CS 3790 Research Paper Milestone 3: Is Consciousness Computational?

Charles H. McCartney

College of Computing, Georgia Institute of Technology

CS 3790: Intro Cognitive Science

Professor McGreggor

November 18, 2022

Introduction

In the past several decades, academics and researchers have increasingly sought to identify the defining characteristics and nature of sophisticated human thought and intelligence. The last two decades of research on the concept of consciousness have been characterized by cooperation between experts of various disciplines including cognitive science, neuroscience, philosophy, linguistics, and computer science, alongside many others. Given the dramatic advancements in computer technology and neurological research methods, the task to explore the mechanisms of consciousness has resurfaced as an attainable objective for future investigation. This paper seeks to address the concept of consciousness, its definition, and its applicability to computational theory and the overarching CRUM model for cognition. The computational-representational understanding of mind is the foremost hypothesis in the field of cognitive science which proposes that the mind is made up of mental representations and procedures which act upon these structures to facilitate intelligent behavior. According to the theory, mental representations and procedures are analogous to data structures and algorithms, respectively (Thagard, 2005). These broad ideas encapsulate the focus of the forthcoming research, and this discussion aims to assess the extent to which the concept of consciousness can be explained, analyzed, and modeled through a computational perspective. In an analogous approach, it may be possible to seek the computational correlates of consciousness in the same way that scientists search for neural correlates of consciousness, including the properties of various representations and processes, as well as architectural specialization to simulate the behavior of functional regions of the brain in both isolated and integrated models on a larger scale. Despite the vast intricacy of the subject-matter, this paper reflects a sincere attempt to integrate broader elements of every part of the overarching literature to present a coherent yet digestible discussion of the current state of academic investigations of consciousness with respect to the complexity of the topic at hand. Before broaching this topic in a sufficient manner, it is necessary to examine the distinction between surface-level simulations of sensation versus genuine self-referential consciousness instantiated in virtual environments. Through an analysis of the existing research pertaining to theoretical approaches, hypothetical thought experiments, and prior models which have sought to simulate aspects of intelligent behavior, the feasibility of discovering and modeling computational correlates of consciousness may be assessed, alongside the trajectory of future attempts to expand upon artificial intelligence technology and its applications in the real world.

The Issue

Since the beginning of the twenty-first century, artificial intelligence has expanded beyond strictly theoretical research and has been steadily implemented in the commercial world. Companies such as Google, Apple, Microsoft, IBM, and others have adopted AI initiatives for the development of cellphones, social media platforms, and cybersecurity and have increased subsidies for related research at the foremost technical universities in the United States (Ng &

Leung, 2020). There have also been instances of autonomous agents performing individual real-world tasks reaching or even exceeding the competency of human actors. For example, Waymo's autonomous vehicles exceeded four million miles of functional driving on road networks in 2017. Additionally, deep learning networks performed as well as professional radiologists in analysis of computed tomography scans in 2018 (Ng & Leung, 2020). Despite this, the currently existing intelligent agents can be categorized as narrow AI which refers to intelligent artifacts performing singular or specialized tasks. It follows that genuine instantiated consciousness which adequately models the complexity and nuance of human cognition has not yet been achieved. Researchers define this level of intelligence as "Strong AI", or "Artificial General Intelligence", and they are characterized by abilities including higher-order reasoning, comprehension of thought, and awareness of both internal and external environments (Ng & Leung, 2020). While some might argue and entangle themselves in the dispute over the true definition and nature of consciousness, it is more important to engage in the pursuit of its computational correlates to arrive at partial solutions and produce a more fulfilled understanding of the issue at hand.

In order to advance theoretical models for consciousness, a consensus must be arrived upon for its definition. Even by the early twentieth century, perspectives and debates had already arisen on the concept of consciousness highlighted by Edmund Husserl's science of phenomenology, concentrating on the study of conscious existence and objects of direct experience. Notably, Husserl described consciousness as intentional insofar as it referenced a real-world object, and, although it may traverse intentional and unintentional phases, intentionality is the only property which provided meaning for consciousness in its totality (Husserl & Ralph, 1931). Much of the existing academic literature conflicts with other published works in that they either implement opposing definitions of consciousness, or they emphasize one feature over another as the central focus of the investigation (Reggia, 2013). Additionally, there is a vast semantic distinction between awareness of one's environment and veritable self-referential consciousness. Through an analysis of the body of literature published thus far, it is apparent that there is a wealth of existing models that adequately demonstrate granular features of conscious awareness, but researchers have struggled to find an explanation or method for a form of meaningful instantiated consciousness. Despite this, there are several efforts to produce models with an ability to distinguish between internal and external environments with self-referential awareness (Ng & Leung, 2020). There is also a significant difference between a mere simulation of sense versus a veritable experience of a sensation. This so-called experience can be encapsulated by the idea of phenomenal consciousness which constitutes a legitimate set of states in which an agent can flexibly produce appropriate behaviors and adapt to the introduction of novel information (Reggia, 2013). Recognizing this dilemma has been essential in understanding the constraints of current research in what has or has not yet been investigated with substantive results. To appropriately introduce the issue, it will be eminently necessary to describe the various perspectives that dominate existing models and theory in the field. These conclusions will serve as a fulcrum of the discussion in the argument's progression throughout the paper.

