Exploring the Boundaries of Artificial Intelligence Using Cognitive Science: Final Report

Karan Bhasin
CS 6795: Intro to Cognitive Science
Georgia Institute of Technology
karanbhasin@gatech.edu

***Abstract*— Artificial Intelligence's role in society grows each day. It is vital that we better understand AI's capabilities to plan the role it plays in society. This literature review aims to look at AI through the lens of cognitive science, to understand its potential, and whether it can replicate the human mind in its emotional and creative capabilities. Four areas were studied to gauge the capabilities of AI: the current relationship between AI and cognitive science, cognitive science theories on concepts like empathy and creativity, current AI development related to these concepts, and ethical considerations as AI evolves. Ten research papers were carefully selected, reviewed and summarized to provide a holistic review of these four areas. It was found that AI continues to be inspired by cognitive science, however, not for less formal concepts like emotions and creativity. While cognitive science provides promising ideas for these concepts, there is further research required. There is active development and research into AI displaying capabilities in these domains, however, they are still limited. Stronger models and theories in cognitive science could lead to breakthroughs in AI. As AI evolves, ethical and moral considerations need to be examined and addressed for systems that are as, or more intelligent, than humans. Frameworks and standards need to be established for how we treat AI, and how AI treats us.**

Keywords—cognitive science, artificial intelligence, literature review, empathy, creativity, AI ethics

I. Introduction

The capabilities of Artificial Intelligence (AI) have evolved greatly in recent times. AI can be used to diagnose diseases [1]. It is used to drive cars [2]. It can also be used to make films that are showcased at film festivals [3]. AI has started to replace the human mind in accomplishing many tasks. This raises an important question however: Is there something AI cannot do? AI can diagnose diseases, but can it sympathize with a patient? It can drive cars, but can it make moral and ethical decisions when it encounters a dilemma? It can be prompted to make a film, but can it draw from experience to create personal and meaningful art? There is also growing anxiety on what role AI can potentially play in society and how it can replace humans [4].

To ensure that AI be used for good, and integrates into society harmoniously, it is important to understand its capabilities. Can AI truly replace humans in all areas where humans function? Is AI capable of uniquely human abilities like empathy, creativity, and self-awareness? What are the repercussions of the use of AI that is as intelligent as a human? This leads to the four research questions that are explored in this paper to understand AI's current capabilities and limitations through the lens of cognitive science.

What is the relationship between AI and cognitive science currently? – It is important to understand how AI is currently being inspired by cognitive science, and what human cognitive abilities it is able to replicate.

How does cognitive science explain empathy, creativity, and self-awareness in the human mind? – It is crucial to gauge if cognitive science is able to explain these processes in the human mind. A strong understanding of these processes in the human mind could inspire AI systems in the future.

Is it currently possible to create AI that can empathize and be creative? – Answering this question can help understand the current capabilities and limitations of AI in replacing the human mind.

If possible, what are the possible repercussions of making AI replicate the human mind so closely? – If AI is capable of being as emotional, creative and aware as a human, then what are the societal, ethical and moral consequences. If AI can replace the human mind, then how do we treat the AI, and how can we ensure there is no malice towards us?

II. Experiment Design

The **preliminary phase** consisted of formulating the four research questions posed in the introduction. These four questions were chosen after careful consideration so that a holistic view of AI's capabilities could be constructed. An initial batch of ten research papers were selected to explore the four questions. They were found on Google Scholar by using search phrases such as 'AI cognitive science relationship', 'cognitive science creativity', 'cognitive science empathy', 'cognitive science self-awareness', 'AI empathy', 'AI creativity', and 'AI ethics'. The initial batch was selected based on the paper's 'Abstract' and 'Conclusion'. A preliminary read through of the initial batch was conducted to confirm the depth and relevance expected for the literature review. To align with the scope and timeline, papers that provided a more holistic were preferred. Papers not considered relevant to the question or providing a broader picture of the question were removed from consideration. Any interesting papers referenced in the literature were also reviewed. Consequently, some of the papers in the final batch were referenced in papers from the initial batch. Any text from a paper that was eventually removed from the final batch but was still relevant to the questions was included in the results and appropriately cited. Additionally, to ensure the readings were relevant only papers published within the past 15

years were selected. This process lead to the selection of the following final batch of ten papers.

What is the relationship between AI and cognitive science currently?

- The Research of the Relationship Between Artificial Intelligence and Human Brain by Xin [6]
- A Standard Model of the Mind: Toward a Common Computational Framework Across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics by Laird, Lebiere, Rosenbloom [7]
- Neuroscience-Inspired Artificial Intelligence by Hassabis et al. [28]

How does cognitive science explain empathy, creativity, and self-awareness in the human mind?

- The Science of Empathy by Riess [29]
- Cognitive Bases of Human Creativity by Sweller [32]
- The cognitive neuroscience of self-awareness: Current framework, clinical implications, and future research directions by Mograbi et al. [33]

Is it currently possible to create AI that can empathize and be creative?

