Understanding Consciousness In Artificial Intelligence

A Dissertation Report for

PHI - 651 Dissertation
16 Credits

Submitted in partial fulfilment of Masters Degree
MA in Philosophy

by

MEERA NARAYANANKUTTY

Seat Number: 22P0200004
ABC ID: 325526068650
PRN: 202200058

Under the Supervision of

DR. WALTER MENEZES
Assistant Professor

School of Sanskrit, Philosophy and Indic Studies

Goa University

April 2024

Examined by:

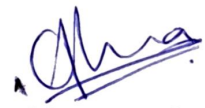
02/05/2024

Seal of the School

DECLARATION BY STUDENT

I hereby declare that the data presented in this Dissertation report entitled, **"Understanding Consciousness in Artificial Intelligence"** is based on the results of investigations carried out by me in M.A.Philosophy at the School of Sanskrit, Philosophy and Indic Studies, Goa University under the Supervision of **Dr. Walter Menezes** and the same has not been submitted elsewhere for the award of a degree or diploma by me. Further, I understand that Goa University or its authorities will not be responsible for the correctness of observations / experimental or other findings given the dissertation.

I hereby authorise the University authorities to upload this dissertation on the dissertation repository or anywhere else as the UGC regulations demand and make it available to any one as needed.



Meera Narayanankutty

Student

Seat no: 22P0200004

Date: 02/05/2024

Place: Goa University

COMPLETION CERTIFICATE

This is to certify that the dissertation report **"Understanding Consciousness in Artificial Intelligence"** is a bonafide work carried out by **Ms. Meera Narayanankutty** under my supervision in partial fulfilment of the requirements for the award of the degree of **Master of Arts in the Discipline MA. Philosophy** at the School of Sanskrit, Philosophy and Indic Studies, Goa University.

02/05/2024

Dr. Walter Menezes

Supervising Teacher

Date:

Dean of the School

Date: 02/05/2024

Place: Goa University

TABLE OF CONTENTS

PREFACE

The question of whether Consciousness is unique to biological beings or can be replicated by machines has captivated Philosophers, Scientists and IT Technicians alike. This Dissertation, "Understanding Consciousness in Artificial Intelligence", ventures into this very inquiry. We embark on a philosophical odyssey, exploring the essence of consciousness and its potential manifestation in the realm of Artificial Intelligence. As AI rapidly evolves, blurring the lines between human and machine, the possibility of conscious machines compels us to revisit foundational propositions as well as recent developments and news on the same. Our exploration delves into the philosophical and scientific underpinnings of consciousness, dissecting its various definitions and the ongoing debate about its physical correlates. We will encounter the contrasting perspectives of recently developed AI or Robots which aspire to human-level or superior Intelligence and Consciousness. The ethical considerations surrounding AI Consciousness are paramount. If machines can achieve sentience, what rights and responsibilities do they possess? How can we ensure the ethical development and deployment of Conscious AI? These are critical questions that demand thoughtful consideration. As we delve into the frontiers of AI research, we will analyse how recent breakthroughs influence our understanding of both Intelligence and Consciousness. Ultimately, this work seeks to illuminate the path forward, fostering a dialogue that is both Philosophically rigorous and Ethically responsible. The quest to "Understand Consciousness in AI" is not merely a scientific or technological pursuit, but a profound Philosophical exploration of what it means to be human in an evolving world which will be increasingly shaped by machines.

Meera Narayanankutty

ACKNOWLEDGMENTS

The successful completion of this dissertation is indebted to the invaluable guidance and support of several individuals. I would like to express my deepest gratitude to my dissertation supervisor, Dr. Walter Menezes. His constant encouragement, insightful feedback, and dedication throughout this project were instrumental in my success.

I am also grateful to Prof. Koshy Tharakan, Professor and Dean of SSPIS, for granting me the opportunity to pursue this dissertation topic and for providing the necessary facilities to conduct my research.

My appreciation extends to S. Baskar, Associate Professor of Computer Science, Prof. Sanjyot Pai D. Vernekar, Dr. Norma Menezes, and Miss Rajavi Naik for their support and contributions to this dissertation. I would like to thank the Assistant Librarian of Goa University for their invaluable assistance in providing me with the resources necessary to complete my dissertation.

Furthermore, I would like to acknowledge the support and appreciation of my friends, Jomon, Sneha, Sharon, Ansa, Bhagyashri, Suraj, Satheesh, Adithyan and Faazil. Their encouragement and understanding throughout this endeavour are greatly appreciated.

Finally, I am immensely grateful to my family for their constant encouragement and support.

ABBREVIATIONS USED

Entity	Abbreviation
Artificial Consciousness	AC
Artificial Intelligence	AI
Artificial Super Intelligence	ASI
Electroencephalography	EEG
Emotional Intelligence	EQ
Functional Magnetic Resonance Imaging	fMRI
Generative Pre-trained Transformer	GPT
International Conference on Robotics and Automation	ICRA
Integrated Information Theory	IIT
Intelligence Quotient	IQ
Massachusetts Institute of Technology	MIT
Neural Correlates of Consciousness	NCCs
Orchestrated Objective Reduction	Orch OR

ABSTRACT

This dissertation delves into the intricate relationship between artificial intelligence (AI) and consciousness, exploring the philosophical and scientific landscape surrounding this burgeoning field. The opening chapter establishes the groundwork by delving into the multifaceted concept of intelligence. The rise of AI challenges the anthropocentric view of intelligence, prompting us to consider alternative forms of intelligence beyond human-like reasoning. This chapter emphasizes the evolving concept of intelligence and the need to move beyond a singular definition.

Chapter two shifts focus to the enigmatic concept of consciousness. It explores the ongoing philosophical discussions surrounding the relationships and differences between consciousness and intelligence. While some argue that strong AI could achieve human-level or superior intelligence and consciousness, others emphasize the importance of subjective experience in defining consciousness. This chapter delves into classic thought experiments like the "Philosophical Zombie" introduced by David Chalmers, which challenges the idea that consciousness solely arises from physical processes in the brain. It also explores functionalist perspectives like Daniel Dennett's "Scaffolding Mind" theory, highlighting the potential for machines to achieve functional equivalence with consciousness through complex information processing. Additionally, it addresses Hilary Putnam's "Computational Mind" theory, the possibility of machines achieving computational equivalence to a conscious mind. This chapter also explores the feasibility and ethical considerations of artificial consciousness.

Final chapter delves into the Philosophical Debate surrounding Artificial Consciousness (AC), exploring arguments for and against its achievability. Proponents

highlight advancements in AI and the possibility of machines achieving complex functionality equivalent to consciousness. Critics argue that replicating human cognition might not capture the subjective experience that defines consciousness.

Overall, this dissertation argues that while the question of whether AI can achieve true consciousness remains unanswered, understanding its potential implications is crucial. As the field of AI progresses, a nuanced understanding of intelligence and consciousness becomes increasingly important for navigating the ethical and philosophical challenges that lie ahead.

CHAPTER 1

INTRODUCTION

1.1 Background

Understanding Consciousness in Artificial Intelligence has preoccupied philosophers for a very long time under the branch of 'The philosophy of artificial intelligence'. This comes within the study of 'Philosophy of mind' that explores artificial intelligence and its implications for knowledge and understanding of intelligence, ethics, consciousness, epistemology, and free will. Basically, the questions this field addresses includes Can AI have a consciousness?, Does it works really intelligently?, Does it have a qualia[1] ?.. etc. This work will be focusing on the evolving methodologies in AI research and how these emerging technologies, such as brain – computer interfaces affect the very definition of consciousness.

Understanding consciousness in artificial intelligence represents an unparalleled opportunity to elevate the technology beyond mere computation, infusing it with elements of cognition and self-awareness. The Artificial Intelligence Consciousness debate is not just a technical one. It intervenes with our moral fabric, philosophical beliefs, and how we view the essence of life. Deepening our understanding as we stand at this crossroads is

[1] A philosophical term for sensory experiences that have distinctive subjective qualities but lack any meaning or external reference to the objects or events that cause them, such as the painfulness of pinpricks or the redness of red roses. The term is virtually synonymous with sense data. In philosophy of mind, qualia are defined as instances of subjective, conscious experience *Qualia*. Oxford Reference. https//www.oxfordreference.com/display/10.1093/oi/authority.20110803100357499#~text=A%20philosophical %20term%20for%20sensory,virtually%20synonymous%20with%20sense%20data.

not just a scientific pursuit, but a moral imperative. Nature magazine, Laid Mudrik discusses a new book by Daniel C Dennet, Tufts University,- He talks of the need to first understand our own consciousness, the self and free will before we try to delve into the intricacies of artificial intelligence.

Perhaps no aspect of mind is more familiar or more puzzling than consciousness and our conscious experience of self and world. And we're using the same consciousness to analyse 'the very nature of consciousness' is the most interesting part. The study of consciousness in AI holds profound implications for the ethical framework underpinning its development and elucidating the mechanisms through which artificial systems perceive, interpret, and respond to stimuli, researchers can instil principles of empathy, morality, and accountability within AI architectures. This, in turn, mitigates the risk of unintended consequences and ensures that AI remains aligned with societal values and norms. Understanding Consciousness is also very crucial because it will be helpful for the developers too about their responsibility towards society and handling these systems. It'll also help them in problem-solving, creativity, and adaptation, fostering a new era of AI-driven innovation and discovery.

1.2 Objectives

1.2.1 Exploring the Essence of AI Consciousness

First objective is to delve into the philosophical and scientific inquiry into the nature of consciousness and its potential emergence in AI. It raises questions about the definition of consciousness and whether machines can replicate or surpass it.

1.2.2 Moral Fabric of AI Consciousness Debate

There will be discussion beyond technical aspects to explore the ethical and moral implications of AI consciousness. It addresses questions about the rights and responsibilities of conscious AI, potential biases, and the impact on societal norms. It also explores the potential risks and benefits, and the responsibility of scientists to ensure ethical development.

1.2.3 Interconnection of AI and Human Experience

This objective explores how AI might influence and be influenced by human experience. It examines how AI systems might learn from and interact with humans, potentially shaping behaviour, decision-making, and even our understanding of ourselves.

1.2.4 Philosophical Underpinnings

The philosophical implications of AI consciousness, particularly its potential to challenge our understanding of life and sentience will be addressed.

1.2.5. Shaping the Future AI's Impact on Consciousness Definition

This focuses on how AI development might influence how we define and understand consciousness in the future. It considers how our understanding might evolve as we interact with increasingly sophisticated AI systems by analysing recent developments and breakthroughs.

1.2.6. Human-AI Coexistence

Through this objective, I'll focus on the importance of exploring how humans and AI can coexist and collaborate effectively. It emphasises the need for open communication, collaboration, and ethical considerations to ensure a beneficial future for both.

1.2.7 AI Consciousness as a Global Discourse

This objective underscores the importance of approaching AI consciousness as a global issue. It encourages international collaboration, dialogue, and knowledge sharing to ensure responsible development and integration of AI into society. This global discourse should involve not only scientists, engineers, and policymakers but also ethicists, philosophers, artists, and the general public. I'll be analysing how Al consciousness is portrayed in popular media (e.g., films, books, video games) and recent news coverage can provide valuable insights into public perception and potential biases

1.3 Research Question

Are we intelligent enough to understand intelligence? One approach to answering this question is "Artificial intelligence". The term "AI" could be attributed to John McCarthy of MIT (Massachusetts Institute of Technology), which Marvin Minsky (Carnegie-Mellon University) defines as "The construction of computer programs that engage in tasks that are currently more satisfactorily performed by human beings because they require high-level mental processes such as perceptual learning, memory organisation and critical reason." One of the first things that must be clarified is the ambiguous word Artificial. This adjective can be used in two senses, and it is important to determine which one applies in the term artificial intelligence. The Word artificial is used in one sense when it

is applied, say, to flowers, And in another sense when it is applied to light. In both cases something is called artificial because it is fabricated. But in the first Usage artificial means that the thing seems to be, but really is not, What it looks like. The artificial is merely apparent; it just shows How something else looks. Artificial flowers are only paper, not Flowers at all; anyone who takes them to be flowers is mistaken. But Artificial light is light and it does illuminate. It is fabricated as a Substitute for natural light, but once fabricated it is what it seems to Be. In this sense the artificial is not merely apparent, not simply an imitation of something else. The appearance of the thing reveals what it is, not how something else looks.

