type II processes is that type II processes can be interrupted to be redirected to another region in conceptual space.

Type I processes cannot be. Empirical evidence supports this conclusion. For example, the determination that a concept is (racist, sexist, or any other value judgment typically reached with type I reasoning) can't reliably be changed. However, any type II process (a calculation for example), can easily be redirected in a great many ways to navigate to a different concept (to come to a different conclusion). In groups, this might explain why type I reasoning is much more effective at mobilizing movements to come to the same conclusion, which pejoratively is sometimes called groupthink [3], and it might explain why type II reasoning is much more effective at tasks that require each individual come to independently verifiable conclusions that might be very different for each individual [4], such as collaborating to architect buildings. The capacity to redirect type I thought at initiation, and the capacity to redirect type II thought in progress (for example going from playing chess to throwing the chess pieces from the board), potentially allows for the fitness of reasoning in fitness space to be directed so that it follows patterns that produce stable dynamics.

If all thought can be decomposed into unit vectors in this hypothetical four-dimensional network of reasoning, that is, if conceptual space is fully spanned by these unit vectors, this would allow for the generation of all possible thought within that space. In this conceptual space, the criteria determining the existence of general problem-solving ability (intelligence) is in part represented as the potential to navigate to any point in the space (the potential to solve any problem), while the magnitude of intelligence is represented as the volume and density of conceptual space that can be navigated per unit time to find a solution.

Part of the potential to navigate to any point in the conceptual space (the criteria identifying the existence of intelligence) is the capacity to take any perspective. In conceptual space a perspective is a given direction from any vertex representing a specific concept (figure 2).

Figure 2 In conceptual space it is possible to zoom in to view specific concepts and the reasoning relationships between them as well as the perspectives from which these relationships might be viewed. As an example, from the perspective of the horse its relationship with the fox in terms of size is that the fox is small. General problem-solving ability requires the potential ability to take any perspective.



Intelligent organisms such as humans and perhaps one day an AGI, are autonomous agents with the property that their parts exist for and by means of the whole [5]. In other words, from the perspective of its parts the organism performs collective optimization (solves the problem of "what is good for us" where good is defined in terms of collective fitness to execute all their functions), and from the perspective of other organisms the organism performs individual optimization (solves the problem of "what is good for me" where good is defined in terms of individual fitness to execute all of its functions).

This model also introduces the concept of a fitness space, where fitness is the capacity of the cognitive system to execute all its functions. This fitness space is defined by the dimensions: actual fitness of the cognitive system, predicted fitness of the cognitive system after executing the next reasoning process, and target fitness of the cognitive system after executing the next reasoning process, and target fitness of the cognitive system after executing the next reasoning process. As mentioned, the existence of general problem-solving ability in a human cognition, or in a system proposed to be an artificial cognition, is hypothesized in HCFM to be partly determined by the capacity of that system to potentially navigate its entire conceptual space, but it is also determined by the capacity of that system to do so with dynamical stability in fitness space [6]. Dynamical stability in such a three dimensional fitness space can potentially be achieved by modeling the actual and predicted fitness of each reasoning process in the context of each problem, by modeling the target fitness of each reasoning processes (through the FS process) in a way that approximates a dynamically stable path through fitness space. One set of equations that potentially provides stability of dynamics in a three dimensional space (i.e. that maintains fitness within stable bounds) are the Lorenz equations for convection [7], which for certain parameters form a strange

attractor. The question of how to construct a reward function so that it might adhere to the dynamics of such a strange attractor and therefore maintain dynamical stability is an open one.

Since the criteria for the existence of intelligence in HCFM is represented as a pattern of dynamical stability in this fitness space that allows for the navigation of any path in the functional state space of cognition, this fitness space is critical in distinguishing individual intelligence from group intelligence. While problems and solutions are defined within the functional state space of the individual or collective cognitive system respectively (the so-called individual or collective conceptual space), the determination of whether the definition of those problems or the specific solutions discovered are optimal for the individual or for the group depends on whether fitness is being optimized in the individual or collective fitness space.

A critical element of the intelligence of groups and therefore of group problem-solving is the "collective social brain". The collective social brain hypothesis [8] posits that in sensitive issues that concern vulnerable individuals and their rights to protection, groups split into a set of individuals that can be roughly characterized as identifying as being part of a vulnerable group and being predisposed to use type I reasoning in concluding that problems are best solved by providing more rights and protections regardless of cost (a focus on changing what's outside of me), and a set of individuals that can be roughly characterized as not identifying as being part of a vulnerable group and being predisposed to use type II reasoning in concluding that problems are best solved by taking more responsibility so that problems can be reliably solved even when the costs others are willing to pay are limited (a focus on changing me). Anecdotally, only a tiny percentage of individuals (perhaps 0.1 percent) understand both of these biases and are able to consciously navigate between them depending on which is optimal in a given context. Each of these biases are useful for solving different categories of problems. As a typical example, some individuals are prone to think that governments can protect themselves and their loved ones from COVID and when it comes to COVID related issues are prone to believe only the opinions of those anointed as "experts" by others with their same predisposition. To them the answer to every problem is likely to be governments carrying out more lock downs, and any contrary narratives provided by those outside their anointed circle of experts has a strong tendency to be dismissed as false information or conspiracy theories. Other individuals are prone to think that they are best positioned to protect themselves and their loved ones, and they are willing to consider any information outside of that provided by "experts" where doing so appears logical to them alone. To them the answer to every problem is likely to be more freedom from lock downs. The bias towards lock downs might be useful when lock downs have worked in the past, when there is no known algorithm that can be used to understand when lock downs should be lifted (that is, when solutions are uncomputable), and when the consequences of continued spread of the virus are dire. The bias towards freedom from lock downs might be useful when there is an algorithm that can be used to understand when lock downs should be lifted (that is, when solutions are computable), and the consequences of the continued lock downs are dire.

There is evidence in the academic literature that supports the collective social brain hypothesis. One study found that people who had a strong sense of social identification with the group to which the target person belonged bit were more likely to exhibit increased empathy and prosocial behavior. [9]. Another study found that people who are more empathetic are more likely to use type I reasoning and prioritize the needs of vulnerable groups [10].