- Arguments for the Rise of Artificial Intelligence Art: Does AI Art Have Creativity, Motivation, Self-awareness and Emotion? by Liu [36]
- The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study by Chin et al. [49]

If possible, what are the possible repercussions of making AI replicate the human mind so closely?

- The Ethics of Artificial Intelligence by Bostrom and Yudkowsky [50]
- Are we preparing for a good AI society? A bibliometric review and research agenda by Wamba et al. [51]

The **review phase** consisted of reading the final batch carefully. Any text considered relevant was highlighted, and notes were taken to formulate the theme of the paper and the relevant points supporting that theme. The research questions were kept in mind to avoid deviating from the topic and scope of the project. Any new concept or term was researched to better incorporate the literature in the results. The notes and highlighted text were used to form a general summary of the paper that were used as a base for the results. Any assumptions and thoughts about each question were noted prior to the review, not only to look for similar themes but also to find literature that challenged my views and avoid bias. The original source is cited whenever the text used in the results referenced to another source in the paper. However, an in-depth review for all the original sources was not possible due to time constraints.

III. Results

The review yielded the following results for each research question:

1) What is the relationship between Artificial Intelligence and cognitive science currently?

It is integral to understand how the current knowledge of the human brain and the concept of mind is used in developing AI. AI is the use of computerized systems to simulate human intelligence to solve problems and complete tasks [5]. The brain's neural networks have inspired machine learning, deep learning and artificial neural network technologies [6]. AI aims to mimic human cognitive processes [6]. The human mind processes, organizes, interprets and reasons based on incoming information which leads to its understanding of things [6]. AI uses memory, algorithms, and models to perform similar cognitive operations [6]. These cognitive processes, which require perception, memory, thinking, and decision-making, are the main similarities between AI and cognitive science [6]. Some key differences are the higher processing speeds of AI compared to the human brain, and AI's superior capabilities performing tasks such as image recognition, and natural language processing [6]. Contrastingly, the brain processes sensory information such as vision, sound, and touch better than AI [6].

Hassabis et al. [7] provide examples of recent developments in AI inspired by cognitive neuroscience. The first example is the use of **attentional mechanisms** inspired by the brain's ability to focus on relevant information and dedicate processing resources to that information [7, 8, 9, 10]. These mechanisms have improved the effectiveness and efficiency of AI in tasks like object classification and recognition [11, 12]. The second example is the use of **episodic control and experience replay** inspired by the brain's reliance on multiple memory systems that use both gradual learning through repeated experiences and one-shot learning [7, 13, 14]. A deep Q-network with a replay buffer is similar to the complementary learning system in the hippocampus and neocortex in storing experiences and prioritizing highly rewarding events [7, 15, 16, 17, 18]. These developments have helped overcome slow learning in deep networks and allowed new concepts to be learned with fewer experiences [19, 20]. The third example is the development of long-short-term memory (LSTM) networks and differential neural computers (DNC) inspired by the brain's **working memory**, which relies on interactions between a central controller and separate domain specific memory buffers [7, 21, 22]. DNCs, which more closely model the brain's working memory by separating control and storage, can perform a variety of complex memory and reasoning tasks using this approach [7, 22]. The fourth example is how AI algorithms support **continual learning** by using elastic weight consolidation, which slows down learning in network weights crucial to solving previous tasks [7, 23]. This approach is inspired by lower rates of plasticity found in synapses that protect previous knowledge from interference [7, 24, 25]. [7, Fig. 1] shows the similarities between the AI systems and neural models across the aforementioned four examples: attention (A), episodic memory (B), working memory (C), and continual learning (D).

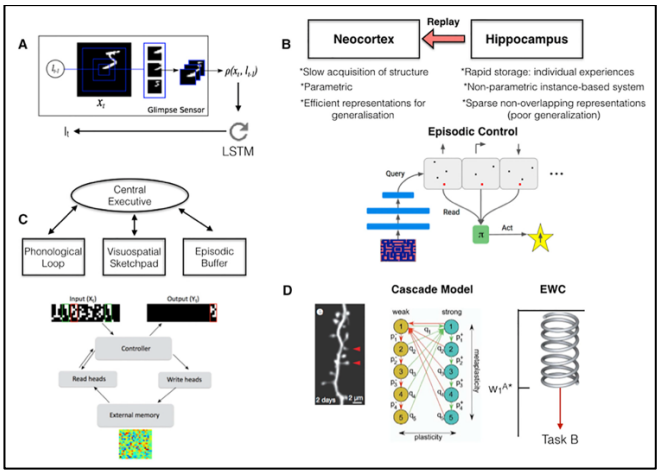

Fig. 1. Similarities between AI systems and neural models of behavior [7, 11, 19, 26, 22, 25, 27, 23]

Laird, Lebiere, and Rosenbloom [28] developed a standard model of the mind, seen in [28, Fig. 2], using the shared features of the three major cognitive architectures ACT-R, Soar and Sigma. The model intends to provide a basis for the development of AI models and systems that are closely similar to human cognition [28]. However, the authors note that their model doesn't account for metacognition, emotion, and mental imagery, which are important factors being explored in this paper [28].