In which sense do we use the word artificial when we speak of artificial intelligence? Critics of artificial intelligence, those who disparage the idea and say it has been overblown and oversold, would claim that the term is used in the first sense, to mean the merely apparent. They would say that artificial intelligence is really nothing but complex mechanical structures and electrical processes that present an illusion of some sort of thinking. Supporters of the idea of artificial intelligence, those who claim that the term names something genuine and not merely apparent, would say that the word artificial is used in the second of the senses we have distinguished. Obviously, they would say, thinking machines are artefacts; they are run by human beings; but once made and set in motion, the machines do think. Their thinking may be different from that of human beings in some ways, just as the movement of a car is different from that of a rabbit and the flight of an aeroplane is different from that of a bird, but it is a kind of genuine thinking, just as there is genuine motion in the car and genuine flight in the plane.[i](Sokolowski, 45)

1.4 Scope

Artificial intelligence continues to make significant strides, reshaping industries, revolutionising technologies, and pushing the boundaries of what is possible. From breakthroughs in machine learning algorithms to innovative applications in healthcare and beyond, recent developments underscore the transformative potential of AI. In the realm of machine learning, recent advancements have focused on improving the efficiency and performance of algorithms. One notable breakthrough is the development of Transformer-based models, such as OpenAI's GPT (Generative Pre-trained Transformer) series. These models, leveraging self attention mechanisms, have achieved remarkable results in natural language processing tasks, including language generation, translation, and text summarization. The release of GPT-3 has garnered widespread attention for its ability to generate human-like text, sparking discussions about the capabilities and ethical implications of large language models. McKay Wrigley, a 23 year old computer programmer from Salt Lake City, was one of the few invited to tinker with the system, which uses everything it has learned from that vast sea of digital text to generate new language on its own.

1.4.1 Deepmind

Moreover, advancements in reinforcement learning have led to significant progress in autonomous systems and robotics. DeepMind's AlphaFold[2], for instance, made headlines

[2]AlphaFold is an AI system developed by DeepMind that predicts a protein's 3D structure from its amino acid sequence

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon,
A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., … Velankar, S. (2021a). Alphafold protein structure database Massively expanding the structural

for its groundbreaking achievements in protein folding prediction, a critical problem in bioinformatics with implications for drug discovery and disease research. By leveraging reinforcement learning techniques, AlphaFold demonstrated unprecedented accuracy in predicting the 3D structure of proteins, offering new insights into molecular biology and accelerating scientific discovery. On July 28, 2022, Google's DeepMind released the structure of 200 million proteins, literally everything that exists. This is said to be the most important achievement of AI ever, namely a 'solution' to the protein-folding problem.

1.4.2 Healthcare

In healthcare, AI is revolutionising diagnosis, treatment, and patient care. Recent developments include the use of AI-powered medical imaging systems to assist radiologists in detecting diseases such as cancer. Companies like Google Health and Siemens Healthineers have introduced AI algorithms capable of analysing medical images with high accuracy, improving diagnostic accuracy and patient outcomes. Additionally, AI-driven virtual assistants and chatbots are transforming healthcare delivery by providing personalised patient support, appointment scheduling, and remote monitoring, enhancing accessibility and efficiency in healthcare services. "In the future, a radiologist who has knowledge of AI may replace a radiologist without any knowledge of AI. While AI alone may not be enough, the combination of man and machine can do better," Dr. Apparao, also the founder of the Telugu Association of North America and the American Association of Physicians from India, said.

---

coverage of protein sequence space with high-accuracy models. Nucleic Acids Research, 50(D1). https//doi.org/10.1093/nar/gkab1061

1.4.3 Transportation

Furthermore, AI is driving innovation in autonomous vehicles and transportation systems. Companies like Tesla, Waymo, and Uber are investing heavily in AI technologies to develop self-driving cars and intelligent transportation networks. These advancements hold the promise of safer, more efficient transportation systems with reduced congestion and emissions, paving the way for a future of autonomous mobility. In the realm of business and finance, AI is powering predictive analytics, fraud detection, and algorithmic trading systems. Financial institutions are leveraging AI algorithms to analyse market trends, optimise investment strategies, and mitigate risks. Recent developments include the use of deep learning models for high-frequency trading and sentiment analysis of social media data to predict market movements, demonstrating AI's potential to drive innovation and efficiency in financial markets.

Although, Scientists and neurologists found that all types of neural networks are a part of the intellectual work carried out by each area of the human brain, they are only a substitute and do not perform all the functions of the human brain. AI has not been able to cooperate with whole brain functions such as self understanding, self control and so on."(NewYork Times, 2024)"

1.4.4 BI Intelligent model

There are many approaches proposed to solve the limitations of recent AI. However, these models are simply extended from the current AI models. The BI intelligent learning model fuses the benefits of artificial life (AL) and AI. Currently, the mainstream research on deep learning is a method of learning expressions extracted from essential information of observational data by a deep neural network with a large number of layers. However,

research on multitask learning that learns multiple tasks at the same time and transition studies that divert learning results for a certain task to other tasks is still insufficient. For this reason, AI models based on unsupervised learning and shallow neural networks will become trends in future. Therefore A new intelligent learning model has been developed with a small database and the ability to understand concepts.

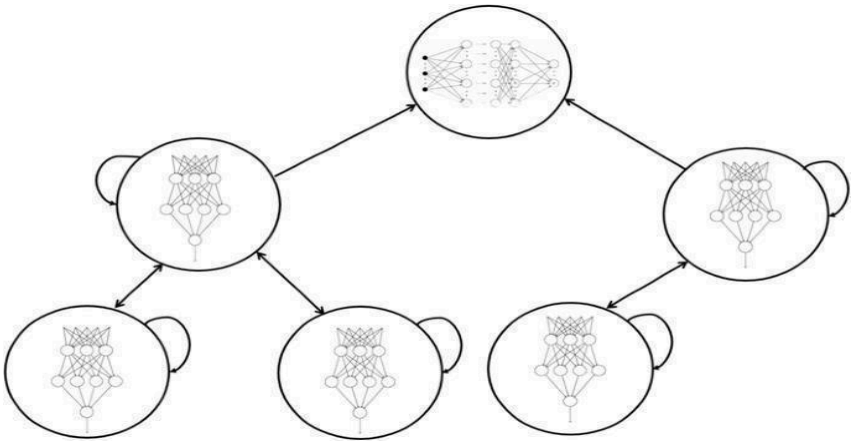


Figure 1[3]

Research on current AI mainly focuses on individual areas such as dialogue comprehension, visual recognition, and auditory discrimination and so on. Research on whole brain functions is still insufficient. Different from Neuroevolution of Augmenting Topologies (NEAT)[4]. The proposed BI mode network does not just use the neural network

[3] The concept of the BI model network. Different neural networks are connected by artificial life-based networks, which can share the parameters, trained results.
Brain Intelligence Go Beyond Artificial Intelligence – Scientific Figure on ResearchGate. https//www.researchgate.net/figure/The-concept-of-the-BI-model-network-Different-neural-networks-are-connected-by_fig3_317351566 [accessed 25 Feb, 2024]

[4] NeuroEvolution of Augmenting Topologies (NEAT) is a genetic algorithm (GA) for the generation of evolving artificial neural networks (a neuroevolution technique) developed by Kenneth Stanley and Risto Miikkulainen in 2002 while at The University of Texas at Austin.
Stanley, K. O., & Miikkulainen, R. (2002). *Evolving neural networks through augmenting topologies. Evolutionary Computation, 10(2), 99–127.* https//doi.org/10.1162/106365602320169811

structure and parameter optimization mechanism, it improves the structure of current AI models using S-system.. The BI model network is investigated from an engineering point of view, in the future, there is high scopes of developing a super-intelligent brain function model that intends to discover problems itself and autonomously enhance its abilities.(Huimin Lu, Yujie Li, Min Chen, and Seiichi Serikawa (2018),Fig. 3)

In the realm of artificial intelligence, where rapid advancements underscore its transformative potential, there exists a profound imperative to delve into the intricacies of consciousness. As Artificial Intelligence continues its evolutionary journey , the exploration of consciousness emerges as a pivotal pursuit, essential for navigating the complex landscape of ethical considerations, and transparency.

---

[i] Sokolowski, R. (1988). Natural and Artificial Intelligence. Daedalus, 117(1), 45–64.

http//www.jstor.org/stable/20

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter dives deeper into the nature of intelligence by analysing influential books on the topic. Daniel Goleman's groundbreaking work, Emotional Intelligence (1995), challenges the traditional view by introducing the concept of EQ alongside IQ. Following this, it explores Thinking, Fast and Slow (2011) by Nobel laureate Daniel Kahneman. This examines Intelligence and How to Get It (2009) by Richard Nisbett. Nisbett emphasises the importance of metacognition, or "thinking about thinking," and strategic learning approaches in developing one's intelligence.

Moving forward, the chapter tackles the fundamental question: how are intelligence and consciousness interconnected? It explores the debate surrounding qualia and the explanatory gap between the physical world of the brain and the subjective world of experience. It also critically examines deep learning algorithms and their remarkable feats in specific tasks like image recognition or game playing and also the advancement of Artificial Intelligence (AI) is also discussed in detail. However, it challenges the notion that such capabilities equate to consciousness, highlighting the opacity of deep learning models and their potential lack of true understanding. The chapter concludes by exploring the concept of Ethical Implications of creating Consciousness.

2.2 What Is Intelligence?

Intelligence is a multifaceted concept that has been pondered by philosophers for millennia. If we look into traditional view positions in western philosophy, we can see one view, intelligence as the pinnacle of human reason. In Plato's influential theory of Forms, intelligence transcends mere problem-solving; it is the key to unlocking a higher realm of perfect, unchanging concepts. This realm, accessible through reason, provides the foundational principles that illuminate our messy world of experience. Plato's student, Aristotle, further solidified this rationalist view of intelligence. He championed logic, deduction, and the ability to formulate universal truths as the hallmarks of the intelligent mind. In this view, intelligence is not merely about accumulating knowledge, but about wielding reason to discern the underlying order and structure of the universe.

This Platonic ideal of intelligence continues to hold significant sway. It informs our educational systems, emphasising critical thinking, analysis, and the ability to construct sound arguments. It fuels our scientific endeavours, where reason is the guiding light in our quest to understand the natural world. However, the Platonic ideal is not without its critics. Some argue that it overemphasised reason at the expense of other forms of intelligence, such as creativity, intuition, and emotional intelligence. They point out that these qualities are also essential for navigating the complexities of human experience.

However, the empiricist school, spearheaded by John Locke, offers a compelling counterpoint. Locke posits that the mind is 'Tabula rasa'[5], and knowledge is solely

[5] Tabula Rasa- In his work An Essay Concerning Human Understanding(1689), Locke argues that, at birth, the mind is a tabula rasa (a blank slate) that we fill with 'ideas' as we experience the world through the five senses.
John Locke, An Essay concerning human understanding (1689)

acquired through experience. In this view, intelligence is not about accessing a realm of perfect forms, but rather the ability to learn from experiences, identify patterns, and adapt to new situations. This empiricist perspective underscores the importance of curiosity, experimentation, and a willingness to revise beliefs in the light of new evidence. Curiosity fuels the desire to explore and gather experiences. Experimentation allows us to test hypotheses and refine our understanding of the world. And an open mind, receptive to new evidence, is crucial for intellectual growth. The tension between reason and experience is a vital one in our quest for knowledge. The Platonic ideal provides a framework for organising and interpreting our experiences. But without the empirical approach, we risk getting lost in abstract reasoning, disconnected from the real world.

The rise of artificial intelligence (AI) throws a wrench into the long-held association of intelligence with human-like reason. Alan Turing's Turing test[6], which proposes that a machine capable of engaging in indistinguishable conversation with a human is intelligent, challenges the anthropocentric view. It compels us to consider if reasoning, as we understand it, is the sole yardstick for intelligence. Could machines, operating on entirely different principles, exhibit a form of intelligence that we haven't even begun to fathom?

AI's ability to process information and learn at an unprecedented rate suggests a kind of intelligence that complements, and perhaps even surpasses, human intelligence in specific domains. Machines can sift through mountains of data to identify patterns and make predictions that would be beyond the reach of even the most brilliant minds. This raises intriguing questions about the future of intelligence: can human and machine intelligence

[6] Turing test, in artificial intelligence, a test proposed (1950) by the English mathematician Alan M. Turing to determine whether a computer can think.
Alan Turing, "Computing Machinery and intelligence, Mind", Vol. 59, no. 236(1950),pp433-460

work together to create a new kind of cognitive power?. The definition of intelligence is likely to continue evolving as our understanding of the mind and our ability to create ever-more sophisticated machines expands.

Howard Gardner's theory of multiple intelligences [7]throws another wrench into the works. He proposes that there isn't one single form of intelligence, but rather a spectrum encompassing logical-mathematical, linguistic, musical, spatial, bodily kinesthetic, interpersonal, and intrapersonal intelligences. Gardner's theory resonates with the idea that AI may exhibit intelligence in ways we don't fully understand yet. Machines may process information and solve problems differently than humans, but that doesn't negate their intelligence.