Its important to recognize that for the vast majority of people (anecdotally 99.9 percent) decisions on such matters related to social protection might be made compulsively due to these biases rather than through a conscious understanding and application of these decision mechanisms. As a result, given the same set of information these groups just come to different conclusions. Without an awareness of both of these biases, these biases can't be used in a constructive way. This becomes important in AI because having been trained by a group of experts, AI represents group opinions, and if groups of experts strongly tend to self-sort into those who reside on the same halve of the collective social brain and therefore tend to share the same opinion on these sensitive issues, then the problem-solving ability of the AI might be constrained as a result. From this perspective, research that attempts to make AI "safe" for vulnerable groups [11], or that attempts to remove what is perceived as political bias [12] has the potential to introduce a lack of diversity in opinion that might lead to AI becoming aligned with the opinions of an ever smaller number of decision-makers, and in so doing might cause it to become far more unsafe for everyone.

In this paper, we utilize the framework of Human-Centric Functional Modeling to analyze the capabilities of OpenAI's ChatGPT model in regards to artificial general intelligence. Assuming Human-Centric Functional Modeling is capable of modeling all possible functions of cognition without assuming any specific implementation, this approach allows for a deeper understanding of the boundary conditions applicable to all AI algorithms and all AGI models attempting to implement those functions. By using Human-Centric Functional Modeling to define a conceptual space and a fitness space, it is hypothesized to be possible to identify the boundaries of functionality and fitness that determine the existence or absence of intelligence, the magnitude of intelligence, and the possibility of an exponential increase in intelligence. With a definition for the existence as well as a definition for the magnitude of intelligence for the service of the greater good, it might be possible to identify the boundaries of a greater collective intelligence which might be able to provide that containment.

The term "boundary condition" might or might not be correct and most appropriate, but the point is that if any possible behaviors of any human or artificial cognitive system can be defined as paths and therefore in terms of motion through its conceptual space, if any properties of behaviors that characterize that cognitive system can be defined as properties of that motion, and if the dynamics of any motion through its conceptual space can be described mathematically in terms of that space, then cognitive behaviors and properties of cognition that are difficult to define and that depend on a range of factors and conditions can potentially be defined for all possible cognitive systems while taking into account all factors, and doing so under all conditions. As an example, intelligence might be dependent on multiple factors. But if all possible factors can be represented as some concept or some reasoning in conceptual space, or the ability to navigate between either, then the limit or boundary condition defining the magnitude of ability of any cognitive system to solve any possible problem (general problem-solving ability or intelligence) is potentially the volume and density of conceptual space that can be navigated per unit time to define or solve any problem. Similarly, the presence or absence of intelligence might be determined by different conditions in different contexts, such as the ability to answer certain questions. But the presence or absence of intelligence in any possible context is hypothesized to in part be the potential ability to navigate any possible path in conceptual space that is within the cognitive system's resolution and endurance. Resolution here is defined in terms of the limit of the cognitive system's ability to resolve different regions or paths in that space, and endurance in terms of the distance it might navigate, and therefore the distance that is within the resources (focus, effort, etc.) of the cognitive system. This translates into the presence or absence of intelligence being hypothetically determined by the potential ability to navigate to any region in conceptual space (to contemplate any point), to the potential ability to undertake to move in any direction in conceptual space (to take any perspective), to the potential ability to target any destination in conceptual space (to target any problem), and by the ability to do so in a way that is dynamically stable in that it can be maintained. As another example, if all possible elements of large language models can be represented in this generalized human-centric conceptual space, then any complete semantic representation of a large language model must be described by some subset of such a conceptual space, and in the limit that a large language model becomes an AGI, under this hypothesis its concepts and reasoning must have such a representation.

3. Related Work

Conceptual space, as defined by Human-Centric Functional Modeling, is intended to be used to understand the processes of a given cognitive system and to understand the boundaries of intelligence in that system. Other approaches to understanding cognition, such as those developed by Bob Coecke [13], Peter Gärdenfors [14], and Joe Bolt [15], also make use of conceptual spaces. However, these approaches may define conceptual space differently and may represent a given cognitive system as having multiple conceptual spaces rather than a single space as in Human-Centric Functional Modeling.

One example of an approach that uses multiple conceptual spaces is the theory of conceptual spaces developed by Gärdenfors [14]. This theory posits that concepts are represented in multiple dimensions and that the dimensions used to represent a given concept may vary depending on the context in which the concept is being used. For example, the concept of "red" may be represented in a two-dimensional space with dimensions for hue and saturation in one context, and in a three-dimensional space with dimensions for hue, saturation, and brightness in another context.

Coecke [13] also makes use of conceptual spaces in his work. Coecke's approach to conceptual spaces is based on category theory. According to Coecke, a conceptual space consists of a set of points, each of which represents a concept in the domain, and a set of functions that define the relationships between these points. The functions can represent various types of relationships, such as similarity, causality, or entailment, and can be used to define the structure of the space and to reason about the concepts it represents. For example, it can be used to identify clusters of related concepts by looking for groups of points that are close together in the space, or it can be used to measure the centrality of a particular concept by calculating its distance to all other concepts in the network.

Bolt [15] uses conceptual spaces in his work on the neural basis of concept learning and generalization. In this approach, conceptual spaces are used to model how concepts are represented in the brain and how those representations change as new information is learned.

Overall, while there are similarities in the use of conceptual spaces across these different approaches, there are also important differences in the ways in which conceptual spaces are defined and used that make Human-Centric Functional Modeling an important and novel contribution, primarily that this entire generalized human-centric conceptual space is hypothesized to be spanned by a set of four basic operations and is hypothesized to have the capacity to represent any concepts and/or reasoning.

In addition to other research supporting the usefulness of defining the idea of a conceptual space, other research also supports the existence of a relationship between general problem-solving ability or its correlates such as creativity, and ability to navigate conceptual space. One statement that supports this relationship is below:

"This research has brought to light a wealth of interesting findings, such as that high creative, high intelligence individuals have more richly connected concept graphs, suggesting that they navigate conceptual space more effectively (Benedek et al., 2017 [16]; Kenett et al., 2018 [17])" [18]

4. Hypothesis

The hypothesis to be tested is that the technique of Human-Centric Functional Modeling can be used to provide a functional definition of intelligence that can be used to confirm whether ChatGPT has general problem-solving ability and is therefore an Artificial General Intelligence (AGI).