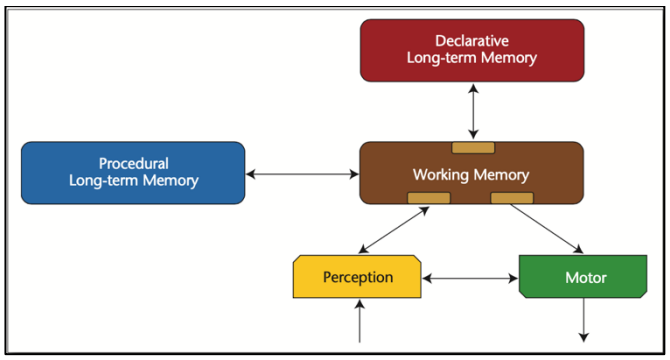

Fig. 2. Structure of the Standard Model [28]

2) How does cognitive science explain empathy, creativity, and self-awareness in the human mind?

Empathy is an important human ability that allows us to view things from somebody else's perspective. It fosters prosocial behavior by encouraging the sharing of experiences, needs, and desires amongst individuals [29]. Empathy was considered an innate trait, however, research has shown it is mutable and teachable [29]. Inner imitation of the others' actions has been supported by research [29]. Functional magnetic resonance imaging has revealed evidence of neural relay mechanisms that enable emphatic individuals to exhibit unconscious mimicry of others' postures, mannerisms and expressions to a greater degree than unemphatic individuals [30]. The same motor and sensory areas in the observer's brains are simulated as the person being observed [29]. Studies have shown an activation of a similar network of brain areas occurs in the brain of the observer when they are imitating or observing emotional facial expressions [29, 31]. Apart from inner representations of these facial displays, shared neural circuits have also been observed for other stimuli like tone, touch, pain and disgust [29]. It has been concluded that neural action representation modulates the observer's emotional content so they feel others' emotions to an attenuated degree. This allows us to empathize but not be overwhelmed by somebody else's distress [29].

Creativity is the unique ability of humans to conjure novel ideas and thoughts. However, there is a lack of advancement in understanding creative processes [32]. Sweller [32] posits that evolution by natural selection is a creative force that can be used to understand human creativity. The author proposes a cognitive architectural framework for creativity based on evolution which can be used to study and enhance human creativity [32]. More specifically, the randomness as genesis principle within this framework is postulated as the source of our creativity [32]. The randomness as genesis principle, inspired by mutations seen in evolutionary biology, is a random generation and test for effectiveness procedure used by humans when solving a problem [32]. This procedure must be backed by a knowledge base that helps reject generated moves that are not viable [32]. This principle's role in creativity is evidenced by the generation of novel ideas by a human when there is a lack of knowledge or an unsolvable problem [32]. Creativity can, thus, be enhanced by expanding an individual's knowledge base, which leads to the generation and testing of ideas that would not even be considered by someone with an inferior knowledge base [32]. Finally, creative solutions can be generated when brainstorming a large number of ideas without sufficient knowledge about the problem or reducing the specificity of goals [32].

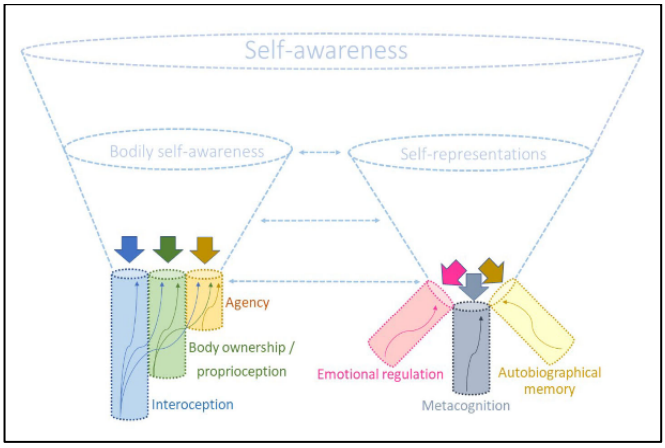

Fig. 3. Integrated model of self-awareness [33]

Self-awareness is the "ability to take oneself as the object of awareness" [34]. Mograbi et al. [35] state that self-awareness should be considered a collection of processes related to the body and cognition. Self-awareness abilities related to bodily processes include internal visceral states awareness (interoception), body ownership (proprioception), and action generation and control (agency) [33]. Abilities related to the representation level include cognitive abilities (metacognition), emotional abilities (emotional regulation) and recording of self-related information (autobiographical memory) [33]. There are potential relationships between these processes and levels that

combine to form the feeling of self-awareness [33]. [33, Fig. 3] depicts these possible relationships in an integrated self-awareness model. Predictive coding is a crucial framework for understanding self-awareness [33]. It proposes that compelled by survival instinct, the brain creates models in an attempt to predict the causes of sensory information [33]. The brain's cognitive goal is to minimize the mismatch between the incoming sensory information and the top-down predictions [33]. This interaction between sensory data and expectations leads to the emergence of self-awareness [33].

