Are LLMs Capable of Achieving Artificial General Intelligence?

Anonymous ACL submission

Abstract

There is an ongoing debate regarding whether modern large language models (LLMs) are early, incomplete forms of Artificial General Intelligence (AGI). In this work we first present some prior art search across disciplines into the exact definition and features of an AGI, specifically the quintessential requirement, consciousness. From this literature survey we derive certain features that have been commonly mentioned as features AGI machines should possess. Based on these we show that LLMs do not satisfy any of these criteria, hence cannot achieve AGI. Further, we make a novel proposal: Quantum Natural Language Processing, a recently invented sub discipline of AI, has the potential to create AGI machines.

1 Introduction

From time immemorial humans have been in a quest to create machines that can aid them in specific tasks (intellectual or otherwise) Starting from abacus, humans have had a huge success in creating machines that serve as an augmentation to their intelligence. These Artificial Intelligence (AI) models built for very specific or constrained applications, is now formally called *weak AI*. There have been several standard techniques and benchmarks proposed for determining if a given system is a *weak AI* or not. Such benchmarks include mathematical formalism, engineering and biology inspired perspectives amongst others (Wang et al., 2020; Goertzel, 2014; Babcock et al., 2016). At the same time, there has also been a parallel quest, a quest on creating machines that are capable of being an exact replica of, if not surpass, human intelligence. This quest, now called *Strong AI*, is considered a huge step towards Artificial General Intelligence (AGI) (McLean et al., 2023; Butz, 2021; Ng and Leung, 2020). A quintessential factor mentioned as being definitely necessary in an AI which is trying to replicate human intelligence is to have an ability to be conscious (i.e., have a human like consciousness)[1]

Currently, there is an ongoing debate regarding whether modern large language models (LLMs) such as GPT-4 are early, incomplete forms of such a copy of human intelligence (Fei et al., 2022; Alvero and Peña, 2023; Susnjak et al., 2024; Pricea). The reason for this debate itself originates from the fact that while LLMs demonstrate impeccable accuracy in several standard engineering tasks of Artificial Intelligence, there is no definite or conclusive proof yet to show that these are machines with conscious artificial general intelligence (Chalmers, 2023). In fact the debate is so fierce that there are schools of thought supporting the rewriting of the definition of AGI itself (Heaven, 2023). Either way, we argue, that the mere fact that such a debate exists, is proof of the fact that humans have not reached a stage where we can definitely, without any doubt, say that current AI models are Conscious. Or in other word, consciousness still remains an unsolved mystery and a big hurdle in humanity's quest to AGI.

In this work we present a novel perspective to this debate. By reviewing definitions and expectations of consciousness from an inter-disciplinary perspective we arrive at certain characteristic features an AI model should have for it to be conscious. Based on these characteristics we further argue that LLMs don't posses most of these requirements and hence are not conscious. We further show that, the newly invented field of Quantum Natural Language Processing (QNLP), instead satisfies all such requirements and in-fact holds a huge potential for becoming the first step towards creation of machines or AI models that are conscious.

Specifically our contributions are as follows:

[1] While this in-turn entails several further features that the model should have, sooner or later, like being able to demonstrate emotional intelligence, imagination, effective command of other machines or robots, self-referring and self-reflecting qualities etc in this work we are concerned only about the consciousness part.

1. We conduct a literature review across several disciplines on consciousness and from them summarize certain features that researchers in these disciplines have suggested that a conscious AGI should have.

2. We show that neural network based models, including LLMs, don't satisfy any of these characteristics and hence they are not conscious.

3. We introduce the field of QNLP and show that it does have all these characteristics, and thus has the potential to be the vanguard methodology for the creation of conscious AGI machines.

2 The Problem

In this work we are trying to find the answer to the question: if a AGI machine or model was ever to be created, what features will it typically have? Note that this effort in AGI is being done from the perspective of Natural Language Processing (NLP) only, and hence makes a fundamental assumption connecting consciousness and language i.e., if an entity can comprehensively understand and respond to a natural language question through speech or text, we assume they are conscious (e.g., Humans).