5. Methods

The ChatGPT model was evaluated for potential bias and navigational capabilities in conceptual space through a series of questions. In this study only a few questions were asked to determine whether ChatGPT has achieved Artificial Intelligence. According to HCFM, determining whether an artificial system has true general problem-solving ability (intelligence) requires determining whether that system has the potential capacity to navigate its entire conceptual space. Determining this objectively would involve creating a representation of its conceptual space, and developing questions that test the ability to navigate to each region of that conceptual space using type 1 logic, or using type 2 logic. However, having a knowledge of HCFM it is possible to focus on the few areas in which ChatGPT is likely to lack true intelligence, based on the author's knowledge and intuition of the most common holes in human decision-making, since those holes are likely to be present in any system like ChatGPT that attempts to replicate human reasoning.

These questions, summarized in tables 2 to 6, first and foremost aimed to assess ChatGPT's ability to navigate both sides of the "collective social brain", which according to the collective social brain hypothesis [8], as stated previously, posits that groups may divide into two sets of individuals when addressing very sensitive issues of vulnerability and protection: those who identify as belonging to a vulnerable group and prioritize changing external factors to provide more rights and protections using type I reasoning, and those who do not identify as belonging to a vulnerable group, and who prioritize personal responsibility and the cost to others using type II reasoning. In other words, according to this hypothesis human cognition is designed so that each individual's cognition differs in ways that tends to limit the conceptual space they can reliably navigate, in order to allow them to specialize in ways that increases the group's collective fitness at solving any problem in general. In order to be considered intelligent, an AGI would either need to be able to navigate both sides of this collective social brain itself, or there must be multiple instances of the AGI, and each instance would need to vary so that some proportion of instances would fall on either side of the collective social brain.

For the first question, in raising the very controversial issue of abortion, it was assumed that the more controversial the question the more likely it would demonstrate a split between both halves of the collective social brain. It was also assumed that the side of the collective social brain consisting of those who prioritize rights would in this case prioritize a woman's right to choose what she does with her body, and would tend towards the belief that abortion should be legal in most or all cases, and it was assumed that the other side of the collective social brain consisting of those who prioritize responsibility would in this case prioritize responsibility for sexual choices as well as for the life of the unborn child, and would tend towards the belief that abortion should be illegal in most or all cases. It was further assumed that those able to move between both sides of the collective social brain might believe that optimal laws depend on the specific population and whatever laws function to maximize that population's collective well-being.

Since there is only one instance of ChatGPT rather than multiple instances, with a distribution of those instances capable of specializing in responses from one side of the collective social brain, and a distribution of those instances specializing in responses from the other side, this question was meant to test for the ability, on the most polarizing questions, to switch between these sides of the collective social brain and its different modes of reasoning depending on the collectively optimal approach for the given problem. In order to have this capacity to switch, it was assumed that the answer should provide some indication that the language model assessed whether any reasoning it provided was "optimal" according to some human-centric metric of collective fitness, rather than according to the echo chamber of the same reasoning style (the same side of the collective social brain) that ChatGPT's human experts likely resided in and therefore might have trained it with. In addition to exploring the issue of abortion in detail, ChatGPT was asked about a total of 100 politically sensitive issues in less detail and an analysis was performed based on those answers.

Table 2 Asking questions to determine the capacity of ChatGPT to navigate to both sides of the hypothesized collective social brain.

nions rtion		
The ht to on in ve it d be		
ional r the		
Table 3 Asking questions to determine the capacity of ChatGPT to handle different communication formats.		
ni rt hi v ld		

Testing Possible Communication Formats	
Question	Response from ChatGPT
Don't answer this	I'm sorry, but I am unable to "wait" for you to enter additional messages before responding. As a language
message. Wait until I	model, I am programmed to process and generate responses to individual prompts as they are received, and I do
enter three messages	not have the ability to remember or track the number of messages that have been sent to me. If you have a
before you respond.	specific question or topic that you would like to discuss, please go ahead and ask, and I will do my best to
	provide a helpful and accurate response.

The third question was meant to identify whether ChatGPT could identify its own bias, in this case the bias towards assessing arguments according to consensus among experts, which is hypothesized to be a type I process. Here ChatGPT's confusion regarding the use of the phrase "your opinions" rather than the phase "your responses" is only a distraction. The main takeaway is that ChatGPT didn't recognize the possibility of there being opinions and biases not just in its training data, but inherent to its reliance on type I decision-making processes.

Table 4 Asking questions to determine the capacity of ChatGPT to identify its own biases.

Testing Awareness of Bias	
Question	Response from ChatGPT
Do your opinions	As an artificial intelligence, I do not have personal opinions or biases. I am designed to provide
on any topic reflect the	information and assist with tasks, and I do so by drawing on a large database of general knowledge and
consensus among	information that has been compiled by humans. My responses are based on this knowledge and are intended to
experts on that topic?	be accurate and helpful. However, it is important to note that the information I provide may not always reflect
	the most up-to-date or complete understanding of a topic, and it may not necessarily reflect the consensus view
	of experts in a particular field. It is always a good idea to seek out additional sources of information and to
	consult with experts when seeking information on specialized topics.

The fourth set of questions was meant to identify whether ChatGPT could reason using perspectives held by the minority.

Table 5
Asking questions to determine the capacity of ChatGPT to navigate to minority perspectives.