3) Is it possible to create Artificial Intelligence that can empathize and be creative?

Recent developments in AI have enabled it to generate art and literature. Jon McCormack [37] used artificial evolution and generative grammars to algorithmically grow digital plant forms for his Fifty Sisters project. One of the fifty images generated can be seen in [37, Fig. 4]. Similarly, ChatGPT-4 is able to create literary works like stories and plays, and even imitate the styles of famous authors [38, 39]. However, these AI systems either generate the content using prompts from humans, or by copying existing artistic styles and structures [36]. Hence, these AI generated works can be considered novel, something derived from something else, but not original [36]. Creative Adversarial Networks (CANs), unlike the imitative Generative Adversarial Networks (GANs), try to solve this problem by developing creativity through deviation from the learned style to enhance the appeal of the generated art [40]. Creative AI systems can be considered to be as technically adept as humans, and are not restricted by human emotions and motivations [36, 41]. However, these systems still have their limitations [36]. AI art lacks the capacity for large-scale narrative in literature and art, and the art is still generated using selected sample data [36, 42]. AI art was developed based on the idea that creativity lies in understanding and constantly discovering connections and relationships between things [36]. The creativity in AI systems is based on processes and algorithms that rely on the logic that the creation of art is decomposable, processable, relational, representable, and datable [36]. However, to develop more sophisticated AI technology, it is crucial to understand the working mechanisms of the brain with regards to creativity, awareness and emotion, so they can be reverse engineered [36].

Imagination is possible because humans are able to construct fictitious scenarios in their mind by reconfiguring existing and familiar elements in new ways [43, 44, 45]. Efficient representations that can be generalized and transferred make it possible for plans made in one setting be used by humans in new, but similar environments [7]. Additionally, humans plan hierarchically, by considering terminal solutions, interim choice points, and gradual steps toward the goal simultaneously [46, 47, 48]. These flexible, combinatorial aspects of planning can form the basis to develop an agent that can plan hierarchically, is creative and can solve problems currently unsolvable by humans [7].

Chin et al. [49] studied the potential of AI chatbots providing emotional and mental health support to users. The authors analyzed user conversations discussing mental health issues with a widely used AI chatbot not trained for mental health advice [49]. They found that a number of users felt safe sharing personal information as they sought a comfortable environment with an active listener [49]. They also noticed differences in user behavior based on culture [49]. Based on their findings, they encouraged the development of AI chatbots the provide cheap and readily available mental and emotional support [49]. However, they highlighted that factors such as age groups, cultures, and languages should be taken into consideration to provide a more inclusive environment for users [49].

Fig. 4. One of the Fifty Sisters that were algorithmically generated [37]

4) If possible, what are the possible repercussions of making AI replicate the human mind so closely?

Society needs to be proactive and prepared as AI technology continues to expand and its capabilities increase. All the criteria that apply to humans performing social functions such as responsibility, transparency, auditability, incorruptibility, and predictability should be applied to the AI replacing humans in performing the same functions [50]. Most AI technology today is specialized, developed to perform a specific task, however, we also need to consider Artificial General Intelligence (AGI) [50]. AGI is more generally applicable, it can operate across many new contexts, some of which cannot be predicted or even imagined by its developers or users [50]. To account for this, developers building AGI should specify good behavior as any action be taken as long as the consequence of that action is not harmful to humans [50]. The AGI should be required to think like an ethical engineer rather than be seen as a product of ethical engineering [50]. Current AI systems lack moral status, they can be edited, copied, terminated or deleted based on human requirements [50]. However, in the future AI systems could be granted moral status if they are sentient and sapient [50]. Sentience is the capacity of an entity to experience phenomena such as pain and suffering, and sapience is the capacities associated with higher order thinking and intelligence such as self-awareness [50]. Bostrom and Yudkowsky [50] propose two principles if AI systems have sentience and sapience like human beings. The first is the principle of substrate non-discrimination which states that if two entities have the same functionality and conscious experience but differ in the substrate of their implementation, then both entities share moral status [50]. The second is the principle of ontogeny non-discrimination which states that if two entities have the same functionality and conscious experience but differ only in how they came to exist, then they share moral status [50]. These principles can guide us in treating sentient and sapient AI systems the same way humans

are treated in familiar contexts [50]. For superintelligent systems that are capable of being smarter, stronger, and more trusted than humans, it is imperative that they be developed with the capability of forming superior ethics, values and niceness [50]. Much of our ethics today is based on our current understanding of the world, but there are contexts and problems we haven't yet encountered or imagined that may require us to reconsider and adapt our morals and values [50].