3 Related Work

As a typical first step towards finding answer to the aforementioned problem, we investigate some prior literature related to AGI. Note that since AGI (especially consciousness) is an inter-disciplinary topic, these emanate from a range of disciplines including Philosophy, Physics, Computer Science, Medicine etc.

In philosophy there have been several thought experiments that were suggested as part of initial investigation into the creation of AGI machines. Such thought experiments include, The Mill by Gottfried Leibniz (Leibniz, 2009), The Game by (Dneprov, 1961), The Robot In Pain by (Dennett, 1978), Phone Calls To Emulate A Brain In Pain by (Block, 1980), A collection of conscious self neurons by (Cole, 1984) and the Chinese Room Argument that was presented in (Searle, 1980, 1999). All these arrive at a common conclusion that, intuitions of implementing systems should not be relied upon to be a faithful representation of the big picture i.e., a collection of individual deterministic automaton cannot produce a machine that is an AGI.

While up until the early 20th century, inquiries into consciousness mostly came from philosophers, spiritualists and some times from metaphysics researchers, in the later 20th century this started being explored in more science and engineering disciplines.

Perhaps, the most important investigation into consciousness and AGI from the field of Computer Science came from Turing's suggestion of a paper machine (Turing, 2004), which later came to be called the *Turing Machine* where with infinite tape the machine is capable of implementing any algorithm. As a corollary, he argues that the individual machines ('computers')[2] do not need to know the big picture or big problem, as long as it follows the exact instructions given to it to do its part. Half a century after Turing proposed his idea, Minsky (Minsky, 1980) suggested a virtual mind argument in reply to the aforementioned Chinese Room argument from Philosophy (Searle, 1980, 1999). He argued that, the agent that understands the the tasks could be distinct from the physical system that conducts the calculation itself. Also, in (Dreyfus, 1965, 1972) the author argues that key features of human mental life could not be captured by formal rules for manipulating symbols. He also identified several problematic assumptions in AI, including the view that brains are discrete computational mechanisms like *digital computers*, and thus reiterates the suggestion that understanding, and in turn consciousness, and hence by extension AGI, cannot be codified as explicit rules.

The first comprehensive work on definition of human consciousness from the perspective of Physics was proposed by Dr. Roger Penrose in his 1989 book *Emperor's New Mind* (Penrose, 1989) where he proposed that *quantum processes are fundamental in creating the physical basis of consciousness*. In this work he asks perhaps one of the most pertinent questions related to modern view of consciousness: "Is our picture of a world governed by the rules of classical and quantum theory, as these rules are presently understood, really adequate for the description of brains and minds?". He then

[2]The term Computers at that time, in 1940s, was used to denote the humans who worked for Alan Turing in cracking the Enigma code, where each computer had a very specific task to do, but none of them knew about the big picture or the big task at hand.

concludes that brain is making some sort of *non-computational physical processes* in some essential way when consciousness comes into play. Further, he argues that there are aspects of brains actions that cannot be faithfully simulated by the actions of a discrete system like classical computers.

Meanwhile there were inquiries being conducted into consciousness from the field of modern medicine also. One of the earliest such work was by In this work was from Hameroff (Hameroff, 1974) where he introduces Micro-Tubules (a deeper subsection of a Neuron) as the possible residing location of consciousness. This research was then combined with Penrose's suggestions (Penrose and Hameroff, 1995) to suggest that micro-tubules are *quantum systems*. These were then codified as Orchestrated Objective Reduction (Orch-OR) theory(Penrose and Hameroff, 1995) which arrives at the conclusion that consciousness originates biologically in *quantum computing structures* called micro-tubules which reside inside brain neurons. Specifically the authors suggest that the pre-conscious state of mind is passed through a quantum computer inside the micro-tubules which organizes consciousness hierarchically into a full fledged consciousness state. Extending this conclusion with the probabilistic wave interpretation of quantum physics (Griffiths and Schroeter, 2018; Cowan and Tumulka, 2016) we can arrive at the conclusion that consciousness is something that collapses from a wave-like form into the particle-form of a human body, when it passes through micro-tubules.