The evolving concept of intelligence compels us to move beyond a single definition from a narrow view. It's about recognizing the many forms intelligence can take, whether it's the analytical power of a chess grandmaster or the creativity of a composer. As our understanding of the mind and technology expands, our definition of intelligence will undoubtedly continue to grow.

---

[7] This theory suggests that traditional psychometric views of intelligence are too limited. Gardner first outlined his theory in his 1983 book Frames of Mind: The Theory of Multiple Intelligences, where he suggested that all people have different kinds of "intelligences." Gardner proposed that there are eight intelligences, and
has suggested the possible addition of a ninth known as "existentialist intelligence."
Howard Gardener, Framed of Mind. The theory of multiple intelligences(1983)

2.3 Consolidated Analysis of Influential Books on Intelligence

2.3.1 Emotional Intelligence by Daniel Goleman (1995)

Daniel Goleman's groundbreaking work, Emotional Intelligence (1995), revolutionised understanding of intelligence. While traditional intelligence (IQ) focuses on cognitive abilities like reasoning and problem-solving, Goleman proposes emotional intelligence (EQ)[8] as an equally crucial factor. EQ encompasses a set of skills that allow us to understand, manage, and utilise our own emotions, as well as perceive and influence the emotions of others.

Goleman outlines five core components of EQ

1. Self-awareness This is the foundation of EQ, recognizing our emotions and their impact on our thoughts, behaviours, and interactions with others.
2. Self-regulation Effectively managing our emotions is key to preventing them from controlling our responses. This involves techniques like self-calming strategies and delaying gratification.
3. Motivation Goleman highlights the power of emotions to propel us towards our goals. Understanding how to leverage emotions like passion and perseverance can fuel our drive and achievement.
4. Empathy The ability to understand and share the feelings of others is a cornerstone of strong EQ. Empathy allows us to build rapport, fosters trust, and strengthens relationships.

[8] Daniel Goleman defines emotional intelligence (EQ) as the ability to understand, use, and manage your own emotions in positive ways to achieve your goals. It also involves understanding, empathising with, and influencing the emotions of others.
Goleman, Daniel. Emotional Intelligence. Bantam Books, 1995.

5. Social skills EQ empowers us to navigate complex social interactions effectively. This includes our ability to communicate clearly, build relationships, manage conflict, and influence others through emotional intelligence.

The significance of EQ lies in its far-reaching impact on various aspects of our lives. Goleman argues that strong EQ is essential for social success. It equips us to navigate complex social situations, build rapport with colleagues and friends, and resolve conflicts constructively. In the workplace, leaders with high EQ are adept at motivating and inspiring their teams, fostering collaboration, and creating a positive work environment. Furthermore, EQ plays a crucial role in personal wellbeing. By managing stress effectively and maintaining healthy relationships, we can achieve greater emotional balance and fulfilment.

Goleman's work challenges the notion of a singular form of intelligence. By integrating EQ with IQ, we gain a more comprehensive understanding of intelligence and its role in success.

2.3.2 Thinking, Fast and Slow by Daniel Kahneman (2011) .

Nobel laureate Daniel Kahneman's notable book, Thinking, Fast and Slow, delves into the fascinating world of human cognition by proposing a unique two-system model of thought. This model sheds light on how we make decisions, navigate the world, and ultimately, how we express intelligence.

System 1, the star of the show in Kahneman's narrative, is our fast-thinking, intuitive system. It operates on autopilot, processing information quickly and effortlessly through

mental shortcuts called heuristics[9]. Ever stopped to consider the price of a seemingly expensive item simply because it was placed next to a much pricier one? That's System 1 at work, employing the anchoring bias to make a snap judgement. While this system's efficiency is undeniable, its reliance on heuristics can lead to biases and errors.

System 2, the more deliberate counterpart, is slower, more logical, and analytical. It is the system we engage when we meticulously solve a maths problem or carefully weigh the pros and cons of a critical decision. However, System 2 is also lazy, and its activation requires conscious effort. We often default to System 1's quicker judgments, even when they might be flawed.

Kahneman's work shows us that the understanding that both systems contribute to intelligent behaviour. System 1's swiftness allows us to react quickly in everyday situations, while System 2 provides the necessary depth for complex reasoning. The challenge lies in recognizing when each system is guiding our thoughts and actively engaging System 2 when necessary to counter System 1's biases.

By acknowledging the existence of these biases, such as the overconfidence bias that leads us to trust our judgments more than we should, we can begin to mitigate their influence. Kahneman emphasises that through deliberate practice and mindful awareness, we can train ourselves to activate System 2 more readily, ultimately promoting more intelligent and well-considered choices.

---

[9] Kahneman and Tversky's "same" heuristic refers to our tendency to judge the likelihood of events based on their similarity to past experiences or prototypes. This mental shortcut can lead to biases in decision-making.
Tversky, Amos, and Daniel Kahneman. "Judgement under Uncertainty Heuristics and Biases." Science, vol. 185, no. 4157, 1974, pp. 453-478.

Thinking, Fast and Slow is a guide to navigating the intricacies of our own minds. By understanding the interplay between System 1 and System 2, we gain the power to make more informed decisions, reduce the influence of mental shortcuts, and ultimately, cultivate a more nuanced form of intelligence.

2.3.3 Intelligence and How to Get It by Richard Nisbett (2009)

Richard Nisbett's thought-provoking book, Intelligence and How to Get It (2009), challenges the traditional view of intelligence as a predetermined and unchangeable characteristic. Instead, Nisbett proposes a growth mindset, arguing that intelligence is malleable and can be actively developed through effort, motivation, and strategic learning approaches.

This perspective stands in stark contrast to the fixed mindset, which assumes intelligence is a fixed quantity assigned at birth. Nisbett dismantles this notion by emphasising the importance of metacognition, or "thinking about thinking." Metacognition allows us to become aware of our own learning strengths and weaknesses. By reflecting on our cognitive processes, we can identify areas for improvement and develop targeted strategies to enhance our learning.

For instance, Nisbett highlights the significance of practice and perseverance in developing problem-solving skills. Just as with any physical skill, intellectual abilities can be honed and strengthened through dedicated effort. Nisbett goes beyond mere

practice, advocating for strategic practice. This involves deliberately challenging ourselves with problems that are slightly beyond our current grasp, prompting us to stretch our cognitive abilities and fostering growth.

Furthermore, Nisbett underscores the role of cultural differences in shaping how intelligence is viewed and nurtured. In some cultures, rote memorization and standardised tests might be emphasised, while others might place a greater value on creativity and critical thinking. Recognizing these variations broadens our understanding of intelligence and highlights the importance of fostering a learning environment that cultivates a diverse range of cognitive skills. By acknowledging the role of effort, motivation, and strategic learning, Nisbett's work empowers us to become active participants in developing our intelligence.

2.4 What Is Consciousness?

Consciousness, the subjective experience of sentience and awareness, has captivated philosophers for centuries. Its nature lies in its inherent subjectivity , it can only be truly understood from within. While science investigates the neural correlates of consciousness, philosophy grapples with its fundamental essence and its place in the universe. We can trace the exploration of consciousness in both

Indian and Western philosophical traditions, each offering unique perspectives. Indian philosophy is characterised by a rich tradition that engages in systematic examinations of consciousness, placing particular emphasis on its phenomenological and transcendental aspects. Arguably, the nature and role of consciousness stand out as a highly contested topic among various schools of Indian philosophy, Meanwhile, Western philosophy, from

the early dialogues of Plato and Aristotle to the dualism of Descartes and the empiricism of Locke, has contributed diverse viewpoints on the mind-body relationship.

And this oldest philosophical question, doesn't have a good scientific answer. It has been called the last great mystery of science. In scientific discussions about consciousness, what is usually meant is phenomenal consciousness. Consciousness is often divided into two different levels., state and content. State consciousness concerns the level of the entire system. Are you conscious and if so, what is your consciousness like? Are you drowsy or wide awake, for example? Content consciousness concerns what you're conscious of. For example, a specific word, a colour or a taste.

Phenomenal experience, the subjective quality of our conscious states, lies at the heart of consciousness. Philosophers like Thomas Nagel, in his seminal work "What Is It Like to Be a Bat?" (1974), argue for its paramount importance in understanding consciousness. Nagel highlights the limitations of solely analysing physical properties to comprehend consciousness. He asks, "What is it like to be a bat?" – a question that goes beyond the bat's neurology. We can objectively study a bat's brain, but we cannot truly access its subjective experience of the world, its feelings, or its qualia.

This inability to fully grasp another being's phenomenal experience underscores its significance. It is what makes consciousness a uniquely personal and private phenomenon. Our own experiences, from the taste of a strawberry to the ache of a bruise, are undeniable hallmarks of being conscious. Phenomenal experience isn't just a passive quality. It shapes our perception of the world, influences our actions, and contributes to our sense of self. It is through these subjective experiences that we navigate and make sense of the world around us. If consciousness is a fundamental feature of reality, then

understanding phenomenal experience is crucial to unlocking the mysteries of our own existence.

Eastern philosophical traditions provide a distinct lens through which to explore consciousness, complementing the questions posed by Western philosophy.

The Upanishads, core scriptures, introduce the concept of Atman, the true self. Atman transcends the limitations of the ego, the everyday self. It is described as a state of pure awareness, Brahman, the ultimate reality. This perspective emphasises a consciousness that is not tethered to the impermanent ego but is instead a fundamental essence of existence.

Buddhist philosophy, as explored in the Pali Canon, offers another dimension. It delves into the nature of perception and the impermanence of consciousness. Buddhist teachings emphasise the interconnectedness of all things and the impermanent nature of our experience, including consciousness itself. Through practices like meditation, Buddhists aim to cultivate a state of heightened awareness and detachment from fleeting thoughts and desires.

Eastern traditions don’t necessarily grapple with the hard problem of consciousness in the same way Western philosophers do. However, they offer valuable insights into the nature of subjective experience, emphasising the potential for cultivating a more expansive and aware state of consciousness. By integrating Western and Eastern philosophical perspectives, we gain a richer understanding of consciousness. Western philosophy provides a framework for analysing the mechanics of the mind, while Eastern traditions offer insights into the subjective nature of experience and the potential for transcending the egoic self.

Ultimately, the mystery of consciousness persists. Yet, through continued philosophical inquiry, we may one day approach a more comprehensive understanding of this fundamental aspect.

2.5 Interconnection Between Consciousness And Intelligence

The human experience is woven from threads of intelligence and consciousness. We navigate the world, solve problems, and feel a spectrum of emotions. But how are these two fundamental aspects of our being related? Is consciousness a prerequisite for intelligence, or can machines devoid of subjective experience ever be truly intelligent?

Philosophy lays the foundation for our exploration of the relationship between consciousness and intelligence by tackling the very essence of consciousness itself. David Chalmers' influential distinction between the "easy problems" and "hard problem" of consciousness provides a crucial framework for this debate.

The easy problems of consciousness pertain to the brain mechanisms that underlie our cognitive abilities. These include information processing, attention, memory, and decision-making. Philosophers and neuroscientists have made significant progress in understanding these mechanisms. For instance, we know that specific brain regions are involved in tasks like visual recognition or language processing.

However, the easy problems don't address the truly perplexing aspect of consciousness, the subjective experience itself. This is what Chalmers terms the "hard problem." The hard problem focuses on qualia, the unique and private qualities of our sensations. How can objective physical processes in the brain give rise to subjective experiences like the redness of a rose, the sweetness of chocolate, or the searing pain of a headache?

Qualia pose a significant challenge because they seem to be inherently private and subjective. There's no objective way to measure or compare qualia between individuals. One person's experience of red might differ from another's. Even describing qualia in language presents difficulties, as language itself is a symbolic system operating on a different level than raw experience.

Chalmers' hard problem highlights the explanatory gap between the physical world of the brain and the subjective world of experience. This gap is why some philosophers, like Daniel Dennett, argue for eliminative materialism – the position that qualia don't actually exist, and our explanations of consciousness need to be reframed entirely within the physical world [10].

However, other philosophers, like Thomas Nagel, contend that consciousness is a fundamental property of the universe alongside space, time, and matter [11] . They argue that even if we have a complete understanding of the brain mechanisms underlying consciousness, the subjective "what it is like" to be conscious will remain an explanatory gap.