Wamba et al. [51] conducted a bibliometric analysis of AI research to determine whether the field is progressing towards building a society that uses AI for good. Though most of the research was technical, they discovered ten areas for social good that represented 2.6% of all AI research [51]. These areas were crisis response, economic empowerment, educational challenges, environmental challenges, equality and inclusion, health and hunger, information verification and validation, infrastructure management, public and social sector management, and security and justice [51]. The authors highlight that these areas constitute a small minority of all AI research and a number of research questions regarding the impacts of AI on society remain unanswered [51].

IV. Discussion

The results highlight the deep relationship between cognitive science and artificial intelligence. To substitute humans in accomplishing tasks and solving problems, AI research has to get a deeper understanding of the human brain and mind. Currently, with a better understanding of the brain through cognitive neuroscience, AI systems are evolving to perform more complex tasks more efficiently and effectively [7]. However, these systems are being developed to solve a specific problem or accomplish specific tasks. There does seem to be a lack of development in generalized intelligence. There also seems to be a limited use of cognitive science in AI research to develop empathic, creative or self-aware systems. This issue may be exacerbated by the exclusion of emotions and mental imagery in cognitive architectures designed to better understand the mind [28]. AI continues to be inspired by cognitive science, and this relationship needs to remain strong for continued development and understanding of the human mind and building systems that can replace humans.

Cognitive science research in empathy, creativity, and self-awareness shows that there are promising theories and ideas about how the human mind represents, connects, and processes these concepts. The literature provides a strong foundation and a general direction for future research. However, further research is required to confirm these theories and expand on these underlying ideas. A deeper understanding of how the mind and brain empathize, create and become aware can inspire the development of AI with these capabilities.

There seem to be a limited number of AI systems capable of being creative or empathic. AI systems are capable of creating original works of art and literature. However, these appear to be influenced by the sample data on which it was trained or by the input of the user's vision and ideas through prompts [36]. While AI art may not seem truly creative or original, the development of systems like CANs is striving towards that goal [40]. Additionally, existing cognitive science theories about concepts like imagination that can be adopted by AI research to develop creative agents [7]. There is also a strong case for building AI systems that can provide empathic support to users seeking mental and emotional support [49]. However, careful consideration needs to be taken, including accounting for linguistic and cultural differences, to effectively and safely provide optimal support in a sensitive environment [49]. The studies and research indicate efforts are being made to encourage and develop creative and emphatic AI. To drive this development, further research is needed and cognitive science needs to expand on how emotions and creativity are processed in the mind. A better understanding of these concepts in cognitive science can lead to more breakthroughs in this type of AI.

As AI systems evolve and become more complex and intelligent, we need to proactively build ethical and moral frameworks and standards to guide their development and integration into society. There are different considerations for AI systems with different levels of intelligence, and guidelines and standards need to be set on how we treat AI, and how AI treats us [50]. Certain considerations that we may not have right now due to the view of AI being another piece of software, may not be true in the future [50]. It is imperative that any future AI development strongly consider the ethical and moral repercussions of the technology. Additionally, such a powerful tool should be applied in areas where humans are vulnerable and need assistance [51]. Researchers should ensure that AI is being used to tackle major social issues faced by society.

V. Conclusion

This literature review explores the current status and possibility of AI replacing humans in social functions where creativity, empathy, and self-awareness maybe required. The results highlight how AI already mimics many cognitive processes of the human mind [6]. AI research is currently inspired by cognitive neuroscience to develop systems that are more efficient and effective in solving problems and completing tasks [7]. There is, however, a lack of development in making a generalized AI, or a system capable of being emphatic or creative. Additionally, efforts to standardize cognitive architectures do not account for emotions or mental imagery [28]. Cognitive science presents some promising ideas about how the human mind is emphatic, creative and aware, but further research is needed to expand on these theories [29, 32, 33]. There is research and development taking place to create emphatic and creative AI systems, however, these systems are still limited in their capabilities [7, 36, 49]. A broader understanding of how the human mind and brain process these concepts can lead to advancements in AI in these fields. As AI systems evolve, the ethical and moral repercussions of intelligent systems functioning in society need to be proactively discussed. There are different considerations based on the intelligence level of the AI. Additionally, AI needs to be used for the good of the world, helping the most vulnerable areas of society [51].

AI can become a powerful tool that permeates through human society and lives. However, there is also a long way to go. A deeper understanding of how the mind creates, emotes and is self-aware is needed for AI to truly start replacing humans for a myriad of social functions. The results show AI is still inspired

by cognitive science, and that an advancement in either field can benefit the other. Cognitive science is exploring how the mind empathizes, creates, and is self-aware, and further research is necessary to gain a more concrete understanding of the human mind. This better understanding can potentially lead to the development of AI systems that can replicate these concepts. Ethics and morals need to be taken into consideration as AI becomes more intelligent and performs more social functions.