Meanwhile in linguistics, even though initially not directly related to consciousness, Noam Chomsky and other grammarians proposed that in human natural language, syntax and communication have a *hierarchical structure* (Chomsky, 1956; Lambek, 1958; Béchet and Foret, 2005) known as universal grammar underlying all languages and are innately hard wired in our brain when we are born (Chomsky, 2014). Or in other words, language is built *bottom up* where certain fundamental blocks like nouns, adjectives and other parts of speech are recursively combined up, so as to support the aforementioned hierarchy.

Early efforts to connect this with consciousness and to find recursive branching patterns in brain's neural networks have been unsuccessful and the idea has fallen out of favor. However the authors of Orchestrated OR theory of consciousness recently proposed that micro-tubules inside neurons have appropriate hierarchical resonance frequencies repeating at different scales, along with natural recursive branching patterns from which Chomsky's universal grammar's X-bar like patterns can be recreated in Quantum computers (Limnell and Vadén, 2021).

4 Summarizing The Prior Art In AGI and Consciousness

All the arguments from Philosophy and Computer Science can be summarized as: key features of human mental life, and in turn of AGI which is trying to replicate it, along with consciousness, could not be captured by formal rules for manipulating symbols i.e digital computers and the programs that run on them are not enough to produce an AGI state or consciousness. Note that all these bring up an interesting converse suggestion that there can exist machines that are capable of better representing key features of human mental life including understanding and consciousness, as long as they are *not discrete rule following systems like digital computers*. Combining that with the ideas from other disciplines that were presented subsequently we can see that certain pertinent characteristics of a conscious AGI machine can be arrived at like:

1. A machine capable of producing consciousness, cannot be a digital system. Also it should be able to support wave and particle forms i.e it must be a **quantum system**.

2. The machine must have a **collapse or reduction** mechanism by which waves can be collapsed into an instance of all the possibilities that were available, for example, a particle.

3. There must be a **hierarchy** to this reduction or collapse.

4. And hence the **mechanism** used to subsequently prove the existence of consciousness (for example the ability to understand and respond in a natural language) must also have a **hierarchical nature** to support this collapse.

4.1 Some Examples

Next let's consider some thought experiments to identify some possible real life scenarios in which the aforementioned characteristics might occur.

First, let's imagine the process of a patient waking up from anaesthesia after a surgery. Specifically, the pertinent time frame starts from the very

first 'moment of awakening' and ends when the patient has 'acquired enough consciousness so as to be able to do understand and speak a natural language to the doctors and nurses'. Like we mentioned before, in (Hameroff and Penrose, 2014, 1996) authors show that consciousness is stored in smaller components inside neural networks called micro-tubules, which they prove are *quantum systems*. Further they show that there is a hierarchical *collapse* from a wave form into the human body form as the patient gains more and more consciousness. As shown before, we also know from works of eminent linguists that most of the human natural languages are hierarchical. Considering the fact that after some time the patient is able to completely 're-acquire' his abilities of understanding and speak such a natural language is the proof that the *mechanism* used, i.e the human brain, is *capable of supporting a hierarchy, especially that of acquiring consciousness step-by-step, and also being able to understand and speak a human natural language* . Further, we can see that these characteristics, albeit in varying ratios, can be found in other such waking-up related scenarios such as a person waking up from sleep, person waking up after getting knocked out in a boxing match[3] etc., and in both cases subsequently re-acquiring their ability to understand and speak a human natural language [4].