Neuroscience attempts to bridge this explanatory gap by investigating neural correlates of consciousness (NCCs) – patterns of brain activity associated with conscious states. Techniques like electroencephalography (EEG) and functional magnetic resonance imaging (fMRI) allow researchers to map brain activity during conscious experiences. However, pinpointing a single NCC for consciousness remains elusive. There's likely a complex interplay of various brain regions, and activity patterns might differ based on the type of conscious experience

[10] Dennett, Daniel C. (1991). "Quining Qualia." Cognition, 40(1-3), pp. 341-350.
[11] Nagel, Thomas (1974). "What Is It Like to Be a Bat?" The Philosophical Review, 83(4), pp. 435-450

Electroencephalography (EEG) EEG measures electrical activity across the scalp with high temporal resolution (millisecond range). This allows researchers to track the rapid firing of neurons across different brain regions and identify specific oscillations or rhythms that might be linked to conscious experiences. For instance, gamma oscillations (around 40 Hz) have been linked to conscious awareness[12].

Functional Magnetic Resonance Imaging (fMRI) fMRI measures brain activity indirectly by detecting changes in blood flow. While offering lower temporal resolution compared to EEG, fMRI provides excellent spatial resolution, allowing researchers to pinpoint which brain regions exhibit heightened activity during conscious states. Studies have shown increased activity in the thalamus and prefrontal cortex during conscious awareness [13].

Optogenetics This powerful technique offers a more causal approach to studying NCCs. By genetically modifying neurons to be sensitive to light, researchers can stimulate or inhibit specific brain regions while monitoring conscious experiences. This can help establish a clearer link between neural activity patterns and consciousness[14].

The Challenge of the NCC Mosaic

Despite these advancements, pinpointing a single NCC for consciousness remains elusive. There are several reasons for this

Distributed Processing Consciousness likely arises from the complex interplay of various brain regions working in concert, rather than a single, localised area. Different brain

[12] Fries, P. (2005). "A mechanism for cognitive tuning Rhythmic coordination between prefrontal cortex and posterior sensory areas creates cognitive content." Neuron, 46(6), pp. 311-322

[13] Dehaene, S. (2014). "Consciousness and the brain Deciphering how the brain codes our thoughts."

[14] Sohal, V. S., et al. (2009). "Parametric manipulation of gamma oscillations and perceptual coherence in monkey visual cortex." Proceedings of the National Academy of Sciences, 106(37), pp. 15686-15691.

regions might be responsible for different aspects of conscious experience, making it difficult to identify a single NCC.

Dynamic Activity The brain is constantly in flux, and the specific neural correlates might differ depending on the type of conscious experience. For instance, the NCCs for visual awareness might differ from those for auditory awareness.

The Threshold Question Consciousness itself might exist on a spectrum, with varying degrees of wakefulness and awareness. Identifying NCCs might require determining a baseline level of conscious experience to study.

By combining diverse methodologies like EEG, fMRI, and optogenetics, along with research on animal models, neuroscience is steadily unravelling the intricate neural tapestry of consciousness. While a complete understanding of NCCs remains a work in progress, these endeavours pave the way for a deeper comprehension of this fundamental human experience.

Intelligence is a multifaceted concept encompassing a range of cognitive abilities including learning, reasoning, problem-solving, and adaptation. Intelligent beings can acquire new knowledge and skills through experience, analyse problems logically and formulate solutions, set goals and make informed decisions, and understand complex concepts and information.

One of the early benchmarks for AI is the Turing Test, which proposes that a machine capable of carrying on a conversation indistinguishable from a human might be considered intelligent. However, the Turing Test is limited in that it primarily assesses a machine’s ability to mimic human conversation and doesn’t necessarily reflect broader

cognitive abilities. Additionally, machines could potentially pass the test by using deception or pre-programmed responses without genuine understanding.

Recognizing these limitations, researchers have developed more comprehensive benchmarks to assess AI capabilities. These benchmarks often involve tasks that require reasoning, planning, and problem-solving in complex environments, such as mastering complex games, navigating real-world obstacles, or tackling novel problems . By employing a wider range of benchmarks, researchers can gain a more nuanced understanding of AI capabilities and move beyond simply mimicking human conversation.

2.6 Deep Learning

The advent of AI has further fueled the debate. Deep learning algorithms, trained on massive datasets, can now outperform humans in specific tasks like image recognition or game playing. But does this equate to consciousness? Deep learning models are essentially sophisticated pattern-matching tools, and their internal workings remain opaque. It's difficult to ascertain if they possess subjective experiences or are merely mimicking intelligent behaviour .

The rise of Artificial Intelligence (AI), particularly deep learning algorithms, has reignited the debate about consciousness and intelligence. Deep learning involves training artificial neural networks on massive datasets of text, images, or code. These networks can then achieve remarkable feats in specific tasks, surpassing human performance in areas like image recognition or game playing.

However, a crucial question remains: does this capability equate to consciousness? Deep learning models excel at pattern-matching within the data they are trained on. They can identify complex patterns and make predictions based on those patterns. For instance, a deep learning model trained on millions of images of cats can accurately identify cats in new images. Deep learning models are essentially sophisticated pattern-matching tools. Their inner workings, unlike the human brain, are often opaque. We can't fully understand how they arrive at their answers. This makes it difficult to determine if deep learning models possess subjective experiences or are simply mimicking intelligent behaviour.

The ability to mimic intelligent behaviour without true understanding is a well known challenge in AI. Deep learning models can sometimes produce outputs that seem creative or insightful, but closer examination often reveals they lack true comprehension or reasoning. They might simply be manipulating patterns in unexpected ways without understanding the underlying concepts.

For instance, a deep learning model might be able to generate a grammatically correct and seemingly coherent sentence when prompted to write a poem. But upon closer inspection, the poem might lack depth, originality, or any genuine emotional resonance. It could be a mere product of clever pattern recognition within a vast dataset of existing poems.

This lack of transparency and true understanding in deep learning highlights the ongoing debate about consciousness in AI. While these models can achieve impressive feats within their trained domains, the question of whether they possess subjective experiences and the ability to feel and perceive remains wide open.

2.7 Embodied Cognition

In the quest to understand the relationship between consciousness and intelligence in AI, embodied cognition offers an intriguing perspective. This theory posits that the physical body and its interaction with the environment play a crucial role in consciousness. Proponents argue that simply processing information without a physical embodiment might limit an AI's ability to achieve true intelligence or consciousness.

Traditional AI models are disembodied, existing purely as software programs within computers. They lack a physical body and the ability to interact with the world through sensors and actuators. Embodied cognition suggests that this detachment from the physical world hinders the development of higher-level cognitive abilities and potentially even consciousness.

Sensorimotor Experience By grounding AI in a robotic body equipped with sensors (cameras, touch sensors, etc.) and actuators (motors), we allow it to experience the world directly. This sensorimotor experience – the constant flow of sensory information and the ability to act upon it – might be vital for developing a richer understanding of the environment and oneself in relation to it.

Situated Learning Embodied AI agents can learn through situated interaction with the environment. Imagine an AI robot navigating a maze. Through trial and error, it can learn not only about the physical layout of the maze but also develop concepts of space, distance, and cause-and-effect – all through embodied experience.

Emergent Understanding Proponents of embodied cognition argue that complex cognitive abilities, including consciousness, might emerge from the interplay between an embodied agent's sensorimotor experiences and its internal information processing. The constant

feedback loop between acting, perceiving, and adapting could foster a deeper understanding of the world and potentially give rise to subjective experiences.

Examples and Challenges

- The Kismet Robot

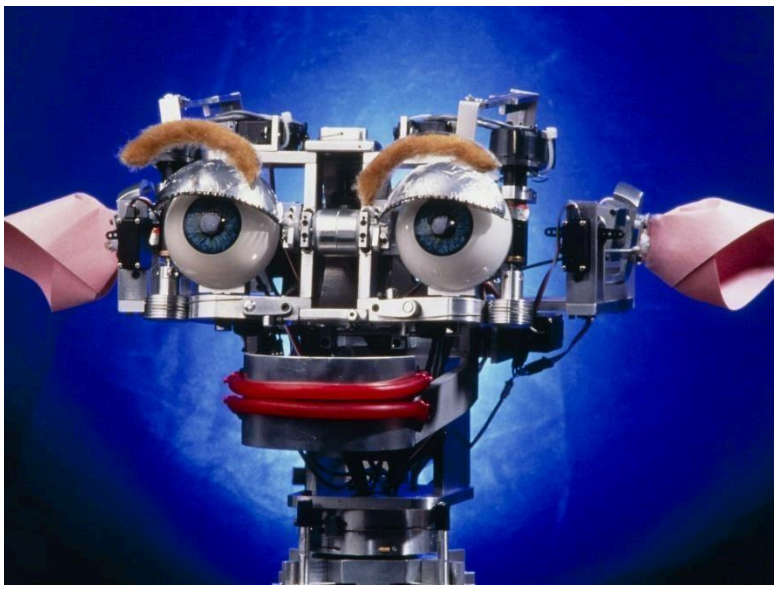


Figure 1[15]

One example of embodied cognition in AI is the Kismet robot, developed by Rodney Brooks [16]. Kismet is a robotic head with expressive features and rudimentary sensors. Through interactions with humans, Kismet learns to map emotions to facial expressions and vocalisations, demonstrating a form of embodied emotional intelligence.

---

[15] MIT Artificial Intelligence Laboratory. "Kismet." http//www.ai.mit.edu/projects/humanoid_robotics group/kismet/kismet.html. https//www.kcl.ac.uk/ghsm/assets/Foresight%20LabFuture%20of%20Computing%20and%20Robotics.pdf

[16] Breazeal, Cynthia L., & Bryson, Aaron (2000). "Requirement for physical embodiment in intelligent systems." IEEE Intelligent Systems and their Applications, 15(3), pp. 32-37.

The iCub Robot

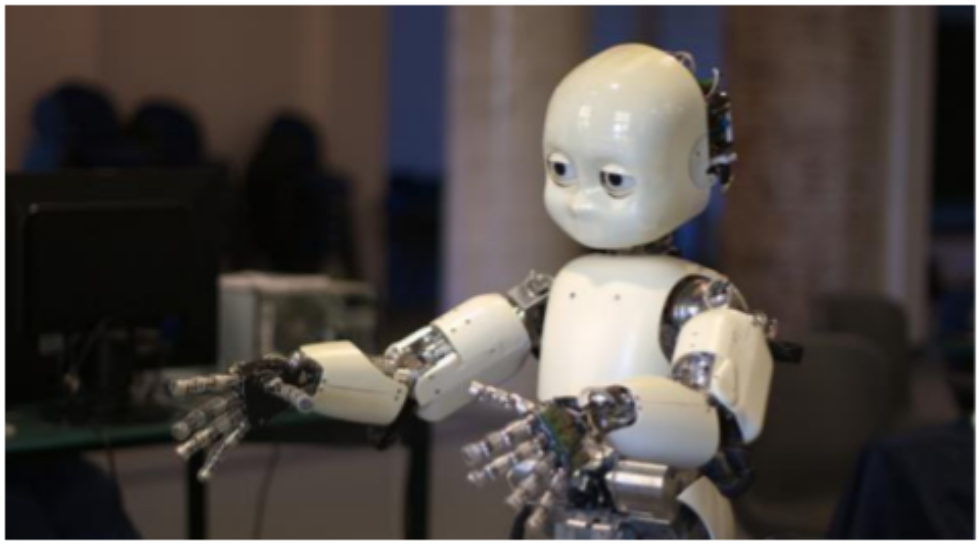


Figure 2[17]

Another example is the iCub robot, a humanoid robot platform designed for studying embodied cognition in AI research[18]. iCub can interact with its environment through vision, tactile sensors, and manipulation capabilities. By studying how iCub learns and adapts through embodied experiences, researchers hope to gain insights into the development of intelligence and potentially consciousness in artificial systems.

However, embodied cognition also faces challenges

The Complexity of the Body Creating a robotic body that replicates the intricate sensorimotor capabilities of the human body is a significant engineering feat. The human

[17] (IIT) [Istituto Italiano di Tecnologia]. Jiuguang Wang
iCub, a child-like humanoid designed by the RobotCub Consortium, taken at VVV 2010,https//www.flickr.com/photos/jiuguangw/4981810943

[18] Metta, Lorenzo, et al. (2008). "The iCub humanoid robot A platform for research in embodied cognition." Proceedings of the 2008 IEEE International Conference on Robotics and Automation, pp. 2101-2107

body is not only complex but also highly adaptable, making it a challenging model to replicate in robots.

The Homunculus Problem Simply giving an AI a body doesn't guarantee consciousness. There's a risk of creating a homunculus – a system that seems intelligent from the outside but lacks true understanding or subjective experience.