VI. Limitations

This literature review just scratches the surface on understanding the current status and limitations of AI. Each question could qualify for an individual literature review. The scope and depth of the review were limited due to constraints in time and expertise. When reading the literature, there were numerous referenced papers that could have been explored to expand on each question. However, this was not possible due to time constraints. The lack of expertise in the area required the author to spend additional time comprehending each paper and understanding new concepts. With a better understanding, the review could have had greater depth. Additional questions that were not included in this review but could have been studied to better understand the topic include '*How cognitive science is inspired by AI?*', '*How does AI differ from cognitive science?*' and '*How is AI replacing humans in various fields today?*'. Other emotions, like happiness, sadness, and stress could have been explored. The paper concentrates on AI's cognitive capabilities, but physical capabilities of AI robots in performing human tasks can also be examined. The review still provides a sufficiently holistic view and points towards research that could help expand on each question and lays a good foundation for further research.

References

[1] S. Kaur *et al*., "Medical Diagnostic Systems Using Artificial Intelligence (AI) Algorithms: Principles and Perspectives," in *IEEE Access*, vol. 8, pp. 228049-228069, 2020, doi: 10.1109/ACCESS.2020.3042273.

[2] G. Biggi and J. Stilgoe, "Artificial Intelligence in Self-Driving Cars Research and Innovation: A Scientometric and Bibliometric Analysis," *papers.ssrn.com*, Apr. 19, 2021. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3829897

[3] L. Manfredi, "OpenAI's Sora to Debut 5 Original Shorts Made Using AI at Tribeca Film Festival," *TheWrap*, May 31, 2024. https://www.thewrap.com/openai-sora-tribeca-film-festival-short-films-debut/ (accessed Jun. 1, 2024).

[4] J. Cox, "AI anxiety: The workers who fear losing their jobs to artificial intelligence," *BBC*, Apr. 18, 2023. https://www.bbc.com/worklife/article/20230418-ai-anxiety-artificial-intelligence-replace-jobs

[5] M. Naveed Uddin, "Cognitive science and artificial intelligence: simulating the human mind and its complexity," *Cognitive Computation and Systems*, vol. 1, no. 4, pp. 113–116, Dec. 2019, doi: https://doi.org/10.1049/ccs.2019.0022.

[6] Q. Xin, "The Research of the Relationship Between Artificial Intelligence and Human Brain," *Frontiers in artificial intelligence and applications*, Nov. 2023, doi: https://doi.org/10.3233/faia230800.

[7] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-Inspired Artificial Intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, Jul. 2017, doi: https://doi.org/10.1016/j.neuron.2017.06.011.

[8] C. Koch and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985, Available: https://pubmed.ncbi.nlm.nih.gov/3836989/

[9] T. Moore and M. Zirnsak, "Neural Mechanisms of Selective Visual Attention," *Annual Review of Psychology*, vol. 68, no. 1, pp. 47–72, Jan. 2017, doi: https://doi.org/10.1146/annurev-psych-122414-033400.

[10] M. I. Posner and S. E. Petersen, "The Attention System of the Human Brain," *Annual Review of Neuroscience*, vol. 13, no. 1, pp. 25–42, Mar. 1990, doi: https://doi.org/10.1146/annurev.ne.13.030190.000325.

[11] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent Models of Visual Attention," *Advances in Neural Information Processing Systems*, vol. 3, 2014, doi: https://doi.org/10.48550/arXiv.1406.6247.

[12] J. Ba, V. Mnih, and K. Kavukcuoglu, "Multiple Object Recognition with Visual Attention," *arXiv.org*, Apr. 23, 2015. https://arxiv.org/abs/1412.7755

[13] E. Tulving, "How many memory systems are there?," *American Psychologist*, vol. 40, no. 4, pp. 385–398, 1985, doi: https://doi.org/10.1037/0003-066x.40.4.385.

[14] C. R. Gallistel and Adam Philip King, *Memory and the Computational Brain*. Wiley-Blackwell, 2009. doi: https://doi.org/10.1002/9781444310498.

[15] D. Kumaran, D. Hassabis, and J. L. McClelland, "What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated," *Trends in Cognitive Sciences*, vol. 20, no. 7, pp. 512–534, Jul. 2016, doi: https://doi.org/10.1016/j.tics.2016.05.004.

[16] J. L. McClelland, B. L. McNaughton, and R. C. O'Reilly, "Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory.," *Psychological Review*, vol. 102, no. 3, pp. 419–457, Jul. 1995, doi: https://doi.org/10.1037/0033-295x.102.3.419.

[17] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized Experience Replay," *arXiv.org*, Feb. 25, 2016. http://arxiv.org/abs/1511.05952

[18] A. C. Singer and L. M. Frank, "Rewarded Outcomes Enhance Reactivation of Experience in the Hippocampus," *Neuron*, vol. 64, no. 6, pp. 910–921, Dec. 2009, doi: https://doi.org/10.1016/j.neuron.2009.11.016.

[19] C. Blundell *et al.*, "Model-Free Episodic Control," *arXiv (Cornell University)*, Jan. 2016, doi: https://doi.org/10.48550/arxiv.1606.04460.