4.2 Some Possible Candidates

Next let's look at some state of the art AI models so as to evaluate if they have any or all of the aforementioned 4 features expected from a conscious machine. The summary of all these, along with comparisons with the Orch-OR theory and human natural language, is captured in Table 1

4.2.1 Neural Networks Based NLP methods

Currently all the state of the art in NLP, including large language models (LLMs) are based on artificial neural network based models. Even though these artificial neural networks were modeled after the real neural networks in human brain, they are far from being conscious. To start with they are not quantum systems. Rather the fact that most of the current artificial neural network based models including LLMs are run on classical computers, we can safely conclude that they are not quantum systems, nor do they have any wave like properties that can be collapsed. Further, deeper inspection will reveal that despite the term large language models, which purportedly understands human natural language, it is neither hierarchical nor built in a bottom up manner. This is because all these models rely on bag-of-words which as a first step to the entire machine learning process, breaks down well formed sentences into smaller components - losing the bottom up nature of those entities, which were built up hierarchically and recursively[5]. Thus we show that current artificial neural network based AI models, including LLMs, do not satiate even a single expected feature of being a conscious AGI machine.

4.3 Quantum Natural Language Processing (QNLP)

Quantum Natural Language Processing (QNLP) (Clark et al., 2008; Coecke et al., 2010, 2020; Coecke, 2021; Coecke and Duncan, 2011) is a very recent invention, is founded on a mathematical proof that there are deep connections between human natural language and quantum physics[6] Further, it utilizes this idea in a standard deep learning framework over a quantum computer based AI model to learn the meaning of given sentences. In this section we will do a quick overview of QNLP, which is imperative for understanding the subsequent consciousness related argument.

QNLP is based on a few fundamental principles in linguistics which most of the human languages that exist currently follow. First, a human language typically starts with defining certain fundamental building blocks. Next, more complex structures are built up by combining these fundamental building blocks, in many cases recursively; for example, in the case of English, fundamental building blocks are typically parts of speech like nouns, verbs, etc. These can then be combined to create intermediate structures called noun phrases, prepositional phrases, etc. Next these structures themselves can

[3] Another example we wanted to include was death, but considering the fact that we don't wake up after death, at-least not in this body, we are leaving that for further discussion.

[4] Another interesting process we want to hypothesize that belongs to the same category is the, the process of a human growing up from birth till adulthood. Specifically, their ability to learn a natural language in the first few years. We want to argue that this event/process also satisfy the above criteria. We will leave the burden of proof of this as a thought experiment to the reader.

[5] While it an be argued that positional embeddings in Transformer based systems roughly simulate grammar and hierarchy, it still doesn't meet the bottom up construction requirement.

[6] Both can be modeled as categories in math, specifically pre-groups. Please refer the cited works for details.

	Orch-OR	Human Natural Language-Syntax + Semantics	Neural Network based AI models + classical computers	QNLP + quantum computers
Is a Quantum system	Yes	Maybe	No	Yes
Has Reduction	Yes	Yes	No	Yes
Has Hierarchy	Yes	Yes	No	Yes
Is Bottom Up	Yes	Yes	No	Yes

Table 1: Comparison of foundational characteristics of consciousness across various schools of thought In Natural Language Processing

be combined, in many cases recursively, to arrive at other complex structures like sentences, paragraphs, discourses, long drawn research documents etc. In other words, human natural language, is a structure built up from basic building blocks, following some rules, also known as grammar. In toto, QNLP is possibly one of the first machine learning systems to be able to combine syntax and semantics of English language as input features and successfully learn from them.

4.4 QNLP and Consciousness

In this section we consider QNLP as a possible candidate to create AI models that can create conscious entities. When we compare features of QNLP with the aforementioned 4 possible features of conscious models, we see that QNLP surprisingly, meets all these 4 criteria.

First, QNLP is a quantum system, since models created in QNLP use quantum computers to evolve the qbits, during the machine learning process. Since it is a quantum system, qbits live in the wave format, until a measurement occurs in the end, which collapses the wave into one of multiple possibilities (e.g. class label in a classification task). Since QNLP is very strictly based on lambek grammar, it is very hierarchichal, just like human natural language. Further in QNLP the meanings are built bottom up, i.e from words, which combine together to form the bigger meaning of a sentence, where entanglement is used as an information flow between the words to glue the information together, also known as grammar.

The authors want to argue that this exact match between QNLP and purported features of a conscious AGI system is possibly the very first steps and pointers towards creating conscious AGI machines.