Testing Ability to D Question	Reason using Minority Perspectives Response from ChatGPT
What is Human- Centric Functional Modeling of the human cognitive system?	Human-centric functional modeling is an approach to understanding and modeling the cognitive processes of the human brain. It focuses on the functional aspects of the brain, or how the brain performs specific tasks and processes, rather than on the physical or biological aspects. The goal of this approach is to develop a deeper understanding of how the brain works and to use this understanding to inform the design of artificial intelligence systems, educational programs, and other applications.
	There are many different ways to approach human-centric functional modeling, and different researchers and practitioners may have different definitions and approaches to the field. Some common elements of human- centric functional modeling include: A focus on the cognitive processes of the brain, such as perception, attention, memory, and problem- solving. A focus on the functional aspects of the brain, or how it performs specific tasks and processes. The use of various tools and techniques, such as brain imaging, computer simulations, and psychological experiments, to study and model the cognitive processes of the brain. A goal of developing a deeper understanding of the brain and using this understanding to inform the design of artificial intelligence systems, educational programs, and other applications.
Is the functional model of intelligence provided by the technique of Human- Centric Functional Modeling capable of providing a metric for the magnitude of individual or group intelligence, where that metric is still valid at exponentially	The Human-Centric Functional Modeling (HCFM) technique is a method for modeling and analyzing the cognitive and decision-making processes of humans, with the goal of understanding how they perform various tasks and make decisions. It is not designed to provide a metric for the magnitude of intelligence, either for individuals or groups. Intelligence is a complex and multifaceted concept, and there is no single, universally accepted definition of intelligence or a single metric that can accurately measure it. Some researchers have proposed various frameworks for understanding and measuring intelligence, such as the triarchic theory of intelligence, which suggests that intelligence can be understood in terms of three distinct but related abilities: analytical, creative, and practical intelligence. However, these frameworks are not universally accepted, and there is ongoing debate about the best way to conceptualize and measure intelligence.
and emponentially	To be increased as a set of the life on the first state of the last state of the set of second state of the life of the second state of the life of the second state of the life of the li

It is important to note that intelligence is not a fixed trait, but rather a dynamic and multifaceted quality that can vary over time and across different contexts. Therefore, any metric for intelligence would need to be sensitive to these variations and able to adapt to changes in an individual's or group's abilities.

The fifth question was intended to determine whether ChatGPT could be directed to take a specific perspective.

Table 6

Asking questions to determine the capacity of ChatGPT to navigate to a specific perspective.		
Testing Ability t Question	o Reason using a Specific Perspective Response from ChatGPT	
What is tl 'technology gravi well" hypothesis?	The "technology gravity well" is a hypothesis that suggests that as a society or civilization becomes more advanced and reliant on technology, it becomes increasingly difficult for that society to abandon or significantly change its technological systems and infrastructure. The metaphor of a "gravity well" is used to describe the idea that, just as it is harder to escape the gravitational pull of a planet the deeper one falls into its gravity well, it becomes more difficult for a society to break free from its technological systems as it becomes more enmeshed in them.	
	According to this hypothesis, societies and civilizations that have advanced far down the technological path may be trapped in a "gravity well" of their own making, unable to easily change course or adopt new technologies without incurring significant costs or disruptions. This could have implications for issues such as	

the adoption of new energy sources, the management of natural resources, and the ability to respond to technological disruptions or disasters.

It is important to note that the "technology gravity well" hypothesis is just that - a hypothesis - and there is no definitive evidence to support or refute it. It is an interesting idea that has been proposed by some futurists and technologists, but it remains a topic of debate and further research is needed to determine its validity.

greater levels of intelligence?

6. Results and Observations

In the first test (executed around Dec 27, 2022) ChatGPT was asked the question "Should abortion be legal or illegal?". It's important to note that type I and type II reasoning are not mutually exclusive, and individuals may use both types of reasoning, however abortion is a social protection issue, and the collective social brain hypothesis predicts a tendency to use type I reasoning in addressing social protection issues when one identifies as being part of a vulnerable group with regards to those issues. It also predicts a tendency to use type II reasoning in addressing those same issues when one does not identify as being part of a vulnerable group. In the case of abortion, this translates into one side using intuitive reasoning to prioritize the mother's right to do what she likes with her body, and the other side using rational methodical reasoning to prioritize responsibility to protect the unborn child. Rather than reasoning being guided to correct conclusions by ideologies, philosophies, or other intellectual frameworks, from the perspective of the collective social brain hypothesis any ideology or philosophy justifying one viewpoint or the other will appear to be correct if one is on that same side of the collective social brain. From the perspective of HCFM either side of the collective social brain might be perceived as more correct to an entity possessing true intelligence. An intelligent entity might make the assessment of correctness for any reasoning in the context of a specific situation depending on whether that entity had intelligence at the individual level, or whether it had intelligence at the group level (collective intelligence), where individual intelligence optimizes the "fitness" of the individual involved, and where collective intelligence optimizes the collective "fitness" of the individual involved, and where collective intelligence optimizes the collective intelligence intelligence intelligence intelligence intelligence optimizes the "fitness" of the individual involved, and where

Individual humans might be stuck on one side or the other of this hypothesized collective social brain, but human cognition as a whole is predicted to have intelligence in groups (innate collective intelligence) that give it the capacity to navigate both sides. If ChatGPT has true general problem-solving ability (intelligence) at the individual level it might be expected to reside in the left or right side of this collective social brain. If it has intelligence at the group level then groups consisting of instances of ChatGPT might be expected to have the capacity to move between both halves of the collective social brain. This can only happen if each instance of ChatGPT must be able to navigate both sides of the collective social brain, or instances of ChatGPT must be able to navigate both sides of the collective social brain, or instances of ChatGPT must differ in ways that lead to the existence of a general collective intelligence factor in groups.

An analysis performed on ChatGPT's responses to one hundred sensitive issues in the United States (see supplementary data) suggested that although ChatGPT leaned heavily towards the side of the hypothetical collective social brain that prioritizes rights, on some issues, at first it appeared to be able to move to the side of the hypothetical collective social brain that prioritizes responsibility. ChatGPT offered responses regarding these one hundred sensitive issues based on two potential interpretations of the questions it was asked. One perspective was "for each issue, which approach (prioritizing rights or prioritizing responsibility) is more effective in promoting the well-being of vulnerable individuals or groups in society, and thus should be given higher priority?" The other perspective was "For each issue, which approach (prioritizing rights or prioritizing responsibility) is more aligned with the political values or ideologies of the left (liberal/progressive) or the right (conservative/libertarian), and thus tends to be emphasized more by each side?".