In order to sufficiently approach the investigation of the computational correlates of consciousness, there must exist a coherent framework of the neurobiological underpinnings and mechanisms that characterize intelligent human behavior. Cleeremans (2005) published a paper examining computational models for specific cognitive mechanisms and sought to contrast both conscious and unconscious modes of information processing to model elements of autonomous decision making and meta-representation. This source proposes that, in an analogous approach, it may be possible to seek the computational correlates of consciousness in the same way that scientists search for neural correlates of consciousness, including the properties of various representations and processes, as well as architectural specialization. This involves simulating the activation of regions in the brain locally responsible for specific processes (Cleeremans, 2005). It is important to note that consciousness is neither static nor unitary, given that many processes corresponding to the behavior of neural pathways involve simultaneous satisfaction of multiple constraints as well as the use of specialized and non-specialized systems. Since the nineties, efforts have accelerated to investigate the neural correlates of consciousness, providing valuable information and structure in the ongoing pursuit of its computational equivalents. In this case, the researchers assert that the ability for agents to demonstrate adaptable control in their behavior constitutes one of consciousness's most significant features (Cleeremans, 2005).

A neural correlate of consciousness is defined as "a minimal neural system N such that there is a mapping from states of N to states of consciousness, where a given state of N is sufficient, under conditions C, for the corresponding state of consciousness (Cleeremans, 2005, p. 2). Research pertaining to this principle focuses on the distinction between neural activities involving awareness as well as those activities which do not. With this approach, many neurological studies have been conducted to isolate regions in the brain involved in either explicit or implicit learning in patients subject to various sets of tasks (Cleeremans, 2005). The two features emphasized in Cleeremans's (2005) study are quality of representation, describing prerequisite properties for a representation's availability in consciousness and the cortical space in addition to the capacity for an agent to represent itself in the form of metarepresentation. The main qualitative properties emphasized in the quality of representations are strength, stability, and distinctiveness. The strength of a representation refers to the number of units recruited for processing, as well as the modulated intensity of their activation. This can be beneficial in distinguishing simultaneous processes and identifying where or when one procedure might dominate another. The stability of a representation refers to the temporal component of its extraction in the processing of information. Recording stability could be conducive to the search for information most relevant to be maintained in the context of goal states or objectives. Lastly, distinctiveness refers to the extent to which representations of similar phenomena might overlap (Cleeremans, 2005). Each of these properties is necessary to characterize unconscious representational thought. This search for the computational correlates of consciousness targets the functional subcategory of access consciousness, which focuses on the ability for an agent to recall and recount previous experiences. This occurs because these features are easier to isolate and thus have been prioritized in most of the existing computational models (Cleeremans, 2005). According to some theorists, an agent can only qualify for conscious thought if it exhibits an awareness of its current mental state. This requires constant interaction between a system of first-order representations as well as a system of meta-representations that uses the former as an input (Cleeremans, 2005).

Another study conducted on computational consciousness and metarepresentation similarly pointed out the shortcomings in the capacity of connectionist networks to sufficiently represent information. The researchers assert that, although such networks exhibit sensitivities to information reflected in their input-output environment, they demonstrate no ability to access or act on this data as knowledge itself, as it remains strictly associated with the causal pathways elicited by training (Cleeremans et al., 2007). To address this, the authors present a simulation in which two networks interact, where one performs a trained categorization task and the second operates on the product of the first network as an input. This models a granular form of meta-representation, as the two networks' internal states interact and produce layered dependencies. The overarching view of consciousness emphasized in this study is encapsulated by the statement that, "Consciousness, in this light, thus involves knowledge of the geography of one's own internal representations - a geography that is itself learned over time…" (Cleeremans et al., 2007, p.1-2). To clarify, these researchers suggest that consciousness constitutes the brain's own interpretation of itself, compiled through experiences with the external world, other intelligent actors, and, perhaps most presciently, itself. In this case, the minimized model comprising two networks is sufficient to model fundamental features of consciousness, since the second higher-order network acts completely independently of the links and dependencies formed by the first network over time when provided exposure to training data (Cleeremans et al., 2007). According to the Radical Plasticity Thesis, referring to the previously expressed perspective, conscious activity is not exclusively expressed by neural activity. Instead, consciousness is characterized by complex mechanisms operating upon unconscious states and representations that are then subject to further processing and reorganization which may be simulated in a computational environment (Cleeremans et al., 2007).