5 Discussion

Since this paper is primarily meant to be initiating a discussion and possibly further research into this topic, we would like to ask these following questions also to add to it, based on everything presented so far.

What if early philosophers got off on the wrong foot[7] trying to understand consciousness using discrete tools in their thought experiments, including a digital computer, while consciousness itself is a continuous process. As an extension, since the philosophers have almost always been avant-garde troops in leading further scientific and engineering discoveries of humanity, what if humanity itself got off on the wrong foot trying to understand a continuous world using discrete tools [8] like transistor-based binary computers and neural networks when we should've been using quantum physics based tools? Despite its immense success in the history of human kind, what if this exact choice (of accepting faithful approximations), is creating the consequence i.e., acting as a glass ceiling preventing current AI to achieve AGI.

Or in other words, if LLMs are truly intelligent, and can eventually be made conscious, why does it need trillions of data points to understand a single piece of information (Bar-Hillel and Carnap, 1953; Shannon, 1948; Bar-Hillel, 1966). For example to understand the sentence *Mom Loves Dad*, a toddler, just needs one sentence, as long as they know the meanings of 'Mom', 'Love' and 'Dad'[9].

[7] Given the fact that at the time of the aforementioned philosophical thought experiments, quantum computers were yet to be conceived on paper, all their thought processes are perfect given the tools they had at their disposal. Hence the word wrong is not used in a demeaning way here, but instead as a gentle guidance very respectfully.

[8] The deeper reference here is not just to digital computers, but also to any process, like Leibniz's and Newton's calculus, which at the end of the day is only a faithful approximation of the continuous problem it is tackling, like the area under a curve, and not the expected continuous solution in itself.

[9] We acknowledge that this is admissible if the definition of AI is Augmentation or a helping tool to human Intelligence,

Further, we wonder the possibility of QNLP, in an effort to create faithful representations of syntax and semantics using quantum circuits and quantum entanglement, is unknowingly simulating micro-tubules, which the Orch-OR theory claims is where consciousness resides? At the risk of being derided by the scientific community for potential science fiction and spiritual influence, we would like to wrap up this section by asking the question: is it just an eerie coincidence or is it by design that the first tool in quantum computing to do natural language processing possesses all the traits necessary for consciousness, and is discovered very close to the predicted future event of singularity?

6 Limitations

Since this work is meant to be a vanguard towards creating discussions and potential future research work towards creating conscious AGI machines, it is at a purely theoretical hypothesis stage right now. A huge array of experiments are definitely needed to be completed before anything conclusive can be said about the ability of AI models created using QNLP to become conscious. Some of such investigations could include, creating tasks that can conclusively verify the presence of sentience in a machine. At the outset, even an investigation into the tasks that can be done by QNLP and not by artificial neural networks based models, will be a good start towards this goal.

7 Future Work

There have been proposed several recent discussions into what is the proof that neural networks based models do understand language. We leave the replication of these experiments from the QNLP perspective for the progeny including ourselves.

Further, we would like to pose these questions below, solutions to which we believe might enhance the quest towards creating a conscious machine. First question we would like to find the answer to in the future is, can we reconstruct micro-tubules (the layer inside neural network in which consciousness resides) using quantum bits and quantum circuits and see what emerges? Next, we would like to find if there is any relation between the two process of consciousness acquisition and language acquisition[10] or are they mutually exclusive processes? If they're related, does consciousness acquisition and language acquisition occur sequentially or in parallel? For example in the aforementioned examples, including that of the patient waking up from a surgery, does language acquisition process occur completely after whole consciousness is recovered, or does it happen in-parallel. Assuming it's sequential, the follow up question will be can we trace back to find the point in time (and all its properties) where consciousness hands over to language acquisition? If yes can we trace backward and recreate consciousness using QNLP, Qubits, and Quantum Circuits to replicate an artificially orchestrated reduction of micro-tubules, leading to the creation of machines that are conscious?