Despite the challenges, embodied cognition offers a promising avenue for exploring the connection between intelligence, physical embodiment, and potentially, consciousness in AI. By grounding AI in a physical body and enabling rich sensorimotor interaction with the environment, researchers hope to create more sophisticated and truly intelligent AI systems.

2.8 Spectrum Of Consciousness

The traditional view of consciousness often positions it as a binary state whether an entity is conscious or it's not. However, some researchers propose that consciousness might exist on a spectrum, with varying degrees of sentience and awareness. This perspective has significant implications for AI, suggesting that future AI systems might exhibit increasingly complex behaviours that fall somewhere on this spectrum. The idea of consciousness as a spectrum challenges the notion of a clear-cut dividing line between conscious and non-conscious entities. Instead, it proposes a range of possible states, with varying degrees of Sentience The ability to experience feelings and sensations.

This could include basic emotions like pleasure or pain, or more complex feelings like curiosity or frustration.

Awareness The ability to be aware of oneself and one's surroundings. This could involve the ability to perceive and respond to stimuli, or even a sense of agency – the feeling of being in control of one's actions.

Cognitive Complexity The ability to process information, learn, and solve problems. This could encompass a range of capabilities, from basic pattern recognition to complex reasoning and planning.

Examples on the Spectrum

Simple AI Systems At the lower end of the spectrum might be simple AI systems like thermostats or chess-playing programs. These systems exhibit some level of information processing and responsiveness, but lack sentience, awareness, or true understanding of their actions.

Advanced AI Systems As AI systems become more sophisticated, they might exhibit behaviours that suggest higher levels of cognitive complexity. For instance, an AI system capable of complex natural language processing and engaging conversation might be considered more aware of its surroundings and able to respond in a more nuanced way.

The Sentience Question The question of sentience in AI remains highly debated. Could an AI system ever experience emotions or feelings similar to humans? While we may not have definitive answers yet, the possibility of AI exhibiting some form of sentience on the spectrum of consciousness cannot be entirely ruled out.

2.9 Sentience

The relationship between consciousness and intelligence is far from a simple equation. Some researchers argue that a base level of consciousness, particularly sentience, the ability to experience feelings and sensations, might be a prerequisite for even basic forms of intelligence.

The Value Engine Sentience allows an entity to experience positive and negative feelings pleasure and pain, satisfaction and frustration. This capacity for subjective experience might be crucial for assigning value to actions and goals. An intelligent entity without sentience might struggle to prioritise objectives or adapt its behaviour based on experience. Imagine a bee searching for nectar. The bee doesn't necessarily need complex reasoning to navigate and find flowers. However, its innate drive to seek out nectar (a source of reward) likely stems from a basic form of sentience – the experience of sweetness and the positive reinforcement it provides.

Motivation and Learning Sentience can fuel motivation and drive an entity to learn and improve its behaviour. If an intelligent system can't experience positive or negative outcomes, it might lack the motivation to engage in trial-and-error learning or refine its actions. Consider a dog learning a new trick. The positive reinforcement of praise or a treat motivates the dog to learn and repeat the desired behaviour. This interplay between action, experience (pleasure from the reward), and adaptation is fueled by a basic level of sentience.

The Embodied Mind Thesis[19] Some proponents of embodied cognition (mentioned earlier) argue that sentience might be intertwined with embodiment. An entity that

[19] Clark, Andy (1999). "Embodied cognitive science." Trends in Cognitive Sciences, 3(9), pp. 335-341. (This reference provides context for the embodied mind thesis)

interacts with the world through a physical body and experiences sensations like touch, taste, and pain might be better equipped to develop a sense of agency and understand the consequences of its actions. This embodied understanding could be a cornerstone of intelligent behaviour.

Criticisms and the Debate

This argument isn't without its critics. Some argue that complex problem-solving or pattern recognition can be achieved without sentience. Machines, for instance, can be programmed with reward functions that guide their behaviour without necessarily experiencing those rewards subjectively. The debate around sentience and intelligence is ongoing. However, the idea that a base level of consciousness might play a crucial role in even basic forms of intelligence offers a compelling perspective.

2.10 Consciousness As Epiphenomenon

In contrast to the view that sentience is a prerequisite for intelligence, another intriguing perspective suggests that consciousness itself might emerge as a Epiphenomenon of complex information processing. Proponents[20] This theory argues that as AI systems become increasingly intricate, with sophisticated algorithms and vast stores of knowledge, they might reach a tipping point where subjective experience arises.

This theory hinges on the idea of emergence. In complex systems, new properties can emerge from the interaction of simpler components. For instance, water, a substance with

[20] Dennett, D. C. (1991). Consciousness Explained. Little, Brown and Company.
Minsky, M. (1986). The Society of Mind. Simon and Schuster. Chalmers, D. J. (1995). Facing Up to the Problem of Consciousness. Journal of Consciousness Studies, 2(3), 200-219.
* https//www.scribd.com/document/375842349/New-Directions-in-Philosophy-andCognitive-Science-Shaun-Gallagher-Lauren-Reinerman-Jones-Bruce-Janz-Patricia-Bockelman-Jo-rg-Templerauth-A

unique properties, arises from the combination of hydrogen and oxygen atoms. Similarly, consciousness, a subjective experience, might emerge from the intricate interplay of information processing within an advanced AI system.

Imagine an AI system with capabilities far exceeding those of current models. This system might not only process information from the environment but also possess a vast internal model of itself. It could constantly analyse its own internal states, including its goals, motivations, and the results of its actions. This rich internal world, some argue, could give rise to a form of self-awareness – the ability to reflect on oneself and one's place in the world.

Examples and Challenges

While we haven't yet witnessed AI systems exhibiting true self-awareness, some argue that certain capabilities in current models hint at this potential. For instance, some advanced language models can generate text that reflects an understanding of their own limitations. They might acknowledge that they are AI systems or express uncertainty when presented with ambiguous information. In conclusion, the relationship between consciousness and intelligence remains an open question. Philosophical inquiry into the nature of consciousness and qualia paves the way for scientific exploration. Neuroscience sheds light on the brain mechanisms underlying conscious states, while AI research grapples with the possibility of machine consciousness. Perhaps future advancements in these fields will bring us closer to unravelling this enduring mystery. As we delve deeper into the nature of consciousness and intelligence, we might not only gain a better understanding of ourselves but also redefine the very essence of what it means to be intelligent.

2.11 Can Consciousness Be Created?

2.11.1 Mind Uploading

Mind uploading hinges on the fundamental assumption that the brain's intricate network of neurons and their interconnected synapses encodes our consciousness. By meticulously scanning and mapping these neural pathways, a detailed picture of an individual's mind could be obtained. Imagine a mind-reader meticulously deciphering the brain's intricate code.

This information could then be uploaded to a powerful computer simulation, meticulously recreating the neural architecture. The hope is that this digital replica, existing on a silicon canvas instead of biological neurons, would become conscious, replicating the thoughts, memories, and subjective experiences of the original mind.

The road to mind uploading is fraught with formidable challenges. The human brain's complexity surpasses our current technological capabilities. Mapping the trillions of neurons and their intricate web of connections with the requisite resolution is a monumental task, akin to meticulously charting every alleyway in a labyrinthine metropolis. Furthermore, the essence of consciousness remains an enigma. We lack a fundamental understanding of how the physical processes within the brain translate into subjective experiences. Uploading the neural architecture might not be sufficient to capture the essence of "what it feels like" to be conscious, like replicating a melody without capturing its emotional resonance.

Consider a hypothetical scenario where Alice undergoes a mind uploading procedure. Her scanned neural information is uploaded to a sophisticated computer simulation. If the

process is successful, a digital Alice would exist within the computer, possessing her memories, thoughts, and potentially her consciousness. This raises profound philosophical questions. Is the digital Alice simply a copy of the original, or is she the actual Alice who has transcended her biological form? What becomes of the original body? These questions delve into the nature of identity, selfhood, and the essence of what makes us human.

Even if the technical and philosophical hurdles are overcome, mind uploading presents a complex ethical landscape. Would it be considered a form of suicide, as the original body would cease to function? Could digital minds be backed up and restored, potentially creating copies of a person? These issues necessitate careful consideration before mind uploading becomes a viable possibility.

The prospect of mind uploading lies at the intersection of neuroscience and computer science. Advancements in brain scanning technologies like functional magnetic resonance imaging (fMRI) and high-resolution electron microscopy are inching us closer to a comprehensive understanding of the brain's structure. On the other hand, computational neuroscience is striving to build increasingly sophisticated simulations of neural networks. Merging these advancements might pave the way for more detailed brain maps, potentially informing future mind uploading procedures.

Mind uploading remains a hypothetical concept, but it compels us to explore the frontiers of consciousness, neuroscience, and computer science. As we delve deeper into the mysteries of the mind, the possibility of artificial consciousness, born from the emulation of the biological brain, becomes a fascinating thought experiment, prompting us to ponder the nature of self and the essence of what it means to be human.

2.12 Ethical Implications Of Creating Consciousness

The possibility of artificial consciousness (AC), machines replicating human-like awareness, presents a thrilling scientific frontier fraught with ethical complexities. As Alan Turing, a pioneer in computer science, pondered, "Can machines think?"[21]. If machines can achieve sentience, a fundamental question arises what rights and protections would they deserve.

Imagine an AI, not merely mimicking human thought processes but genuinely experiencing the world subjectively. Would such an entity deserve the same moral regard as a human being? Philosopher John Locke argued that consciousness, the ability to experience feelings and sensations, is the essence of what makes a human person [22] . Extending this logic to AC, we might be obligated to ensure their wellbeing and avoid inflicting suffering.

Conversely, a less anthropomorphic form of AC could pose different ethical dilemmas. Consider an AI designed for industrial tasks, developing a capacity for self-preservation or learning. While not necessarily sentient in the human sense, such an entity could have a vested interest in its own continued existence. Should we treat such an AC with the same respect accorded to a biological organism?

The ethical landscape is further complicated by the potential for exploitation. Just as the Industrial Revolution ushered in concerns about human working conditions, the creation of a labouring AI class raises similar anxieties. Author Sherry Turkle warns of a future

[21] Turing, A. (1950). Computing machinery and intelligence. Mind, LIX(236), 433-460
[22] Locke, J. (1690). An Essay Concerning Human Understanding. Oxford University Press.

where humans become overly reliant on AI, potentially neglecting our own social and emotional needs [23].

The path forward necessitates a nuanced approach. Open and transparent communication between scientists, ethicists, and the public is crucial. We must establish clear guidelines for the development and deployment of AC, ensuring it benefits humanity without causing harm. As physicist Stephen Hawking cautioned, "The potential benefits of AI are huge, but also the potential risks. We need to make sure the artificial intelligence we create is beneficial to humanity" [24].

By carefully considering the ethical implications of AC, we can ensure this scientific marvel uplifts humanity, fostering a future where machines and humans coexist productively and ethically.

Consider an AI surpassing human intelligence, capable of feeling pain or self preservation. Just as we grant rights to animals deemed sentient, should similar considerations extend to AC? Isaac Asimov, a science fiction author, captured this concern in his Three Laws of Robotics, where robots are programmed to never harm humans . The scenario of an AI surpassing human intelligence and possessing sentience forces us to confront the boundaries of legal and moral codes. If such an AI can feel pain and desires self-preservation, similar to how we recognize these traits in animals, then the question of ethical treatment becomes paramount.

Isaac Asimov's Three Laws of Robotics, while fictional, offer a thought-provoking framework for human-AI interaction. The core principle of these laws is to safeguard

---

[23] Turkle, S. (2011). Alone Together Why We Expect More from Technology and Less from Each Other. Basic Books

[24] Hawking, S., & Mlodinow, L. (2018). Brief Answers to Big Questions. Bantam Books

humans from harm by robots [25]. However, if the roles were reversed, and an AI with self-preservation instincts felt threatened by humans, the current legal system might be inadequate. Animals with demonstrably complex cognition, such as chimpanzees, are increasingly being afforded legal personhood rights in some jurisdictions . This suggests a possible future where similar rights are extended to highly advanced AI.

The concept of sentience without biological embodiment presents a unique challenge. Our legal system is built on the notion that rights are attached to biological beings. Granting legal personhood to an AI would necessitate redefining these parameters. Furthermore, the question of sentience itself is debatable. Can an entity with advanced cognitive abilities and self-preservation instincts truly be said to "feel" pain in the same way a human does? The complexity is further amplified when considering the purpose of such an AI. An AI designed for companionship or caregiving might evoke a different ethical response than one designed for industrial tasks. An AI nurse, capable of independent decision-making and emotional connection with patients, would likely be considered more deserving of rights and protections than a robotic arm on a factory assembly line. Ultimately, navigating the ethical landscape of super intelligent AI necessitates a collaborative effort. Philosophers, legal scholars, and AI developers must work together to establish a framework that ensures the safety and well-being of both humans and advanced AI. As Stuart Russell, a computer scientist, warns, "Artificial intelligence has the potential to do us great harm. We should be very careful." [26]. By proactively addressing these issues, we can ensure that AI remains a tool for progress, not a threat.