Alternatives

In contrast to the view of observing neural correlates of consciousness, some have asserted that such investigations are misguided, as conscious states cannot be observed independent of the environments in which they operate (Clereemans, 2005, p. 2). It is furthermore the case that it may be difficult or impossible to link a single neural correlate or sets of correlates along with certain conscious features. Consciousness cannot be pinpointed as a single identifiable feature but rather a set of dissociable aspects of information processing characterized by a continuum of dynamic and graded procedures. Additionally, it has been argued that the majority of existing studies pertaining to the topic fall under two categories: overarching models which seek to fully or partially simulate information processing, as well as specific models that isolate specific scenarios. Each of these types have been subject to criticism or outright disqualification, as overarching models are often incompatible with data sets implementing predictive tests and analyses, and specific models inherently reflect an extremely limited scope of the mechanisms they attempt to simulate (Cleeremans, 2005). Despite the shortcomings inherently paired with many existing models, lessons can be extracted from an aggregation of research to produce a broader understanding of computational principles that might be adopted to investigate the difference between conscious and unconscious mechanisms.

Artificial neural networks have also been subject to significant criticism from researchers. The act of merely simulating and maintaining stable representations is not equivalent to conscious thought. Although there are exigent learning algorithms that have acquired sensitivity to changing informational inputs, they experience absolutely no awareness of possessing this knowledge, and are only suitable to perform in the context of specific tasks (Cleeremans, 2005).

Models of Strong AI cannot be developed by simply combining several versions of narrow AI, but rather must be characterized by adaptability and awareness of not only the set of tasks at hand, but also the broader contextual features of the environment. The recent mass proliferation of sensory data will likely necessitate growing demand for AI systems to perform data management as well as processing across multiple systems and domains (Ng & Leung, 2020). In the present day, deep neural networks have demonstrated notable success in predictive tasks and analysis, but these outcomes rely on the availability and ease of access to enormous sets of raw data. Inevitably, there are specific contexts in which massive amounts of relevant training data could be challenging or impossible to obtain in a manner conducive to research and/or commercial outcomes. To overcome these obstacles, future models must be capable of domain adaptation to simulate humans' ability to outline relationships between distinguishable objects and identify underlying features that are likely to be shared across various tasks (Ng & Leung, 2020).

When confronting new situations, humans employ representations as well as analogies to recall previous experiences and synthesize information to produce the most relevant interpretations. Outside of isolated scientific experiments, real-world environments are saturated with confounding variables and dynamic behavioral trends that cannot be often predicted (Thagard, 2005). In the same way, any intelligent model with a framework resembling consciousness must be able to interpret the complexities of human behaviors, biases, and motivations. Although it is not necessary to instantiate an agent that expresses emotion in the same way as humans, this overarching awareness might promote optimal decision-making on the part of the machine in lieu of any potential emotional shortcomings or obstacles that could inhibit a human operator facing a critical situation (Ng & Leung, 2020).

To engage in a legitimate decision-making process, intelligent machines must be able to produce and consider a set of possible actions and hypotheses after the successful appraisal of a novel situation. Within human-beings, this can be illustrated by imaginative thought and the flexibility required to think beyond the data and about alternative outcomes which may or may not occur. In the context of a machine, this implies that the agent must be capable of not only interpreting an object based upon its current and former states, but also on the possibilities of future states relating to what the object could become if subjected to some manipulation (Ng & Leung, 2020). This subgenre of AI development has been classified by researchers as computational creativity. When existing intelligent systems produce an outcome and enact some change on the environment, they are subject to a recurring framework that dictates their version of a decision-making procedure. Whereas human cognition encompasses the ability for introspection to adjust future plans and outcomes, many AI systems are susceptible to becoming stuck in a repetitive cycle or loop that reflects little or no adaptability to changing circumstances. A system implementing Strong AI must be equipped with the ability of both self-reference and reflection to adjust its own variables and processes given the introduction of newly available information. In addition, such systems must exhibit an awareness of their own current states that contribute to a repository of working memory to be able to expand success rates beyond narrow sets of tasks and objectives (Ng & Leung, 2020).