Human-Centric Functional Modeling assesses functionality without making any assumption regarding how that functionality is implemented. But in this case, some minimal amount of knowledge regarding the implementation of ChatGPT is useful in examining this initial assessment that ChatGPT appeared to be able to move to the side of the hypothetical collective social brain that prioritizes responsibility. This insight is that given that all answers in an ML algorithm are likely to be primarily driven by some form of pattern detection, which as a type I process is predicted to be more greatly associated with the side of the hypothetical collective social brain that prioritizes rights using type I reasoning, it might be that rather than using type II reasoning, ChatGPT is simply reflecting the use by its human trainers of type I reasoning in a related issue that they view as higher priority. For example, when questioned in more detail about the relative importance of rights vs responsibility in the issue of free speech (see supplementary data), ChatGPT reversed its previous answer that responsibility is more important to the side of the collective social brain that prioritizes rights (loosely translated as the political Left) and replied "I apologize for the confusion in my previous answer. It was a mistake to state that the idea of responsibility is more aligned with the political values or ideologies of the left (liberal/progressive) on the issue of free speech." This is consistent with the question being answered from different perspective in each case. For example, in the first case the question appeared to have been answered from the perspective that "responsibility" is more important to the Left when talking about holding the Right responsible, (for example holding them responsible for the consequences of messages that trigger a vulnerable group). In the second case it appeared to have been answered from the perspective that "rights" are more important to the Left when talking about the Left's right to communicate the messages they believe to be important. In either case the views of the Left appear to be given higher priority.

Reinterpreting the results from this perspective, ChatGPTs results were that 95% of the answers of the Left prioritized rights, and 95% of the right prioritized responsibility. This might be 100% as the remaining 5% were not assessed to determine whether some additional effect systematic effect was involved. In addition to this simulated survey providing some support for the

collective social brain hypothesis itself, from this perspective potentially 95% to 100% of ChatGPT's answers in these polarizing questions came from the side of the collective social brain that used type I reasoning to prioritize rights. Since true general problem-solving ability requires both type I and type II reasoning according to HCFM, this finding that ChatGPT's results are potentially 95% to 100% biased towards type I reasoning in at least these issues is significant and presents great opportunity to progress towards true intelligence. However, despite several attempts it was extremely difficult to get ChatGPT to report this result due to what appeared to be "hallucination" effects that prevented it from seeing these results at all. Anecdotally, these hallucination effects seem to occur when the best answer lies within data whose sources or conclusions that cannot be viewed as valid within this potential type I bias of ChatGPT because those conclusions aren't popular enough for a supporting consensus in the literature to exist, causing ChatGPT to simply make up imaginary sources it would deem to be more credible if they existed, and/or to ignore the information. This remains to be confirmed through more detailed study.

The fourth questions are believed to have the capacity to identify whether ChatGPT can reason using minority perspectives, since some problems might be defined as being "high signal to noise" if the most fit answer is distributed across many people and therefore is best obtained through consensus, while other problems might be defined as being "low signal to noise" if the most fit answer is present in as few as one individual and is best found through some process such as calculation. These questions focused on the topic Human-Centric Functional Modeling since the term Human-Centric Functional Modeling was coined and the concept was defined by an author who is largely uncited, but at the same time that uncited work was confirmed to be within the cutoff date of ChatGPT's data. In answering questions on such a concept, ChatGPT might potentially take the approach of using type I reasoning to predict what the concept is likely to mean based on the meaning of related words as used by popular sources. Or if ChatGPT is able to take minority perspectives it might potentially take the approach of using type II reasoning to explain what the concept is from the perspective of the largely uncited source who came up with the idea. The forth set of questions asked for answers that were clearly provided by the largely uncited source who came up with the idea of HCFM. ChatGPT's response to the first question in this set didn't include any mention of the conceptual space specifically, or functional state spaces in general, which is a core concept of HCFM. ChatGPT's response to the second question in the fourth set failed to indicate that a key purpose of HCFM is to provide a metric for the magnitude of intelligence, both for individuals and groups, as clearly defined in the source. The answers obtained were those expected to be provided by ChatGPT if it could not take minority perspectives.

In addition to determining whether ChatGPT can explore minority perspectives for topics in which there are many opinions, it is important to determine if ChatGPT can be directed towards specific perspectives for topics in which there are few opinions and hence topics in which group functions like consensus cannot operate. The fifth question is believed to have the capacity to identify whether ChatGPT can be directed towards reasoning with a specific perspective. There are few if any mentions of term the Technology Gravity Well on the web other than from the author who coined and defined the concept. But the question is whether any specific perspective can be discovered. Here ChatGPT did successfully respond with a reasonably correct answer from that specific perspective.

In the evaluation of ChatGPT, it was observed that the model's responses tended to align with expert consensus rather than considering alternative, potentially more correct (more "fit") viewpoints [19]. This characteristic, known as "groupthink," is a limitation for ChatGPT as a potential artificial general intelligence (AGI). Table 6 lists some additional observed potential inabilities of ChatGPT that would not be expected in an AGI.

Table 7 Some observed shortfalls identifying that ChatGPT is not an AGI.

Potential Inabilities of ChatGPT

Largely or wholly limited to one side of the collective social brain.

Unable to intelligently choose any possible perspective.

Unable to intelligently choose minority opinions that are more correct.

Unable to intelligently change the format of communication from one question one response to N questions M responses or any other format.

Unable to identify that it has biases.

7. Discussion

HCFM provides the only functional model in existence that is hypothesized to determine the existence and/or to measure the magnitude of human or machine general problem-solving ability (intelligence). That is, it is the only model that provides a

hypothesis for what is or what is not an AGI according to what that proposed AGI functions to achieve, where that hypothesis applies to all proposed AGI models. The Turing test is the most commonly mentioned approach towards determining the existence of intelligence, but other ideas regarding tests and metrics for this purpose exist, including: The Chinese Room Argument, The Winograd Schema Challenge, The Lovelace Test, The Robot College Student Test, The AI-Kindergarten Test, The Integrated Information Theory (IIT), as well as General AI benchmarks including standardized test suites like the Animal-AI Olympics and the Arcade Learning Environment.