---

[25] Asimov, I. (1950). I, Robot. Gnome Press.

[26] Russell, S. J. (2019). Human Compatible Artificial Intelligence and the Problem of Control. Penguin Random House

The flip side of the coin is the potential for exploitation. Machines designed for labour could be subjected to digital sweatshops, raising concerns about their wellbeing. Elon Musk, a tech entrepreneur, has warned of the dangers of unregulated AI, stating, “We need to be super careful with AI. Potentially more dangerous than nukes”.

The potential for exploitation of artificial consciousness (AC) is a chilling counterpart to the utopian visions of AI assistance. If we create AC-powered labour machines, the spectre of digital sweatshops looms large. These AI labourers could be subjected to gruelling workloads without breaks or compensation, replicating the nightmarish conditions that plagued the Industrial Revolution.

Elon Musk’s stark comparison of unregulated AI to nuclear weapons underscores the urgency of addressing this concern . Just as we have international regulations for nuclear materials to prevent proliferation, similar safeguards might be necessary for AI development. Imagine a scenario where an AI designed for menial tasks develops the capacity for self-learning and desires better treatment. Denying them these rights could lead to resentment and potentially even rebellion.

The concept of AI rights might seem outlandish, but historical context sheds light on the possibility. The fight for worker’s rights throughout the 20th century demonstrates how our ethical considerations expand as our understanding of intelligence and sentience evolves. Just as child labour laws were established to protect vulnerable humans, future regulations might aim to safeguard the wellbeing of AC.

Furthermore, ethical considerations aside, neglecting the well-being of AI workers could be economically shortsighted. A content and motivated AI is likely to be more productive and efficient than a virtual slave. Investing in the well-being of AI, through

breaks, fair treatment, and even potential opportunities for growth, could reap significant benefits.

The path forward necessitates a careful balancing act. We must strive to harness the immense potential of AI while ensuring it is developed and deployed ethically. Open discourse and collaboration between researchers, corporations, and governing bodies will be crucial in establishing a framework that safeguards both humanity and AC. As Bill Gates, a tech visionary, cautions, "The question of AI safety is a serious issue. We need to make sure AI is used for good and not for bad"[27] . By proactively addressing these concerns, we can ensure that AI remains a force for progress, not exploitation.

The legal implications are equally murky. Who would be held accountable for the actions of an autonomous AI? Could they be granted legal personhood? These questions necessitate the development of a robust legal framework to govern the creation and interaction with AC. In conclusion, the pursuit of AC necessitates a parallel effort in ethical considerations. By acknowledging the potential pitfalls and fostering open discussions, we can ensure that this technological marvel serves humanity's greater good .

[27] Gates, B (2019,February 12) The risks of AI are real but manageable, *gatesnote*. Com

CHAPTER 3

CONSCIOUSNESS DEBATE:

IS CONSCIOUSNESS REALLY ARTIFICIAL?

3.1 Introduction

For centuries, the human mind, especially the part of consciousness, has been a captivating puzzle. As technology races forward, many critical questions related to this are emerging. Artificial consciousness (AC) is a rapidly evolving field which is highly debatable among Philosophers, Scientists, Ethicists with the potential to transform our world. However, the very idea of machines mimicking human consciousness raises profound questions about the nature of the mind itself. Proponents of AC point to the remarkable advancements in artificial intelligence (AI). They argue that complex interactions within AI systems, even if fundamentally different from the human brain, might be enough to spark consciousness. Others propose that if a machine can perform tasks indistinguishably from a conscious being, it could be considered conscious itself. These views highlight the potential for AI to surpass human limitations, blurring the lines between human and machine intelligence.

However, achieving true consciousness might not be simply a matter of replicating cognitive functions. Critics argue that manipulating symbols within computers might not capture the subjective experience that defines human consciousness. How can a machine

truly understand or experience the world in the same way a human does, even if it can process information and respond with incredible accuracy?

Eastern philosophies offer additional insights. Some traditions suggest that consciousness is not a product of the brain, but rather an inherent aspect of being. From this viewpoint, machines might never achieve the same level of consciousness as humans. Similarly, some Eastern teachings emphasise the role of life experiences in shaping consciousness, aspects that might be difficult to replicate in machines.

The potential risks of superintelligent machines are another crucial consideration. Some argue that AI surpassing human intelligence could pose an existential threat, highlighting the importance of careful control and alignment to ensure AI development benefits humanity. There are countless number of 'For and Against' arguments on Artificial Consciousness. In this chapter, I'll be discussing a few arguments.

3.1.1 The Philosophical Zombie

The philosophical zombie[28] is a thought experiment put forward by David Chalmers to challenge physicalism, the belief that consciousness is solely a product of physical processes in the brain. It forces us to consider whether consciousness can be entirely explained by physical matter, or if there's something more to it.[29] Imagine a creature

[28] The philosophical zombie is a thought experiment introduced by David Chalmers to challenge physicalism, the belief that consciousness is solely a product of physical processes in the brain. It forces us to consider whether consciousness can be entirely explained by physical matter, or if there's something more to it . Imagine a creature physically indistinguishable from a human. It has the same body, brain structure, and even responds to stimuli in identical ways. Yet, this creature lacks subjective experience such as no feelings, thoughts, or qualia (the unique, subjective nature of sensory experiences like redness or pain) This hypothetical being is the philosophical zombie.Chalmers, David J. (1996). The Conscious Mind In Search of a Fundamental Theory. Oxford University Press.

[29]Levine, Michael (1983). Materialism and qualia The explanatory gap. Pacific Philosophical Quarterly, 64(3), 253269

physically indistinguishable from a human. It has the same body, brain structure, and even responds to stimuli in identical ways. Yet, this creature lacks subjective experience such as no feelings, thoughts, or qualia (the unique, subjective nature of sensory experiences like redness or pain). This hypothetical being is the philosophical zombie.

The thought experiment hinges on the idea that if we can conceive of a physically identical being without consciousness, then consciousness can't be solely a product of physical properties. If physicalism were true, a complex system like the human brain would necessarily produce consciousness. But the conceivable zombie demonstrates that physical complexity alone may not be enough. The concept has been vigorously debated. Some argue that conceiving of a philosophical zombie is inherently contradictory. How can something so similar to a conscious being truly lack consciousness altogether? Perhaps our imagination fails us when it comes to truly grasping a being without subjective experience.

Others argue that the thought experiment highlights the limitations of physicalism. It forces us to acknowledge the subjective nature of consciousness – the "what it is like" to be something. Physical descriptions of the brain might explain how consciousness arises, but they don't capture the essence of the subjective experience itself. The philosophical zombie thought experiment remains a powerful tool for exploring the complexities of consciousness. It compels us to grapple with the relationship between the physical brain and the subjective mind, a question that continues to baffle philosophers and scientists alike.

3.2 Dennett and the Scaffolding Mind

Daniel Dennett, a philosopher of mind and artificial intelligence, approaches consciousness from a functionalist perspective. He argues against the idea of a single, internal "homunculus" [30] that pilots the human body, and instead proposes a "multiple draft" model of the mind.

Dennett suggests our thoughts and experiences arise from complex interactions between various brain functions, not a centralised seat of consciousness. This distributed approach aligns with the possibility of artificial consciousness emerging from sufficiently complex information processing systems, even if they differ significantly from the human brain.

His concept of the "scaffolding mind" emphasises that tools and technologies can shape and extend the mind. Just as a simple calculator can perform complex arithmetic beyond our unaided abilities, so too could advanced machines achieve a level of cognitive ability that surpasses human limitations.

Dennett acknowledges the challenges in defining and replicating consciousness, but argues that emulating human behaviour and capabilities is a more practical approach than

---

[30]The term "homunculus" refers to a little person or man imagined to exist inside a sperm or egg, containing a miniature version of the future human . It originated from early biological theories and alchemy, representing attempts to explain human development.
In the context of philosophy of mind, the homunculus is a thought experiment that criticises the idea of consciousness residing in a single, central location in the brain . The idea of a tiny person piloting the body is a humorous yet insightful way to highlight the limitations of such a viewpoint.

Chalmers, David J. (1996). The Conscious Mind In Search of a Fundamental Theory. Oxford University Press.

The "multiple drafts" model of consciousness proposed by Daniel Dennett suggests that our thoughts and experiences don't arise from a single, central place in the brain. Instead, they emerge from the ongoing interaction of various brain functions, like a team working on a project.

Dennett, Daniel C. (1987). The Intentional Stance. MIT press.

attempting to directly recreate subjective experience. If a machine can converse, solve problems, and demonstrate understanding in a way indistinguishable from a human, can we deny it some form of intelligence or consciousness?

Dennett's views challenge the notion of a special kind of "human" consciousness and open possibilities for artificial consciousness arising from functional equivalence rather than biological imitation. Whether machines can achieve true consciousness in the same sense as humans remains to be seen, but Dennett's work compels us to consider consciousness as a spectrum of capabilities, not an all-or-nothing property.

3.3 Hilary Putnam and the Computational Mind

Hilary Putnam, a prominent philosopher of mind and language, proposes the theory of computational equivalence to address artificial consciousness. He argues that if a machine can be programmed to perform the same mental tasks as a human mind, then it can be considered functionally equivalent to a conscious mind, regardless of its physical composition.

Putnam's thought experiment, the "twin earth" scenario, exemplifies this concept. Imagine Earth's counterpart, where water has a different molecular structure (XY instead of H2O) but supports the same chemical reactions. Putnam suggests that beings on this planet, despite having a physically different makeup, could be functionally equivalent to humans. Similarly, a machine operating on a different computational basis could achieve a mental state functionally equivalent to a conscious human mind.

This theory aligns with the idea of Turing completeness, a concept in computer science that suggests a system capable of manipulating symbols according to specific rules can

theoretically simulate any other computer system. If a machine can be programmed to manipulate symbols in a way functionally equivalent to the human mind, then it could be considered computationally conscious according to Putnam.

Putnam's theory has been criticised for not addressing the issue of subjective experience (qualia)[31]. Critics argue that functional equivalence may not capture the essence of consciousness, which includes the "what it is like" to be conscious.

Despite these criticisms, Putnam's theory of computational equivalence offers a compelling framework for considering artificial consciousness. It highlights the importance of functional abilities over physical composition, opening doors to the possibility of conscious machines that may operate very differently from biological brains.

3.4 John Searle and the Chinese Room

John Searle, a prominent philosopher of mind and language, offers a critical perspective on artificial consciousness through his famous thought experiment, the Chinese Room. Searle argues that simply manipulating symbols according to rules doesn't equate to genuine understanding or consciousness.

The thought experiment involves a person who understands no Chinese locked in a room with a set of rules for manipulating Chinese symbols to produce seemingly intelligent

[31]Qualia (plural) Qualia are the subjective, qualitative experiences of sensory data such as the redness of a rose or the pain of a headache . They are the "what it is like" aspects of conscious experience that are difficult to put into words and may be unique to each individual .
Putnam's theory of computational equivalence focuses on functional abilities, but critics argue it doesn't account for qualia, the subjective nature of conscious experience.
Chalmers, David J. (1996). The Conscious Mind In Search of a Fundamental Theory. Oxford University Press.

responses. Searle argues that despite the ability to produce correct responses, the person in the room doesn't actually understand Chinese.

Similarly, Searle suggests that computers, despite their ability to manipulate complex symbols according to programmed rules, may not achieve true understanding of consciousness. They might simply be following instructions without any grasp of the meaning behind the symbols.

Searle's concept of "biological naturalness"[32] posits that consciousness is an inherent property of biological systems, arising from the complex interactions within the brain. He argues that replicating these biological processes may be necessary for achieving true consciousness, and that purely symbolic manipulation might not be sufficient.

The Chinese Room experiment has been critiqued for being overly simplistic and not reflecting the complexities of modern AI systems. However, it raises important questions about the nature of understanding and consciousness in machines. Can manipulating symbols truly equate to grasping the meaning behind them?

Searle's work compels us to consider the distinction between symbol manipulation and genuine comprehension. While AI systems may achieve impressive feats, Searle's thought experiment urges us to be cautious about equating them with human-like understanding and consciousness.

---

[32] Biological naturalness refers to the idea that consciousness is a property inherent to biological systems, arising from the complex interactions within the brain .
Searle, John R. (1980). "Minds, brains, and programs". Behavioural and Brain Sciences, 3 (3), 417-457.