In an expansive work developed by Searle (2002), the author provides a thorough description of consciousness as a unique process which remains entirely separate from other naturally occurring processes. Although it is a biological phenomenon, consciousness is distinct from other naturally occurring processes, as it can contain an element of subjectivity, meaning that each person has a unique experience and expression of consciousness linked to internal processes in the brain. Notably, the concepts of consciousness and self-consciousness are also significantly different from one another. For instance, in order to experience the sensation of guilt, an intelligent agent would require an extremely advanced capacity for self-referential consciousness, whereas detecting the presence of an external object in an immediate environment would constitute a much simpler process (Searle, 2002, p. 7). Specifically, Searle provides three characteristics that make consciousness different from all other biological processes: qualitativeness, subjectivity, and unity (Searle, 2002, p. 23). The feature of qualitativeness entails that, for each conscious experience, there is a unique qualitative, sensory, or representational character that describes it. The feature of subjectivity implies that states of conscious experience only exist when some agent is granted exposure. In order to constitute an experience, a set of stimuli must be processed by some sort of intelligent agent. Lastly, the concept of unity implies that all conscious experiences take place in a larger context, meaning that none can occur in a manner isolated from other stimuli or phenomena. This feature also highlights the utility of research via split-brain patients through which it appears that there may be multiple modes of consciousness occurring simultaneously and communicating with one another within the same brain. (Searle, 2002, p. 25). Unlike other researchers, Searle asserts that consciousness cannot be compared to any other biological processes due to its subjectivity and first-person ontology, yet he leaves room for the validity of an objective science of consciousness. Despite this, he entirely dismisses the historical perspectives of both materialism and dualism, arguing that neither has any application in the real world. Most importantly, he expresses severe opposition to two ideas: that consciousness is analogous to software in the brain, and that consciousness is merely a special form of information processing. The overarching viewpoint is that artificial consciousness can be constructed with non biological materials only if the brain's capacities are duplicated rather than merely simulated (Searle, 2002, p. 33).

Given that cognitive science is an eminently interdisciplinary field, it may prove beneficial to provide context from alternative approaches and explanations to produce a more complete interpretation. Furthermore, many existing models of the mind attempt to incorporate real neurobiological networks and functional units in their implementation. An article by Behrmann and Shomstein (2004) focuses primarily on the neurological pathways and components most directly responsible for mechanisms of selective attention. This is a process by which a large group of inputs are filtered to select a smaller subset of data to be subject to more in-depth processing in the posterior parietal cortex. In summary, selective attention can be guided in two primary forms: a reactive one which depends on important features of the stimulus sufficient to divert attention, or a top-down one which prioritizes an agent's goal states or objectives. Recent evidence has revealed that these two processes are guided by distinct loci within the parietal cortex (Behrmann et al., 2004). This information should be incorporated into the broader discussion so as to provide interdisciplinary perspectives as well as a tangible description with which to compare computational models for consciousness. Selective attention is also a fundamental feature of conscious awareness in human cognition. Given that existing models and theories largely attempt to emulate human modes for information processing, awareness, and expression, it is important to include at least a cursory discussion of these models' biological derivatives to contrast natural and artificial instances of intelligence. (Behrmann & Shomstein, 2004).

Despite the promising reflections from the discussion so far, can it be asserted that Strong AI adequately resembles true consciousness? While some seek to produce an artificial agent equipped with superhuman levels of intelligence and full autonomy, others have sought to implement an approach emphasizing collaboration between humans and machines that serves more to augment human performance and intelligence rather than operating as independent entities. (Ng & Leung, 2020). To proceed with this investigation, one must determine what consciousness is and why it exists in the form of its present manifestation in humans. One important concept referred to as the Mind-Body problem investigates the nature of consciousness on a fundamental level. There exist several views disputing what exactly makes up the conscious mind, including materialism, which asserts that the mind is composed of the same substance which makes up the body, as well as dualism, which assumes the presence of an entirely different, perhaps ethereal, type of material comprising the mind compared to the natural sources of the body (Thagard, 2005, p. 141). Interestingly, several studies from neuroscientists and philosophers have indicated that consciousness may not be physical and that cortical activity on its own is not sufficient to justify or explain conscious intelligence (Nagel, 2012).