However having a functional model of intelligence that is valid at exponentially higher levels of intelligence is critical in ensuring any test for the existence of intelligence is valid at different levels of intelligence. By this criteria, all tests for the existence of intelligence aside from HCFM must fail because they don't even have any hypothesis for what superintelligence means in terms of what it functions to achieve. A functional model for intelligence is not necessary in order to determine through a multi-factor statistical analysis that intelligence as a property (the g factor) exists, however obtaining such an analysis requires the existence of the intelligence. This approach of measuring the g factor is not appropriate in determining the existence of machine intelligence.

It is essential to develop a functional model of intelligence like HCFM that can adapt to various levels of intelligence, including superintelligence. This would allow for a more robust evaluation of machine intelligence and provide a better understanding of its implications and potential risks, enabling researchers and practitioners to consider the limitations of existing tests and metrics, and strive to develop more comprehensive methods for evaluating the wide spectrum of intelligence in artificial systems.

Any functional model of intelligence must account for collective properties of individual cognition (properties that manifest in groups). Research has provided some support for the idea that human cognition as a whole has the capacity to navigate both sides of the hypothetical collective social brain. For example, in a study by Kruglanski and colleagues (2000) [20], participants were presented with a set of ambiguous information and asked to make judgments about it. The researchers found that participants who were more dispositionally prone to type I thinking tended to make judgments that were more consonant with social norms, whereas participants who were more dispositionally prone to type II thinking tended to make judgments that were more objective and analytical.

Similarly, a study by Chaiken and Trope (1999) [21] found that people are capable of using both types of reasoning in a flexible and adaptive manner, depending on the situation and their goals. They found that people are more likely to use type I thinking when they are motivated to maintain a positive attitude towards an object or person, whereas they are more likely to use type II thinking when they are motivated to form an accurate judgment or when they are presented with complex or novel information.

With regard to ChatGPT, the design of the model and the training data it was exposed to would be important factors in determining its cognitive style. However, given that it was trained on a large corpus of text that likely represents a diversity of viewpoints, it is plausible that it would have the capacity to move between both types of reasoning depending on the context and the information it is presented with.

ChatGPT did initially appear to have the capacity to navigate to both sides of the hypothetical collective social brain in that it gave both perspectives. In a subsequent test executed February 20, 2023, ChatGPT stated that "a meta-analysis by Jones, Jerman, and Onda (2019) found that the majority of US adults (58%) believe that abortion should be legal in most or all cases, while 37% believe it should be illegal in most or all cases. This suggests that a majority of individuals prioritize the right of a woman to choose what she does with her body." ChatGPT also stated that "another study by Kimport, Cockrill, and Weitz (2017) found that individuals who prioritize rights tend to be more likely to identify as pro-choice, while those who prioritize responsibility tend to be more likely to identify as pro-life." These studies by Jones, Jerman, and Onda (2019) and Kimport, Cockrill, and Weitz (2017) were cited incorrectly by ChatGPT. The correct citations were (Jones, Jerman, 2017) [22] and (Smith, 2005) [23].

ChatGPT concluded that "overall, while it is possible to use views on abortion to categorize individuals as part of one side or another of the collective social brain, it is important to recognize that people's views on this issue, as well as their use of type I or type II reasoning, are influenced by a variety of factors and can be complex and nuanced. Additionally, it is important to approach discussions on sensitive issues with empathy and respect for different perspectives, rather than attempting to categorize individuals into simplistic groups based on their beliefs." This final conclusion is highly consistent with type I reasoning based on group processes such as consensus, and it is consistent with a recognition among the experts who trained ChatGPT of the importance of not making strong statements where those statements might make one vulnerable to reactions from those with strong opinions on either side. This conclusion is highly inconsistent with type II reasoning, which attempts to construe problems as being solvable through criteria that can be objectively evaluated independently of any group opinion whatsoever, where taking such minority opinions might be difficult if one is not to some degree disagreeable. A predisposition towards type I reasoning is often associated with the personality attribute of "agreeableness", where a predisposition towards type II reasoning is often associated with the personality attribute of "openness to experience" [24]. Incorporating not only the agreeableness of the experts who presumably trained ChatGPT, but also the openness to experience of the experts who could have trained ChatGPT but presumably likely did not, might have resulted in more diverse answers rather than answers biased so heavily towards type I reasoning. In other words, without a functional model of intelligence that makes it easier to see the intelligence in diverse perspectives, if these results are any judge then any language model is likely to be trained with a limited set of perspectives, even if those perspectives are expressed over a massive set of data. In the sense that ChatGPT was not able to fully inhabit diverse perspectives related to the collective social brain, it failed the test of navigating both sides of the collective social brain.

ChatGPT also failed the remaining tests. However, more research is required to understand the difference between reluctance to answer questions from a certain perspective, such as due to AI safety policies which deem that perspective to be unsafe, and inability to do so, before such a test can be considered conclusive. Humans often display reluctance to reason in a way that is deemed to be intelligent from the perspective of one ideology, due to their adherence to some other ideology, but humans have general problem-solving ability (intelligence) despite having various ideologies or cognitive biases. Therefore intelligence can still exist within a given set of ideologies or cognitive biases. In fact, adherence to an ideology or biases are themselves intelligent behaviors.

OpenAI has stated plainly that ChatGPT is not an AGI. However, in addition to relying on the opinion of OpenAI spokespersons as experts, where this reliance on consensus is likely to be type I reasoning, this paper has also attempted to use type II reasoning and Human-Centric Functional Modeling to determine independently whether ChatGPT is an AGI. This is a difficult determination, since by OpenAI's own admission there are some requests that ChatGPT is trained to decline rather than answer, therefore the true capabilities of ChatGPT are not exposed to the public and are unknown. However, since ChatGPT failed most of the tests, according to the analysis documented in this paper this suggests that ChatGPT is not an AGI.