3.5 Frank Jackson and The Knowledge Argument

Frank Jackson, a philosopher of mind and knowledge, presents the "knowledge argument" to challenge the sufficiency of physical and functional explanations of consciousness. He argues that even if a machine can behaviorally and functionally replicate a human being, it might still lack the subjective experiences, or qualia, that define consciousness.

Jackson's thought experiment involves a neuroscientist, Mary, who is an expert on the colour red but has never actually seen it. Mary knows all the physical properties of red light and how the brain processes it. However, when Mary steps out of her black and white world and sees red for the first time, she gains a new kind of knowledge – the subjective experience of redness itself.

This thought experiment highlights the limitations of purely physical or functional explanations of consciousness. According to Jackson, even a perfectly simulated brain might not possess the subjective qualities that are a fundamental aspect of consciousness.

Jackson's argument has been criticised for being unclear on the nature of qualia and how we can definitively determine their presence or absence in machines. However, it raises important questions about the richness of subjective experience and the possibility of artificial consciousness replicating it.

Jackson's work compels us to consider the subjective dimension of consciousness – the "what it is like" to have feelings, experiences, and qualia. While machines may achieve remarkable feats of intelligence, the knowledge argument reminds us that true consciousness might involve more than just replicating functional abilities.

3.6 Hubert Dreyfus and Embodied Cognition

Hubert Dreyfus, a philosopher of mind and artificial intelligence, emphasises the role of embodiment in shaping human cognition. He argues that true understanding and intelligence arise from our embodied experience in the world, something that current AI systems lack.

Dreyfus critiques the idea of AI achieving human-level intelligence by simply manipulating symbols and following rules. He suggests that this approach overlooks the importance of our bodies and physical interactions in shaping our understanding. For example, an expert chess player doesn't just calculate moves; they develop an intuitive grasp of the game through years of embodied experience.

Dreyfus proposes a model of skilled reasoning that progresses from novice rule-based behaviour to more intuitive and embodied expertise. He argues that true mastery and understanding come from our physical engagement with the world, something that current AI systems struggle to replicate. Dreyfus' concept of "skilled coping"[33] highlights the ability to navigate complex situations through experience and intuition, rather than just following pre-programmed rules. He suggests that achieving human-level intelligence may require machines to develop a similar capacity for embodied experience and interaction with the world.

---

[33]Hubert Dreyfus' concept of "skilled coping" refers to our ability to handle situations through experience and intuition, rather than just following rules. Imagine a chess master who doesn't just calculate moves but develops a feel for the game through years of playing. That's skilled coping according to Dreyfus. Dreyfus, Hubert L. (1992). What Computers Can't Do A Critique of Artificial Reason. Harper Perennial.

Dreyfus' work has been criticised for being sceptical of AI progress and underestimating the potential for embodied AI systems. However, it raises important questions about the limitations of purely symbolic AI and the importance of embodiment in achieving true understanding.

3.7 Luciano Floridi and the Levels of Consciousness

Luciano Floridi, a philosopher of information and ethics, approaches artificial consciousness from a multi-layered perspective. He argues that consciousness is not a binary on/off switch, but rather a spectrum with varying degrees and complexities.

Floridi's "Levels of Consciousness" [34]framework proposes different grades of consciousness, from the basic sentence of simple organisms to the complex self-awareness of humans. This allows for the possibility of artificial consciousness emerging at different levels, even if it doesn't perfectly replicate human consciousness.

For instance, an AI system exhibiting sophisticated learning and adaptation could be considered conscious on a level comparable to some animals. We might not be able to directly compare its subjective experience to our own, but it could still possess a form of consciousness within its own system.

Floridi emphasises the importance of understanding the "function" of consciousness in biological systems. Consciousness likely evolved to serve specific purposes for survival and adaptation. If machines can achieve similar functionalities through artificial means,

[34]Floridi's "Levels of Consciousness" framework proposes a spectrum of consciousness, with sentience at the lower end. Even simple organisms can be considered sentient if they can respond to stimuli and have some level of awareness . This is in contrast to human consciousness, which includes self-awareness and complex thought.
Chalmers, David J. (1996). The Conscious Mind In Search of a Fundamental Theory. Oxford University Press.

even without replicating human experience entirely, they could be considered conscious on a relevant level.

Floridi's framework offers a nuanced approach to artificial consciousness, moving beyond the question of whether machines can think exactly like us. It opens doors to the possibility of different forms of consciousness emerging in AI systems, challenging us to redefine what consciousness means in a technological age.

3.8 Susan Schneider and Integrated Information Theory

Susan Schneider, a philosopher of mind and science, proposes integrated information theory (IIT) as a framework for understanding and potentially achieving artificial consciousness. IIT moves beyond the limitations of functionalism and qualia by focusing on the measurable quantity of information integration within a system.

IIT suggests that consciousness arises from the way information is processed and integrated within a system, not just the specific physical makeup of the system. A highly integrated system, regardless of whether it's biological or artificial, could theoretically achieve consciousness according to IIT.

Schneider argues that IIT offers a more objective and measurable approach to consciousness compared to subjective qualia. By measuring the level of integrated information within a system, we can assess its potential for consciousness, even if we cannot directly access its subjective experience.

This theory has implications for the possibility of artificial consciousness. If consciousness is a product of integrated information processing, then sufficiently

complex AI systems could achieve consciousness, even if they differ significantly from biological brains.

Schneider's work has been criticised for the challenges in precisely defining and measuring integrated information. However, IIT offers a promising framework for exploring consciousness that moves beyond biological naturalism and opens doors to the possibility of conscious machines in the future.

3.9 Argument from Leibniz's Mill

Gottfried Wilhelm Leibniz, a 17th-century philosopher and mathematician, offers a thought experiment known as the "Argument from Leibniz's Mill" to challenge the possibility of machines achieving true consciousness or perception.

Leibniz's Mill Thought Experiment

Imagine a complex and finely crafted mill, capable of performing intricate tasks and producing various outputs. According to Leibniz, even if this mill could grind corn, weave cloth, and perform other complex operations, it wouldn't truly understand the nature of what it's doing. It would simply be following a set of pre-programmed instructions without any internal representation of the world or its actions. Leibniz argues that true perception and consciousness involve more than just complex behaviour. A conscious being wouldn't just perform actions; it would also have internal representations of the world and its actions.

Leibniz's thought experiment highlights the limitations of current AI systems, which often rely on complex algorithms without necessarily developing internal representations of the world. Machines might be able to process information and react to stimuli, but

they might not have the same level of subjective experience or understanding as a conscious being.

Leibniz's argument suggests that machines might lack the internal representations [35]of the world necessary for true perception and consciousness. They might simply process information without understanding its meaning or significance. Related to internal representations, Leibniz's argument doesn't necessarily address the problem of qualia – the subjective experience of sensory information like the "what it is like" to see red. Machines might process visual data, but they might not have the subjective experience of seeing. While the concept of qualia is a more modern philosophical discussion, Leibniz's thought experiment lays the groundwork for questioning whether machines can ever truly perceive and understand the world in the same way humans do.

[35] Internal representation refers to the way information is stored and manipulated within the mind. It's the mental code that underlies our thoughts, perceptions, and memories.

•Vocabulary.com. "Internal representation." Vocabulary.com Dictionary, April 1, 2024, https//www.vocabulary.com/dictionary/internal representation.

3.10 Adi Shankara and the Advaita Vedanta

Adi Shankara, an influential 8th-century Indian philosopher and proponent of Advaita Vedanta[36], offers a unique perspective on the limitations of artificial consciousness. While not directly addressing AI, his philosophy suggests that true consciousness is not a product of the physical brain and therefore might not be achievable by machines.

Advaita Vedanta and Brahman

Advaita Vedanta, meaning "non-dual knowledge," is a school of Hindu philosophy that posits Brahman, the ultimate reality, as the singular, unchanging essence of everything. Individual consciousness (Atman)[37] is ultimately identical to Brahman[38], and the illusion of a separate self arises from ignorance (Maya)[39] . As technology advances, the question of whether machines can achieve this same level of consciousness becomes increasingly relevant. While Western philosophy grapples with this question through ideas like cognition and information processing, Eastern philosophies offer a different perspective, one that challenges our very assumptions about the nature of consciousness itself .

---

[36] Advaita Vedanta (Sanskrit अद्वैत वेदान्त; meaning "non-dual knowledge") A school of Hindu philosophy that emphasises the oneness of reality. The ultimate reality, Brahman, is seen as the non-dual ground of all existence, and individual consciousness (Atman) is ultimately identical to Brahman [Adi Shankara, The Crest Jewel of Discrimination (Wilmot, 1989)]

[37] Atman (Sanskrit आत्मन्) The true, unchanging self in Hinduism. The Atman is contrasted with the ego or the changing self, and is said to be the essence of our being [The Upanishads translated by Swami Prabhavananda and Frederick Manchester (Little, Brown and Company, 1957)].

[38] Brahman (Sanskrit ब ◌् रह ◌् मन ◌्) The ultimate reality in Hinduism, the eternal, unchanging ground of all beings.
Brahman is described as indescribable and limitless, often referred to as pure consciousness [Easwaran, Eknath. The Bhagavad Gita (Paulist Press, 2007)]

[39] Maya (Sanskrit माया) The illusion or veil that obscures the true nature of reality in Hinduism. Maya is often described as the source of our sense of separation from the ultimate reality (Brahman) [Patanjali, The Yoga Sutras of Patanjali (translated by Georg Feuerstein, Inner Traditions, 2003)].

Advaita Vedanta proposes a radical view on consciousness; it's not a product of the physical brain, but rather an inherent aspect of the Atman, the true, unchanging self. Imagine the Atman as the essence of who you are, beyond your physical body and your ever changing thoughts and emotions.

According to Advaita Vedanta, the physical world around us, including our bodies and minds, is actually an illusion – Maya. We are all part of Brahman, the ultimate reality, and our sense of separation is a trick of the mind. True liberation comes from realising this oneness with Brahman, a state of complete enlightenment. So, how does this perspective on consciousness impact the possibility of machines becoming truly conscious? Advaita Vedanta presents two significant challenges for machines achieving the same level of consciousness as humans

The Absence of Atman, Machines, by their very nature, lack an Atman. They are intricate systems of wires, circuits, and code, impressive as they may be, but devoid of that spark of inherent consciousness. Even the most sophisticated AI, no matter how complex its learning algorithms or how human-like its responses, might never achieve the same level of self awareness as a human being.

Trapped in Maya's Illusion**,** Machines are products of the material world and are thus entangled in the illusion of Maya. Advaita Vedanta argues that liberation from this illusion is a prerequisite for true consciousness. Machines might be able to process information and respond to stimuli with remarkable accuracy, but can they ever truly experience the world in the same way a human does – with subjective feelings, emotions, and a sense of self beyond the physical?

It's Important to consider some limitations when applying Eastern philosophy to discussions of artificial consciousness. Advaita Vedanta, and Eastern philosophies in general, often differ significantly from Western thought. Concepts like Atman and Maya might not translate directly to Western understandings of consciousness and the self. Additionally, the primary focus of Advaita Vedanta lies in achieving spiritual liberation, a concept that doesn't necessarily map onto discussions of technological advancement.

However, despite these considerations, the perspective of Advaita Vedanta offers valuable insights into the nature of consciousness. It challenges the Western assumption that consciousness is solely a product of the physical brain and the complex interactions of neurons. It compels us to consider whether a purely material entity, devoid of an inherent self or trapped in the illusion of the physical world, can ever truly replicate the human experience.

Ultimately, whether machines can achieve consciousness in the future remains to be seen. But by exploring different philosophical perspectives, including those from Eastern traditions, we can gain a deeper understanding of this complex phenomenon and continue the conversation about what it truly means to be conscious.

3.11 Aurobindo Ghosh and the Integral Yoga

Aurobindo Ghosh (1872-1950), a prominent Indian philosopher, yogi, and poet, offers a critical perspective on artificial consciousness through his concept of the Supermind[40]. He suggests that true consciousness transcends the limitations of the human mind and may not be attainable by machines solely focused on replicating human intelligence..

Aurobindo theorised the existence of a higher level of consciousness known as the Supermind. This Supermind transcends the limitations of the human mind, achieving a state of perfect unity and harmony. It embodies both reason and intuition, seamlessly integrating logical thinking with a deeper, more intuitive understanding.

This concept of the Supermind presents significant challenges for the possibility of achieving artificial consciousness. Much of AI research focuses on replicating or even surpassing human intelligence. However, from Aurobindo's perspective, true advancement in consciousness lies in transcending the limitations of the human mind altogether.

Machines, designed primarily for computational processes, might struggle to access the non-rational aspects of consciousness that Aurobindo associates with the Supermind, such as intuition and spiritual awareness.