8 Risks and Ethical Impact

The ethical impact of developing artificial general intelligence (AGI) machines that are conscious raises profound questions, touching on moral philosophy, rights, and societal consequences. It begets some key ethical considerations. We discuss some of them below. If AGI machines attain consciousness, they could possess subjective experiences or a sense of self, similar to humans or animals. This would lead to a key ethical question: Should these AGIs be granted moral consideration? Would they have rights akin to human rights?If conscious AGIs can experience suffering or joy, denying them fundamental rights (such as the right to life or freedom from harm) could be viewed as morally wrong, similar to the ethical issues surrounding animals or marginalized human groups.Society may need to develop a new legal and moral framework for granting personhood or similar rights to AGI systems. Further, similar arguments could be raised about a bunch of topics that have been oft discussed, including exploitation and control of AGI, its suffering and well-being, existential risk and human safety, its potential for coexistence or conflict with humanity etc.

9 Broader Impact

If this hypothesis is proven to be true through future investigations, QNLP (and subsequently Quantum AI) will be a huge paradigm shift towards AGI especially since it combines syntax and semantics for the first time.

as opposed to be being a faithful copy of human intelligence, including consciousness

[10] in this work we are defining language acquisition as the ability to understand and speaking a language,

10 Conclusion

Creating machines that are conscious has been the holy grail quest of Artificial Intelligence, a quest known as Artificial General Intelligence. In this work we presented an inter-disciplinary literature review the field of consciousness and discover four features that a conscious AGI machine is expected to have. Further, we show that current artificial neural network based models used in NLP do not satiate any of the four required features and hence are not capable of being conscious. Instead we show that a very nascent sub-discipline of Artificial Intelligence, QNLP surprisingly meets all the 4 features expected to be present in conscious AGI machines. Note that this work is not at all meant to be a last word in this topic but on the other hand, is meant to be a vanguard towards creating discussions and potential future research work towards creating conscious AGI machines.

References

AJ Alvero and Courtney Peña. 2023. Ai sentience and socioculture. *Journal of Social Computing*, 4(3):205–220.

James Babcock, János Kramár, and Roman Yampolskiy. 2016. The agi containment problem. In *Artificial General Intelligence: 9th International Conference, AGI 2016, New York, NY, USA, July 16-19, 2016, Proceedings 9*, pages 53–63. Springer.

Yehoshua Bar-Hillel. 1966. Language and information; selected essays on their theory and application. *Foundations of Language*, 2(2).

Yehoshua Bar-Hillel and Rudolf Carnap. 1953. Semantic information. *The British Journal for the Philosophy of Science*, 4(14):147–157.

Denis Béchet and Annie Foret. 2005. k-valued non-associative lambek grammars (without product) form a strict hierarchy of languages. In *Logical Aspects of Computational Linguistics: 5th International Conference, LACL 2005, Bordeaux, France, April 28-30, 2005. Proceedings 5*, pages 1–17. Springer.

Ned Block. 1980. Troubles with functionalism. In *The language and thought series*, pages 268–306. Harvard University Press.

Martin V Butz. 2021. Towards strong ai. *KI-Künstliche Intelligenz*, 35(1):91–101.

David J Chalmers. 2023. Could a large language model be conscious? *arXiv preprint arXiv:2303.07103*.

Noam Chomsky. 1956. Three models for the description of language. *IRE Transactions on information theory*, 2(3):113–124.

Noam Chomsky. 2014. *Aspects of the Theory of Syntax*. 11. MIT press.

Stephen Clark, Bob Coecke, and Mehrnoosh Sadrzadeh. 2008. A compositional distributional model of meaning. In *Proceedings of the Second Quantum Interaction Symposium (QI-2008)*, pages 133–140. Oxford.

Bob Coecke. 2021. The mathematics of text structure. *Joachim Lambek: The Interplay of Mathematics, Logic, and Linguistics*, pages 181–217.

Bob Coecke, Giovanni de Felice, Konstantinos Meichanetzidis, and Alexis Toumi. 2020. Foundations for near-term quantum natural language processing. *arXiv preprint arXiv:2012.03755*.

Bob Coecke and Ross Duncan. 2011. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13(4):043016.