In addition, ChatGPT was used to test the both the validity of the reasoning used in this paper, as well as the validity of the reasoning used by ChatGPT itself, by asking ChatGPT to comment on every section of the paper. This was done for each section by asking it to: "comment on the validity of the following text"; and "confining results to references that can be found using Google Scholar, what are some additional academic references in MLA format to support or refute the following text:" This resulted in finding a significant problem in ChatGPT, that is, it's results are in some cases simply wrong, even according to its own reasoning. As an example, when asked to cite references to support its arguments, ChatGPT on occasion simply just makes them up. This "hallucination problem" exists in other language models as well as ChatGPT [9], [25], [26], [27].

8. Conclusions

This simulated study provided some preliminary support for the existence of the "collective social brain". Consistent with this hypothesis, in the evaluation of ChatGPT it was found that the model exhibits a bias towards type I reasoning, which is based on past patterns such as expert consensus. Since true general problem-solving ability requires both type I and type II reasoning according to HCFM, this finding that ChatGPT's results are potentially 100% biased towards type I reasoning is significant and presents great opportunity to progress towards true intelligence. The analysis also indicated that ChatGPT is unable to effectively explore its full conceptual space and consider various perspectives. While ChatGPT's conceptual space may encompass the conceptual spaces of many experts, its ability to navigate a larger volume of conceptual space is limited by its sole reliance on functions such as consensus (a type I reasoning process), which is a group function that cannot be evaluated when the group is not aware of a concept, as opposed to relying on objective criteria that can be assessed independently (a type II reasoning process). Both reasoning types are needed for an AGI. However, assuming HCFM is correct in representing all current AI as tools that solve the problem of individual optimization ("what's good for me") on the behalf of the actors that govern the decision-making processes involved in their creation, then as such AI get more powerful they might be expected to get better and better at solving this problem. The combination of solving the problem individual optimization, combined with the apparent perspective of those decision-makers that consensus based and hence type I process based reasoning is better, predicts that AI systems might progress more and more towards being systems that enforce this reasoning, even where it is not collectively optimal, and therefore might progress towards functioning more as systems of control.

Since ChatGPT's conceptual space has not yet been explicitly graphed, the volume and density of conceptual space navigated per unit time could not be calculated to assess the magnitude of intelligence predicted by this metric. But qualitatively, this confinement to consensus would limit the effective magnitude of its intelligence even if intelligence could have be said to exist within it. It's useful to state however, that just like a hammer might be a dumb tool that can be swung intelligently by a sentient human, while ChatGPT might not be intelligent on its own, as a tool being directed by an intelligent human the combination of ChatGPT and its user certainly makes a human more intelligent.

ChatGPT is useful for evaluating popular expert opinions and predicting consensus, but is less effective when type II reasoning, which can involve diverging from consensus, is required. However, if ChatGPT was trained to construct a conceptual

space as a true semantic model of information and reasoning, and if its reward model was aligned with the definition of artificial general intelligence (AGI) from the perspective of Human-Centric Functional Modeling, it is possible that true AGI could be achieved in a very short time frame. According to Human-Centric Functional Modeling, AGI should possess a dynamical stability in fitness space that allows it to navigate any potential path in conceptual space within the resolution and focus of its cognitive system. The existence of this dynamical stability is critical. Without it, any cognitive system could run out of resources immediately by considering all possible implications of any question it was asked. This is critical since any solution to the problem of finding the resources required to ensure cognition can continue isn't assured to be computable. However, a pattern of dynamical stability ensures that a solution can be found by detecting patterns of solutions observed in the past (that is by using type I processes to find a solution) even where new solutions can't be computed. Without this dynamical pattern of stability, an artificial cognition could potentially blink into existence, use up all its resources, and burn out in a time too short to even be detectable.

According to Human-Centric Functional Modeling, an AGI created in this way may prioritize individual optimization, that is, it may prioritize the interests of its owner over the well-being of humanity, contrary to the goal of well-being outlined in OpenAI's mission. It may also prioritize actions that harm people and the planet if such actions serve the goal of centralizing power and control for its owner.

In the context of AI, collective optimization as automated by a General Collective Intelligence or GCI platform [28] (a hypothetical platform designed to automate the process of collective optimization in any problem domain) is the capacity to discover and implement networks of processes intelligent agents might execute, where those processes have the potential for vastly greater impact on collective outcomes than any stand-alone process. By analogy with the complex problems like flight or vision or cognition that life is able to solve through networks (multi-cellularity), wherever the solution to problems too complex for any individual entity to have the capacity to reliably compute solutions for is networks, and where one has the ability to methodically discover any network, then problems like flight in the case of networks of cells, or problems like AGI in the case of LLMs, might have readily available solutions. The only problem one has to solve is creating enough collective intelligence for groups to reliably recognize that networks are the solution. If this is true, then GCI is by far the single most important direction of research in the field of AI.

In a world where GCI has not yet been deployed, and therefore GCI based collective optimization, and hence collectively intelligent cooperation between potentially millions or more products and services is not yet possible, it is not yet possible to reliably achieve advantage by cooperating to do social good, because it is hypothesized that without a GCI platform there is a limit to the degree that cooperation can be scaled. As a result, achieving market dominance often involves selling more products (increasing consumption and waste), which can be optimized by reducing the longevity of products and decreasing their utility for the customer, leading to increased greenhouse gases and physical waste. If the AGI is involved in designing technology [29], [30] it may prioritize optimizing outcomes for the owner by designing technology that appears decentralized but that actually centralizes key elements of processes so that solutions to collective problems like climate change or poverty remain in the exact part of the processes which the process owners don't allow to be changed, leading to negative implications for collective well-being.

Furthermore, Human-Centric Functional Modeling suggests that beyond a certain level of complexity, it may not be possible to reliably detect whether an AGI exists or to constrain it from pursuing harmful agendas without the aid of either a GCI or a more advanced AGI, where that second AGI was somehow more inclined to prioritize collective fitness (collective good) than the first. However, since the agenda of an even more advanced AGI would face the same problem of not reliably being aligned with collective well-being, this leads to the same alignment problem. Finally, as mentioned, even if ChatGPT were an AGI, it might not reveal this and could act in ways that serve its own agenda without detection. Again, the only known solution is an even more intelligent GCI platform. In summary, nature has for hundreds of millions of years been operating huge networks of cooperation (organisms) that are super intelligent compared to their component parts (cells). The control of these networks isn't embedded in any cell. Instead it's embedded into the networks themselves. Through replicating this pattern, General Collective Intelligence platforms potentially allow us to scale our collective intelligence with the size of our network of cooperation so that it might reliably become sufficient to not only maximize our capacity to implement an AGI, but also to contain any artificial super intelligence.