[20] Oriol Vinyals, C. Blundell, T. P. Lillicrap, Koray Kavukcuoglu, and Daan Wierstra, "Matching Networks for One Shot Learning," *arXiv (Cornell University)*, Jun. 2016, doi: https://doi.org/10.48550/arxiv.1606.04080.

[21] A. Baddeley, "Working Memory: Theories, Models, and Controversies," *Annual Review of Psychology*, vol. 63, no. 1, pp. 1–29, Jan. 2012, doi: https://doi.org/10.1146/annurev-psych-120710-100422.

[22] A. Graves *et al.*, "Hybrid computing using a neural network with dynamic external memory," *Nature*, vol. 538, no. 7626, pp. 471–476, Oct. 2016, doi: https://doi.org/10.1038/nature20101.

[23] J. Kirkpatrick *et al.*, "Overcoming catastrophic forgetting in neural networks," *Proceedings of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, Mar. 2017, doi: https://doi.org/10.1073/pnas.1611835114.

[24] J. Cichon and W.-B. Gan, "Branch-specific dendritic Ca2+ spikes cause persistent synaptic plasticity," *Nature*, vol. 520, no. 7546, pp. 180–185, Mar. 2015, doi: https://doi.org/10.1038/nature14251.

[25] G. Yang, F. Pan, and W.-B. Gan, "Stably maintained dendritic spines are associated with lifelong memories," *Nature*, vol. 462, no. 7275, pp. 920–924, Nov. 2009, doi: https://doi.org/10.1038/nature08577.

[26] V. Mnih *et al.*, "Human-level Control through Deep Reinforcement Learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015, doi: https://doi.org/10.1038/nature14236.

[27] S. Fusi, P. J. Drew, and L. F. Abbott, "Cascade Models of Synaptically Stored Memories," *Neuron*, vol. 45, no. 4, pp. 599–611, Feb. 2005, doi: https://doi.org/10.1016/j.neuron.2005.02.001.

[28] J. E. Laird, C. Lebiere, and P. S. Rosenbloom, "A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics," *AI Magazine*, vol. 38, no. 4, pp. 13–26, Dec. 2017, doi: https://doi.org/10.1609/aimag.v38i4.2744.

[29] H. Riess, "The Science of Empathy," *Journal of Patient Experience*, vol. 4, no. 2, May 2017, doi: https://doi.org/10.1177/2374373517699267.

[30] L. Carr, M. Iacoboni, M.-C. . Dubeau, J. C. Mazziotta, and G. L. Lenzi, "Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas," *Proceedings of the National Academy of Sciences*, vol. 100, no. 9, pp. 5497–5502, Apr. 2003, doi: https://doi.org/10.1073/pnas.0935845100.

[31] I. Morrison, M. V. Peelen, and P. E. Downing, "The Sight of Others' Pain Modulates Motor Processing in Human Cingulate Cortex," *Cerebral Cortex*, vol. 17, no. 9, pp. 2214–2222, Nov. 2006, doi: https://doi.org/10.1093/cercor/bhl129.

[32] J. Sweller, "Cognitive Bases of Human Creativity," *Educational Psychology Review*, vol. 21, no. 1, pp. 11–19, Nov. 2008, doi: https://doi.org/10.1007/s10648-008-9091-6.

[33] D. C. Mograbi, S. Hall, B. Arantes, and J. Huntley, "The cognitive neuroscience of self-awareness: Current framework, clinical implications, and future research directions," *WIREs Cognitive Science*, Dec. 2023, doi: https://doi.org/10.1002/wcs.1670.

[34] A. Morin, "Self-Awareness Part 1: Definition, Measures, Effects, Functions, and Antecedents," *Social and Personality Psychology Compass*, vol. 5, no. 10, pp. 807–823, Oct. 2011, doi: https://doi.org/10.1111/j.1751-9004.2011.00387.x.

[35] D. C. Mograbi, J. Huntley, and H. Critchley, "Self-awareness in Dementia: a Taxonomy of Processes, Overview of Findings, and Integrative Framework," *Current Neurology and Neuroscience Reports*, vol. 21, no. 12, Nov. 2021, doi: https://doi.org/10.1007/s11910-021-01155-6.

[36] B. Liu, "Arguments for the Rise of Artificial Intelligence Art: Does AI Art Have Creativity, Motivation, Self-awareness and Emotion?," *Arte, Individuo Y Sociedad*, vol. 35, no. 3, pp. 1–11, Apr. 2023, doi: https://doi.org/10.5209/aris.83808.

[37] J. G. McCormack, "Working with Generative Systems: An Artistic Perspective," *Electronic Workshops in Computing*, Jul. 2017, doi: https://doi.org/10.14236/ewic/eva2017.47.