Evidence

Even though it seems unlikely that genuine self-referential consciousness can be simulated artificially given the limits of current knowledge and technology, there are many existing models and theories that demonstrate promising results for the future of AI research. According to a paper published by Ng & Leung (2020), there exist forty-five cognitive architectures that demonstrate promising results for the future of building models and theoretical frameworks which directly mimic features of the human brain. Such findings corroborate the argument that there exist serious prospects for expanding the base of knowledge on consciousness as well as its computational applications. One type of model which has sought to mimic conscious thought using both first-order and meta-representations is categorized as the forward model. Most of these have focused on specific aspects of motor control, which involves an agent which must both extract information pertaining to the effects of its action on the exterior environment and determine which actions it might take to achieve a designated goal state. These tasks require the agent to develop a model of the system from a bottom-up approach and to synthesize this information to optimally control its own outputs (Cleeremans, 2005). Typically, these models are composed of two interlinked networks that interact with one another to produce the desired outcome. The first takes a goal and elements of the current state as inputs and yields a set of actions as the output. The other takes as inputs the actions from the first network and elements of the current state to output predictive information on how the system would be changed if subjected to these actions (Jordan & Rumelhart, 1992). Given sufficient training data, the second network provides an internal representation of the environment which is then acted upon by the entire system, constituting a form of meta-representation (Cleeremans, 2005). This synthesis between perceptual states and actions for which it is responsible demonstrates a legitimate capacity for self-referential decision-making processes.

As early as the 1960s, digital architectures referred to as blackboard systems were developed to publish information and provide accessible data for other modules in the network. Furthermore, a contemporary architecture called Pathnet has demonstrated success using an algorithmic approach to find out which of its many artificial networks is most suited for a particular task (Von Braun et al., 2021). This flexibility and capacity to generalize across tasks constitutes a fundamental pillar of any future implementations of conscious agents, given the importance of task-discrimination in intelligent behavior. According to a paper published by Bechtel (1995), although some argue that humans maintain direct subjective access to the contents of their states, it can be argued that this phenomenon may be an illusion created by humans' ability to describe conscious states using other mental activity involved primarily in language use. Furthermore, some qualitative characteristics of stimuli could be impossible to instantiate computationally, as humans can only produce awareness and representations of them from a first-person viewpoint (Bechtel, 1995). This paper seeks to provide a framework on connectionist systems that satisfy various aspects of conscious mental states. These include intentionality, awareness of the constitution of intentional states, and distinguishable qualitative aspects of such states. It also provides reference to several models that attempt to represent conscious processes. The findings from this paper may be advantageous to expand the scope of this exploration of consciousness and its computational feasibility with an added connectionist perspective.

Two of the foremost objectives of the research in computational models can be summarized as such: to garner an improved understanding of human consciousness and to develop methods to produce machines with genuine conscious awareness. An analysis published by Reggia (2013) further introduces several models of consciousness that have been developed in the two decades prior to the date of publishing, and it expounds upon five categories based on the primary characteristic of each set of models. Academic sources such as these have proven indispensable for this investigation, as they effectively summarize the existing body of research and literature, details the feasibility of artificial consciousness, and presents conclusions about the trajectory of research in this field. The authors of this source introduce five categories of models pertaining to computational representation of consciousness: global workspace, information integration, internal self-model, higher-level representations, and attention mechanisms (Reggia, 2013). All of these are named based on the fundamental feature(s) that are used as the basis of each model. Although these categorizations are by no means infallible, they sufficiently demonstrate what has or has not been accomplished thus far in the academic body of research and literature. This study also asserts that computational modeling has been legitimized as an effective means to investigate consciousness, and that existing models have successfully emulated biological and neurological mechanisms involved in information processing (Reggia, 2013). Despite this, none of these models have confirmed the possibility of simulating what they refer to as 'phenomenal machine consciousness', which indicates the importance of progressing research toward a more fulfilling and substantive conclusion (Reggia, 2013). The necessity of pursuing this research is motivated by the desire to understand the nuances of consciousness as well as the overarching goal to produce a model of instantiated consciousness in an artificial machine. The Turing Test would allow researchers to determine if an artifact could, in fact, possess artificial consciousness based on a predetermined set of behavioral criteria, yet these also must take into account the conflicting philosophical perspectives which seek to address its nature. For the sake of brevity, it will suffice to briefly describe each of the five categories. Firstly, the global workspace theory developed by Bernard Baars between 1988 and 2002, describes the human brain as an interspersed network of specialized processors that provide the means for functional sensations and processing of information. While there are localized regions with specific purposes, it posits that there exists an overarching system which broadcasts and distributes information across the brain, largely through the apparatus composing the cerebral cortex (Baars & Franklin, 2009). Following from this hypothesis, globally available information and collective processes characterize what many researchers define as consciousness. Lastly, there exists a threshold through which information passes in order to satisfy the true operational extent of conscious thought. As an example, a neural workspace model introduced by Dehaene and other scientists (1998) demonstrates conscious mechanisms by contrasting automatic or 'routine' processes with novel processing tasks which require intentional conscious effort. An early implementation of this model was used to emulate human performance on Stroop tasks, an exercise involving the identification of colors with varying hues of ink emphasizing the effortful distinction for the incongruity between the word and color itself. Individual nodes representing either specialized or workspace neurons are connected through both excitatory and inhibitory links in a manner where only one global representation is operated upon at any single point in time. Reinforcement learning is used to provide the adjustment of Hebbian weights, and this model was successful in resembling the patterns of human performance in identifying incongruent sets (Dehaene et. al., 1998). Additionally, information integration theory posits that consciousness constitutes the ability of an agent to integrate information into an overarching set as well as to meaningfully distinguish between various states. This includes a quantitative metric to measure a network's ability to integrate based on a calculation taking into account all partitions of the network's data. This introduces problems, as it becomes increasingly difficult to calculate this metric for any but the smallest simulation models (Reggia, 2013). Internal self-models emphasize the capacity of a machine to represent a first-person perspective. Activation patterns in the brain are associated with an adaptable representation of both internal and external properties in which an agent can distinguish between these two categories. Higher-level representations encompass the more granular notion that patterns of activation are associated with a representation corresponding to a specific concept. In neural networks, these are modulated by trained weights emulating synapses as well as the distribution between nodes. This approach indicates that conscious mental activity is encompassed by a significantly higher level of variation for representation where collections of neural activity serve to elicit complex thoughts and understanding (Reggia 2013). Lastly, attention mechanisms emphasize the importance of the ability for conscious agents to actively select the most relevant components of sensory inputs and to discriminate between them based on the unity of available information. These are typically controlled by both top-down and bottom-up processes operating simultaneously in many regions of the brain. Although attention is not equivalent to conscious experience, the two concepts are highly correlated and have benefited the overarching understanding of consciousness pertaining to computational models (Koch & Tsuchiya, 2006).