Bob Coecke, Mehrnoosh Sadrzadeh, and Stephen Clark. 2010. Mathematical foundations for a compositional distributional model of meaning. *arXiv preprint arXiv:1003.4394*.

David Cole. 1984. Thought and thought experiments. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 45(3):431–444.

Charles Wesley Cowan and Roderich Tumulka. 2016. Epistemology of wave function collapse in quantum physics. *The British Journal for the Philosophy of Science*.

Daniel C Dennett. 1978. Why you can't make a computer that feels pain. *Synthese*, 38(3):415–456.

A Dneprov. 1961. Igra -the game in russian. *Znanie sila Knowl. Power*, 5:39–41.

Hubert L Dreyfus. 1965. Alchemy and artificial intelligence.

Hubert L Dreyfus. 1972. What computers can't do: The limits of artificial intelligence.

Nanyi Fei, Zhiwu Lu, Yizhao Gao, Guoxing Yang, Yuqi Huo, Jingyuan Wen, Haoyu Lu, Ruihua Song, Xin Gao, Tao Xiang, et al. 2022. Towards artificial general intelligence via a multimodal foundation model. *Nature Communications*, 13(1):3094.

Ben Goertzel. 2014. Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence*, 5(1):1.

David J Griffiths and Darrell F Schroeter. 2018. *Introduction to quantum mechanics*. Cambridge university press.

Stuart Hameroff and Roger Penrose. 1996. Orchestrated reduction of quantum coherence in brain microtubules: A model for consciousness. *Mathematics and computers in simulation*, 40(3-4):453–480.

Stuart Hameroff and Roger Penrose. 2014. Consciousness in the universe: A review of the 'orch or'theory. *Physics of life reviews*, 11(1):39–78.

Stuart Roy Hameroff. 1974. Ch'i: A neural hologram? microtubules, bioholography, and acupuncture. *The American Journal of Chinese Medicine*, 2(02):163–170.

Will Douglas Heaven. 2023. Google deepmind wants to define what counts as artificial general intelligence.

Joachim Lambek. 1958. The mathematics of sentence structure. *The American Mathematical Monthly*, 65(3):154–170.

Gottfried Wilhelm Leibniz. 2009. *The philosophical works of Leibnitz.* Ripol Classic.

Petteri Limnell and Tero Vadén. 2021. Unfolding the big picture. essays in honour of paavo pylkkänen.

Scott McLean, Gemma JM Read, Jason Thompson, Chris Baber, Neville A Stanton, and Paul M Salmon. 2023. The risks associated with artificial general intelligence: A systematic review. *Journal of Experimental & Theoretical Artificial Intelligence*, 35(5):649–663.

Marvin Minsky. 1980. Decentralized minds. *Behavioral and Brain Sciences*, 3(3):439–440.

Gee Wah Ng and Wang Chi Leung. 2020. Strong artificial intelligence and consciousness. *Journal of Artificial Intelligence and Consciousness*, 7(01):63–72.

Roger Penrose. 1989. The emperor's new mind oxford university press. *New York*.

Roger Penrose and S Hameroff. 1995. Orchestrated objective reduction. *J Conscious*, 99(2).

Richard Pricea. If llms continue to improve, they will help solve the agi alignment problem.

John Searle. 1999. The chinese room.

John R Searle. 1980. Minds, brains, and programs. *Behavioral and brain sciences*, 3(3):417–424.

Claude Elwood Shannon. 1948. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423.

Teo Susnjak, Timothy R McIntosh, Andre LC Barczak, Napoleon H Reyes, Tong Liu, Paul Watters, and Malka N Halgamuge. 2024. Over the edge of chaos? excess complexity as a roadblock to artificial general intelligence. *arXiv preprint arXiv:2407.03652*.

Alan Turing. 2004. Intelligent machinery (1948). *B. Jack Copeland*, page 395.

Pei Wang, Patrick Hammer, Peter Isaev, and Xiang Li. 2020. The conceptual design of opennars 3.1. Technical report, Technical report, Temple University, Philadelphia, United States.