[38] A. Hornaday, "Perspective | AI is a time-honored movie trope. Now it can write its own script.," *Washington Post*, Feb. 17, 2023. Available: https://www.washingtonpost.com/movies/2023/02/17/ai-in-movies-screenwriting/

[39] S. Heritage, "Can an AI program really write a good movie? Here's a test," *The Guardian*, Mar. 24, 2023. Available: https://www.theguardian.com/film/2023/mar/24/chapgpt-movie-script-ai

[40] A. Elgammal, B. Liu, M. Elhoseiny, and M. Mazzone, "CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms," *arXiv (Cornell University)*, Jun. 2017, doi: https://doi.org/10.48550/arxiv.1706.07068.

[41] A. Hertzmann, "Toward Modeling Creative Processes for Algorithmic Painting," *arXiv (Cornell University)*, May 2022, doi: https://doi.org/10.48550/arxiv.2205.01605.

[42] E. S. Mikalonytė and M. Kneer, "Can Artificial Intelligence Make Art?," *ACM Transactions on Human-Robot Interaction*, vol. 11, no. 4, May 2022, doi: https://doi.org/10.1145/3530875.

[43] A. Eslami *et al.*, "Attend, Infer, Repeat: Fast Scene Understanding with Generative Models," *arXiv (Cornell University)*, Jan. 2016, doi: https://doi.org/10.48550/arxiv.1603.08575.

[44] I. Higgins *et al.*, "Early Visual Concept Learning with Unsupervised Deep Learning," *arXiv (Cornell University)*, Jan. 2016, doi: https://doi.org/10.48550/arxiv.1606.05579.

[45] D. J. Rezende, Ali, S. Mohamed, P. Battaglia, M. Jaderberg, and N. Heess, "Unsupervised Learning of 3D Structure from Images," *arXiv (Cornell University)*, Jan. 2016, doi: https://doi.org/10.48550/arxiv.1607.00662.

[46] J. Balaguer, H. Spiers, D. Hassabis, and C. Summerfield, "Neural Mechanisms of Hierarchical Planning in a Virtual Subway Network," *Neuron*, vol. 90, no. 4, pp. 893–903, May 2016, doi: https://doi.org/10.1016/j.neuron.2016.03.037.

[47] A. Solway *et al.*, "Optimal Behavioral Hierarchy," *PLoS Computational Biology*, vol. 10, no. 8, p. e1003779, Aug. 2014, doi: https://doi.org/10.1371/journal.pcbi.1003779.

[48] Q. J. M. Huys, N. Eshel, E. O'Nions, L. Sheridan, P. Dayan, and J. P. Roiser, "Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees," *PLoS Computational Biology*, vol. 8, no. 3, p. e1002410, Mar. 2012, doi: https://doi.org/10.1371/journal.pcbi.1002410.

[49] H. Chin *et al.*, "The Potential of Chatbots for Emotional Support and Promoting Mental Well-Being in Different Cultures: Mixed Methods Study," *Journal of Medical Internet Research*, vol. 25, p. e51712, Oct. 2023, doi: https://doi.org/10.2196/51712.

[50] N. Bostrom and E. Yudkowsky, "The Ethics of Artificial Intelligence," *Artificial Intelligence Safety and Security*, pp. 57–69, Jul. 2018, doi: https://doi.org/10.1201/9781351251389-4.

[51] S. Fosso Wamba, R. E. Bawack, C. Guthrie, M. M. Queiroz, and K. D. A. Carillo, "Are we preparing for a good AI society? A bibliometric review and research agenda," *Technological Forecasting and Social Change*, vol. 164, p. 120482, Mar. 2021, doi: https://doi.org/10.1016/j.techfore.2020.120482.

APPENDIX

TABLE I. TERM PROJECT TASK LIST

Week	Task #	Task Description	Time (hours)	Status
3	1	Create template task list	0.25	Done
3	2	Choose broad topic	0.25	Done
3	3	Finalize research questions	1	Done
3	4	Identify relevant literature	2	Done
3	5	Write Project Pitch	4.5	Done
Project Pitch Due				
4	6	Create lit. review task list	0.5	Done
4	7	Read and review initial batch for questions 1, 2	5	Done
5	8	Search and review additional literature for q. 1, 2	2.5	Done
5	9	Read and take notes for final batch for q. 1, 2	10	Done
6	10	Review progress	0.5	Done
6	11	Read and review initial batch for questions 3, 4	5	Done
7	12	Search and review additional literature for q. 3, 4	2	Done
7	13	Read and take notes for final batch for q. 3, 4	10	Done
7	14	Review progress and plan final report	1.5	Done
8	15	Results and Discussion	12	Done
8	16	Abstract and Introduction	2	Done
9	17	Experiment Design	2	Done
9	18	Conclusion and Limitations	3	Done
9	19	References and Table, Review Final Report	3	Done
Final Report Due				
10	20	Introduction	0.5	Pending
10	21	Design	2	Pending
10	22	Results	3	Pending
10	23	Conclusion	1	Pending
10	24	Review and Edit Presentation	1	Pending
10	25	Prepare script	2	Pending
10	26	Record presentation	1.5	Pending
10	27	Prepare and submit video	2	Pending
Final Presentation Due				
		Total Hours	**80**	