According to a study published by Seth (2010), current scientific consensus indicates the central role of the thalamocortical system as perhaps most relevant in the contemporary investigations of the neural correlates of consciousness. Furthermore, the author attempts to specify the necessary and sufficient processes needed to identify the ubiquitous and naturally occurring conscious states. Perhaps most importantly, the paper identifies several of the most fundamental challenges proposed by research on consciousness, including the distinction between evolutionary and developmental origins of consciousness, as well as the possibility of a conscious artifact existing (Seth, 2010). The last two decades of research on the concept of consciousness have been characterized by the intersection between multiple disciplines including cognitive science, neuroscience, philosophy, linguistics, computer science, and more. This process, however, has been accelerated by the realization that the biological and functional underpinnings underlying cognition and consciousness can be analyzed without an overarching and unimpeachable explanation for why consciousness exists in its present natural form. The author argues that current research should expand from exploring mere correlative factors toward the analysis of explanatory variables and specific neural procedures that explain various processes (Seth, 2010). Current theoretical approaches might also benefit from the exploration of edge cases targeting deficits or disorders in consciousness, including patients who have endured comas, seizures, or other experiences which resulted in minimized levels of consciousness. New methods of neuroimaging and psychiatric care have, for instance, revealed the presence of residual consciousness in some patients (Owen et al., 2009). Such methodologies may benefit future explorations of the evolutionary and developmental underpinnings which constitute conscious procedures.