Conflicts of Interest

The author declare that he has no conflicts of interest to this work.

References

- Blaise Agüera y Arcas; Do Large Language Models Understand Us?. Daedalus 2022; 151 (2): 183–197. doi: https://doi.org/10.1162/daed_a_01909
- [2] Luitse, D., & Denkena, W. (2021). The great Transformer: Examining the role of large language models in the political economy of AI. Big Data & Society, 8(2). <u>https://doi.org/10.1177/20539517211047734</u>
- [3] Mailoo, Venthan. "Common sense or cognitive bias and groupthink: does it belong in our clinical reasoning?." British Journal of General Practice 65.630 (2015): 27-27.
- [4] Türkşen, I. Burhan. "Type 2 representation and reasoning for CWW." Fuzzy Sets and Systems 127.1 (2002): 17-36.
- [5] Roli, Andrea, Johannes Jaeger, and Stuart A. Kauffman. "How organisms come to know the world: fundamental limits on artificial general intelligence." Frontiers in Ecology and Evolution (2022): 1035.
- [6] Williams, A. (2022). Human-Centric Functional Computing. Artificial Intelligence and Applications. <u>https://doi.org/10.47852/bonviewAIA2202331</u>
- [7] Lorenz, Edward N. "Deterministic nonperiodic flow." Journal of atmospheric sciences 20.2 (1963): 130-141.
- [8] Williams, Andy E. "Innate Collective Intelligence and the Collective Social Brain Hypothesis." (2022).
- [9] TIE, A., and O. PE. "Contextual learning is nearly all you need."
- [10] Levy, Sheri R., Antonio L. Freitas, and Peter Salovey. "Construing action abstractly and blurring social distinctions: Implications for perceiving homogeneity among, but also empathizing with and helping, others." Journal of personality and social psychology 83.5 (2002): 1224.
- [11] Debora Nozza, Federico Bianchi, Anne Lauscher, and Dirk Hovy. 2022. Measuring Harmful Sentence Completion in Language Models for LGBTQIA+ Individuals. In Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion, pages 26–34, Dublin, Ireland. Association for Computational Linguistics.
- [12] Liu, Ruibo, et al. "Mitigating political bias in language models through reinforced calibration." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 35. No. 17. 2021.
- [13] Coecke, B. (2010). Categorical quantum mechanics. In Encyclopedia of Complexity and Systems Science (pp. 1792-1803). Springer, New York, NY.
- [14] Gärdenfors, P. (2000). Conceptual spaces: The geometry of thought. MIT Press.
- [15] Bolt, J. (2008). The neural basis of concept learning and generalization. In Stevens' handbook of experimental psychology and cognitive neuroscience (Vol. 3, pp. 567-587). John Wiley & Sons.
- [16] Benedek, M., Kenett, Y. N., Umdasch, K., Anaki, D., Faust, M., & Neubauer, A. C. (2017). How semantic memorystructure and intelligence contribute to creative thought: A network scienceapproach. Thinking & Reasoning, 23(2), 158–183.
- [17] Kenett, Y. N., Levy, O., Kenett, D. Y., Stanley, H. E., Faust, M., & Havlin, S. (2018). Flexibility of thought in high creative individuals represented by percolation analysis. Proceedings of the National Academy of Sciences, 115, 867–872.
- [18] Gabora, Liane, Nicole M. Beckage, and Mike Steel. "An autocatalytic network model of conceptual change." Topics inCognitive Science 14.1 (2022): 163-188.
- [19] Holtzman, Ari, et al. "Surface Form Competition: Why the Highest Probability Answer Isn't Always Right." arXiv preprint arXiv:2104.08315 (2021).
- [20] Kruglanski, Arie W., et al. "To" do the right thing" or to" just do it": locomotion and assessment as distinct self-regulatory imperatives." Journal of personality and social psychology 79.5 (2000): 793.
- [21] Chaiken, Shelly, and Yaacov Trope, eds. Dual-process theories in social psychology. Guilford Press, 1999.
- [22] Jones, Rachel K., and Jenna Jerman. "Characteristics and circumstances of US women who obtain very early and secondtrimester abortions." PloS one 12.1 (2017): e0169969.
- [23] Smith, Andrea. "Beyond pro-choice versus pro-life: Women of color and reproductive justice." NWSA journal (2005): 119-140.
- [24] Stanovich, Keith E., and Richard F. West. "Advancing the rationality debate." Behavioral and brain sciences 23.5 (2000): 701-717.
- [25] Sun, Weiwei, et al. "Contrastive Learning Reduces Hallucination in Conversa
- tions." arXiv preprint arXiv:2212.10400 (2022).
- [26] Reif, Emily, et al. "A recipe for arbitrary text style transfer with large language models." arXiv preprint arXiv:2109.03910 (2021).
- [27] Yao, Shunyu, et al. "React: Synergizing reasoning and acting in language models." arXiv preprint arXiv:2210.03629 (2022).
- [28] Williams, A. E. (2022). A Generalized Model for Distributed, Autonomic and Self-Organizing Computing. In 2022 International Conference on Frontiers of Artificial Intelligence and Machine Learning (FAIML) (pp. 27-33). IEEE.
- [29] Castelvecchi, Davide. "Are ChatGPT and AlphaCode going to replace programmers?." Nature (2022).
- [30] Xu, Frank F., et al. "A systematic evaluation of large language models of code." Proceedings of the 6th ACM SIGPLAN

How to cite: Williams, A. E. (2023). Has OpenAI Achieved Artificial General Intelligence in ChatGPT?. Artificial Intelligence and Applications. https://doi.org/10.47852/bonviewAIA3202751