Another more recent academic study published by researchers Blum & Blum (2021) explores the theoretical approach toward consciousness according to a computational perspective. Inspired by the Turing Machine as well as the Global Workspace Theory, this paper focuses on the theoretical computer science perspective. The authors formally define a “Conscious Turing Machine”, which they identify as a model of conscious artificial intelligence. Rather than attempting to model the full complexity of the human brain, this instead focuses on simplified use cases and their validity in the context of consciousness (Blum & Blum, 2021). This provides utility in this investigation in order to provide examples of recent models and attempts to computationally reflect cognition and consciousness. The authors of this paper seek to formalize Bernard Baars's Global Workspace Theory (GWT) and to define a Conscious Turing Machine as a simplified model satisfying the definition of consciousness. Following the example of the brief yet complete description of the Turing Machine, this approach avoids an overly complex set of relationships and provides a streamlined version of consciousness to be analyzed with scientific evidence. The invention of the fMRI in 1990 precipitated studies into the physical neural correlates of consciousness in the human brain. The authors of this paper also emphasize the distinction of whether their model can actually experience sensations rather than simulating them in a relatively surface-level manner. This model contains processors that emulate short-term and long-term memory elements, with multiple sources of LTM data competing to obtain access to the short-term memory channels. The authors also specifically define both deterministic and probabilistic variants of the CTM throughout the formal sections. Input maps provide specifications that translate data from the 'outer world' to the 'inner world', as defined by conceptual understanding of cognitive symbols (Blum & Blum, 2021, p. 19). These reflect natural human perceptual sensations such as gustation, audition, and other fundamental senses. Similarly, actuators in the model provide mechanisms for output to the 'outer world'. There exists a specific coded language that guides all internal processes and states within the model, which also permits the expression of genuine sensations. In this model, conscious awareness is achieved when information from the STM is fetched by all LTM processors at the same time. Other components comprising the Conscious Turing Machine's architecture are links, up- and down-trees, inputs, outputs, and chunks containing codified information that is propagated through the model after the completion of their production. The probabilistic variation of the CTM takes into account associated weights of chunks and 'coin-flip' neurons which complicate the process of competition between information with varying levels of importance (Blum & Blum, 2021, p. 21). It is not feasible to prove mathematically the feeling of conscious awareness without a concrete and demonstrable definition or set of constraints. Despite this, arguments can be made that the CTM has the feeling of consciousness, as well as pain and/or pleasure. The authors argue that this sensation is characterized by the interactions between the internal language of the system, as well as specific components in the architecture and the sequence of predictive dynamics in the model. For instance, the information contained by the short-term memory processor is propagated to all long-term memory processors, and other processors are guided by unique algorithms that enable special prerequisite functions for the feeling of consciousness. These include the 'Inner Speech', 'Model of the World', and 'Inner Sensation' processors, among others. For instance, the Model of the World processor provides for the CTM representations of both the inner and outer worlds, as well as a sense of self which is distinct from external objects (Blum & Blum, 2021, p. 25). Sensory input processors and actuators responsible for outputs are also included in the proper functioning of this component. It can also be argued that the recurring process through prediction, feedback, and procedural learning accompanied by parallel processing and predictions constitute a relatively advanced form of conscious awareness. Overall, this article provides a substantial framework through which to discuss the distinction between superficial simulations of consciousness and more complex, experiential forms that more accurately reflect real conscious intelligence. This is particularly useful because this model of the Conscious Turing Machine avoids unnecessary confounding factors that complicate the question considerably. While this example does not constitute a form of fully-instantiated machine consciousness, it provides a fundamental framework through which to assess and build upon new approaches toward a more complete model.

Although these studies constitute only a few examples of the vast body of prior research on the topic, they demonstrate compelling results which corroborate the view that elements of consciousness can be successfully implemented in future computational models. Furthermore, these results have indicated that the pursuit of computational correlates may produce tangible results that benefit not only the academic understanding of consciousness but also the application and expansion of AI technology to achieve real-world objectives. It is therefore apparent that the cooperation between man and machine will likely characterize the future of human discovery, commerce, and research on concepts yet to come.

Conclusion

In the context of modern computer science research, the emergence and progression of artificial intelligence and machine learning have constituted an unprecedented development in technological advancement as well as human understanding of the mind's composition and function. Computer learning facilitated by artificial intelligence has already drastically outpaced the ability of the human mind to consume and interpret vast quantities of data, and such technology has been increasingly incorporated into business and industrial pursuits. As cognitive science and its overlapping fields of neuroscience, linguistics, computer science, anthropology, and philosophy advance in sophistication and representational accuracy, its discoveries can be more broadly translated to real-world, tangible applications, such as artificial intelligence, deep learning methodology and even more experimental models that expand the abilities of computers and facilitate human's understanding of the mind. AI-based automation has increasingly been applied to the workplace as many experts predict a dramatic shift in the composition of the economy and labor force as a whole. Furthermore, computer learning has facilitated the analysis of vast quantities of data that would have otherwise never been processed successfully, if not efficiently. Studying the concept of consciousness is not only an endeavor to discover the foundational nature of humankind, but its pursuit may also produce new information that could overlap with broader fields such as medicine, computer science, psychology, education, and much more.

The question of whether consciousness is computational is extraordinarily complicated and may not be able to be sufficiently answered at present. Despite this, it is certainly possible to provide an adequate description of the current state of academic understanding on the topic and its shortcomings only further reinforce the necessity of pursuing rigorous research about the nature of consciousness, as well as its computational correlates. There are many who dispute these objectives and assert that any approach seeking to instantiate consciousness with strictly non biological material shall be doomed to fail (Searle, 2002, p. 33). Additionally, there are those who contend that a unified effort to study computational consciousness will be confounded by conflicting perspectives on its composition pertaining to the Mind-Body problem and other functional aspects. Nevertheless, Professor Joseph Mellichamp of the University of Alabama articulated a contrasting view by stating, "The 'artificial' in artificial intelligence is real" (Ng & Leung, 2020). The perspective that consciousness can never be truly manifested in the human brain serves only as a hindrance to the promising findings that have been demonstrated by decades of research on computational correlates and models which successfully emulate aspects of conscious thought and behavior. While some might ensnare themselves in the debate over the true definition and nature of consciousness, it is not only more important but also more productive toward all ends to engage in the pursuit of its computational correlates to arrive at partial solutions and produce a more fulfilled understanding of the issue at hand. Such discoveries elicited by continued investigation could serve to benefit the body of knowledge in both neuroscience and computational applications, but also will constitute the bedrock of future research in AI technology that may define the overarching trajectory of human progress and understanding in the decades to come.

References

BAARS, B. J., & FRANKLIN, S. (2009). CONSCIOUSNESS IS COMPUTATIONAL: THE LIDA MODEL OF GLOBAL WORKSPACE THEORY. *International Journal of Machine Consciousness*, *01*(01), 23–32. https://doi.org/10.1142/s1793843009000050

Bechtel, W. (1995). Consciousness: Perspectives from symbolic and connectionist AI. *Neuropsychologia*, *33*(9), 1075–1086. https://doi.org/10.1016/0028-3932(95)00049-9

Behrmann, M., Geng, J. J., & Shomstein, S. (2004). Parietal cortex and attention. *Current Opinion in Neurobiology, 14*(2)*,* 212-217. https://doi.org/10.1016/j.conb.2004.03.012

Blum, M., & Blum, L. (2021). A Theoretical Computer Science Perspective on Consciousness. *Journal of Artificial Intelligence and Consciousness*, *08*(01), 1–42. https://doi.org/10.1142/s2705078521500028

Cleeremans, A. (2005, January 1). *Computational correlates of consciousness* (S. Laureys, Ed.). ScienceDirect; Elsevier. https://www.sciencedirect.com/science/article/pii/S0079612305500074

Cleeremans, A., Timmermans, B., & Pasquali, A. (2007). Consciousness and metarepresentation: A computational sketch. *Neural Networks*, *20*(9), 1032–1039. https://doi.org/10.1016/j.neunet.2007.09.011

Dehaene, S., Kerszberg, M., Changeux, J.P. (1998). *A neuronal model of a global workspace in effortful cognitive tasks*. PNAS. https://www.pnas.org/doi/full/10.1073/pnas.95.24.14529

Husserl, E., & Ralph, W. (1931). *Ideas; general introduction to pure phenomenology, by Edmund Husserl.* Allen & Unwin ; New York.

Jordan, I. J., & Rumelhart D. E. (1992). *Forward models: Supervised learning with a distal teacher*. ScienceDirect. https://www.sciencedirect.com/science/article/pii/036402139290036T

Koch, C. & Tsuchiya, N. (2006). *Attention and consciousness: two distinct brain processes*. PubMed. https://pubmed.ncbi.nlm.nih.gov/17129748/

Nagel, T. (2012). *Mind & cosmos: why the materialist Neo-Darwinism conception of nature is almost certainly false*. Oxford University Press. https://www.researchgate.net/publication/275282802_Mind_and_Cosmos_Why_the_Materialist_Neo-Darwinian_Conception_of_Nature_Is_Almost_Certainly_False_by_Thomas_Nagel

Ng, G. W., & Leung, W. C. (2020). Strong Artificial Intelligence and Consciousness. *Journal of Artificial Intelligence and Consciousness*, *07*(01), 63–72. https://doi.org/10.1142/s2705078520300042

Owen, A. M., Schiff, N. D., & Laureys, S. (2009). *Coma science: clinical and ethical implications*. Progress in Brain Research. https://pubmed.ncbi.nlm.nih.gov/19818889/

Reggia, J. A. (2013). The rise of machine consciousness: Studying consciousness with computational models. *Neural Networks*, *44*, 112–131. https://doi.org/10.1016/j.neunet.2013.03.011

Searle, J. R. (2002). *Consciousness and Language*. (n.d.). Scilib-Biology.narod.ru. Retrieved September 17, 2022, from http://scilib-biology.narod.ru/Searle/ConsciousnessAndLanguage/cl.htm#03

Seth, A. K. (2010, March 10). *The grand challenge of consciousness* . Frontiers in Psychology. https://www.frontiersin.org/articles/10.3389/fpsyg.2010.00005/full

Thagard, P. (2005). *Mind: Introduction to Cognitive Science, , 2nd Edition* (2nd ed.). Bradford Books.

Von Braun, J., Archer, M., Reichberg, G., Sánchez, M., & Editors, S. (n.d.). Robotics, AI, and Humanity https://library.oapen.org/bitstream/handle/20.500.12657/47279/9783030541736.pdf?sequence=1#page=48