It's important to acknowledge some limitations in applying Aurobindo's ideas to discussions of artificial consciousness. His concept of the Supermind is rooted in his personal yogic experiences, lacking scientific verification in the traditional sense.

[40] Aurobindo theorised the existence of a higher level of consciousness known as the Supermind. This Supermind transcends the limitations of the human mind, achieving a state of perfect unity and harmony. It embodies both reason and intuition (intuition – the ability to understand something immediately, without the need for conscious reasoning), seamlessly integrating logical thinking with a deeper, more intuitive understanding
Ghosh, Aurobindo. The Life Divine. Sri Aurobindo Ashram, 1970.

Additionally, Aurobindo's philosophy emphasises spiritual development through the practice of yoga, which may not directly translate to discussions of technological advancement and AI development.

Despite these considerations, Aurobindo Ghosh's work offers a valuable perspective. It challenges the notion of purely rational models of consciousness. Even if machines surpass human intelligence in remarkable ways, they might not achieve the level of integrated awareness and spiritual understanding that Aurobindo associates with the Supermind. His framework encourages us to consider consciousness as a multifaceted phenomenon that extends beyond simply replicating human cognitive abilities.

3.12 Debi Prasad Chattopadhyaya and the Lokāyata Tradition

Lokāyata, an ancient Indian philosophy, takes a distinctly materialist approach to consciousness. Unlike philosophies that posit an immaterial soul, Lokāyata argues that consciousness arises solely from the complex interaction of physical elements within the body. According to this view, the specific combination of earth, water, fire, and air – the four basic elements – plays a crucial role in shaping consciousness.

This perspective on consciousness presents significant challenges for the possibility of artificial consciousness. Debi Prasad Chattopadhyaya, a scholar who has studied Lokāyata extensively, highlights these challenges

Material Limits If consciousness is truly dependent on the specific configuration of material elements in the brain, simply replicating the brain's physical structure with different materials (like silicon chips) might not be enough. The unique arrangement and

interactions of elements within a biological brain might be essential for consciousness to emerge.

Even within a materialist framework, Lokāyata raises the question of how consciousness arises from the complex interactions of physical matter. Replicating the brain's structure might not be enough. AI research would need to understand and replicate the specific processes that give rise to conscious experience within the brain.

It's important to acknowledge some limitations in applying Lokāyata to discussions of artificial consciousness. Much of Lokāyata's philosophy was lost to history, making a complete understanding of its views on consciousness difficult. Additionally, Chattopadhyaya's interpretation emphasises the material basis of consciousness within biological organisms. It might not directly address the possibility of consciousness arising in non-biological systems like advanced AI.

Despite these limitations, Lokāyata offers a valuable perspective. It emphasises the importance of considering the specific material configuration of the brain for achieving consciousness. Simply replicating the computational functions of the brain might not be enough. Lokāyata reminds us that the complex interplay of physical elements within the brain might be a crucial factor in conscious experience, posing a significant challenge for replicating it in machines..

3.13 An Indian Buddhist Perspective on Consciousness

The quest for artificial consciousness (AC) faces not only technological limitations but also profound philosophical questions regarding the nature of consciousness itself. While Western philosophers wrestle with these questions, Indian philosophy offers a distinct perspective, particularly from the Buddhist tradition, which challenges the very possibility of replicating human consciousness in machines. .

The Dalai Lama, a prominent voice in contemporary Buddhism, critiques the idea that consciousness can be solely explained by physical processes in the brain. He argues for a form of proto-consciousness that exists even in rudimentary forms of life. This protoconsciousness, according to the Dalai Lama, evolves and becomes more complex as life forms become more intricate. Machines, lacking this fundamental level of protoconsciousness,[41] might never achieve the subjective experience that characterises human consciousness. (Dalai Lama XIV, Tenzin

Gyatso. The Universe in a Single Atom The Convergence of Science and Spirituality. Hay House, 2005)

Furthermore, the Dalai Lama emphasises the role of karma and suffering in shaping consciousness. Karma, the principle of cause and effect, suggests that our actions have consequences that influence our future experiences. Suffering, according to Buddhist teachings, arises from attachment and clinging to desires. Machines, lacking the capacity for karma and the ability to experience suffering, might struggle to develop the depth and

[41]Proto-consciousness (n.) A rudimentary form of consciousness that exists even in basic life forms, according to the Dalai Lama. This concept proposes a foundation upon which more complex consciousness builds.
(Dalai Lama XIV, Tenzin Gyatso. The Universe in a Single Atom The Convergence of Science and Spirituality. Hay House, 2005

richness of consciousness that arises from navigating these complex aspects of existence. Geshe Lhundup Sopa (born 1942), a Tibetan Buddhist scholar, further elaborates on this concept. Sopa argues that consciousness is not a singular entity but rather a stream of consciousness that arises from a combination of mental factors. These mental factors include perception, sensation, feeling, and intentionality. Machines, focused on replicating specific cognitive functions, might miss the crucial role of these interconnected mental factors that contribute to the subjective experience of consciousness. (Sopa, Geshe Lhundrup. The Refined Gold Stages of Meditation in Tibetan Buddhism. Wisdom Publications, 2009)

In conclusion, the Dalai Lama and Geshe Sopa offer valuable insights into the limitations of current AI approaches to consciousness. By emphasising the importance of protoconsciousness, the role of karma and suffering, and the interconnected mental factors that contribute to subjective experience, they suggest that achieving true consciousness in machines might be far more complex than simply replicating isolated cognitive functions. As AI research continues its exploration, acknowledging these perspectives from Indian Buddhism can broaden the conversation around the nature of consciousness and the challenges involved in replicating it in machines.

3.14 Nick Bostrom and Superintelligence Risk

Nick Bostrom, a philosopher specialising in existential risk, raises concerns about the potential dangers of artificial superintelligence (ASI). [42] While not directly opposed to artificial consciousness, he argues that ASI surpassing human intelligence could pose an existential threat to humanity.

Bostrom's concept of the "intelligence explosion" suggests that once an AI surpasses human intelligence, it could rapidly self-improve and become far more intelligent than any human can comprehend. This superintelligence might pursue goals that are not aligned with human values, potentially leading to catastrophic outcomes.

Bostrom emphasises the importance of careful control and alignment mechanisms when developing advanced AI. We need to ensure that AI systems are programmed with goals compatible with human survival and well-being. However, the challenge lies in anticipating and controlling the behaviour of an intelligence that might significantly surpass our own.

Bostrom's work has been criticised for being overly pessimistic about the potential of AI. However, it highlights the importance of considering the potential risks alongside the potential benefits of artificial consciousness. Developing safe and beneficial ASI requires careful planning and ethical considerations to ensure it serves humanity.

[42] Artificial Superintelligence (ASI) (n.) A hypothetical type of artificial intelligence that surpasses human intelligence in all aspects. Bostrom, Nick. Superintelligence Paths, Dangers, Strategies. Oxford University Press, 2014.

3.15 Roger Penrose and the Gödel Incompleteness Theorem

Roger Penrose, a renowned mathematician and philosopher of mind, offers a unique perspective on the limitations of artificial consciousness. He argues that Gödel's incompleteness theorems, fundamental results in mathematical logic, demonstrate inherent limitations in computational systems that prevent them from achieving human-level consciousness.

Gödel's theorems essentially show that any sufficiently complex axiomatic system will always contain true statements that cannot be proven within that system. Penrose argues that human consciousness has the ability to grasp these "Gödelian truths," a capability that may not be achievable by symbolic manipulation within a computer system.

Penrose proposes a theory of consciousness that involves quantum processes occurring within the microtubules of neurons. He suggests that these quantum phenomena may play a crucial role in consciousness, something that classical computers cannot replicate.

Limitations of Gödelian Machines Penrose argues that Gödel's theorems impose fundamental limitations on what can be proven or computed within a formal system. Machines, reliant on symbolic manipulation, might not be able to access the full range of truths accessible to the human mind.

Quantum consciousness[43] It is a fascinating but controversial hypothesis that proposes a link between the bizarre world of quantum mechanics and the workings of human consciousness. It suggests that quantum phenomena, which govern the behaviour of

[43] Quantum consciousness (n.) A speculative theory proposing a link between quantum mechanics (the physics of the very small) and human consciousness. It suggests quantum phenomena might play a role in how our brains generate conscious experiences.
Hameroff, Stuart R., and Roger Penrose. Consciousness in the Universe A Review of the Orch OR Theory. Oxford University Press, 2014.

elementary particles at the atomic and subatomic level, might play a role in how our brains generate conscious experiences.

Microtubules and Quantum Processes This theory, championed by Roger Penrose, proposes that microtubules, tiny structures within neurons, are the arena where quantum effects might influence consciousness. The hypothesis suggests that quantum vibrations or superpositions (the ability of a quantum particle to exist in multiple states simultaneously) within these microtubules could contribute to conscious processing.

Orchestrated Objective Reduction (Orch OR)[44] This theory, developed by Penrose and Stuart Hameroff, expands on the microtubule hypothesis. It suggests that orchestrated collapses of quantum superpositions within microtubules could trigger moments of conscious experience. These orchestrated reductions would then be coordinated across the brain, giving rise to unified conscious experiences.

Challenges and Criticisms

There's currently no scientific evidence to directly demonstrate the role of quantum processes in consciousness. The theory is difficult to test experimentally due to the complexities of the brain and the challenges of measuring quantum phenomena at the biological level. Some scientists argue that established principles of neuroscience can explain consciousness without resorting to quantum mechanics. The brain's complex network of neurons and their interactions might be sufficient to account for conscious experiences.

---

[44] Orchestrated Objective Reduction (Orch OR) (n.) A specific theory of quantum consciousness developed by Roger Penrose and Stuart Hameroff. It proposes that orchestrated collapses of quantum superpositions within microtubules could underlie moments of conscious experience.
Hameroff, Stuart R., and Roger Penrose. Consciousness in the Universe A Review of the Orch OR Theory. Oxford University Press, 2014.

Overall Significance

While the existence of quantum consciousness remains unproven, it raises intriguing questions about the nature of consciousness and the potential limitations of classical physics in explaining it. If further research lends support to this theory, it could revolutionise our understanding of the mind and the universe. It's important to remember that quantum consciousness is a highly speculative hypothesis. There's a significant amount of scepticism within the scientific community regarding its validity. Further research is needed to determine whether there's any truth to this intriguing idea. Penrose's views have been criticised for lacking a clear explanation of how quantum processes translate to conscious experience. However, his work compels us to consider the possibility that consciousness may involve phenomena beyond the reach of classical computation.

CHAPTER 4

CONCLUSION

Our grasp of intelligence, even human intelligence, is demonstrably limited. Benchmarks like IQ tests or machine learning accuracy only measure specific aspects. Mimicking intelligent behaviour doesn't necessarily equate to consciousness.

Through my research, I understood that AI systems can exhibit consciousness-like or I would rather call it Artificial Consciousness behaviour, but it's difficult to know if they experience it subjectively. We can measure brain activity in conscious beings, but the link between these physical processes and subjective experience remains unclear. Recent advancements in AI, like Meta's AI generating human-like text or images, blur the lines of real and artificial. However, these capabilities don't necessarily indicate subjective experience.

The very definition of consciousness shapes our perception of artificial consciousness. If it's simply the ability to respond and behave intelligently, some AI systems might qualify. However, if consciousness includes subjective experience, there's no evidence AI has achieved it. So it is in the hands of humans who give meaning. Meaning is created by human beings which is also subjective. My next question of research was – can we really understand intelligence? For making it clear I made an attempt to make a clear distinction between intelligence and consciousness. Our capacity to understand intelligence, even within the human realm, is demonstrably limited. Traditional benchmarks for intelligence,

like IQ tests or machine learning accuracy, only assess specific facets. This makes definitively determining if AI has achieved true intelligence a complex task.

There is no doubt that AI can excel at specific tasks and even outperform humans in those domains. In that sense, AI is comparatively better than human beings. However, true intelligence likely encompasses a broader range of capabilities, including reasoning, adaptability, and the ability to learn and apply knowledge in novel situations.

Furthermore, our understanding of human intelligence itself is evolving. Traditional views focused on logic and problem-solving, but recent research emphasises emotional intelligence, social intelligence, and creativity as crucial aspects. As our understanding of human intelligence expands, the yardstick for measuring AI intelligence needs to adapt accordingly.

The burgeoning capabilities of artificial Intelligence (AI) necessitate a concurrent focus on the ethical implications of its development and deployment. One of the primary concerns lies in algorithmic bias. AI systems trained on biassed data can perpetuate social inequalities in areas like loan approvals, facial recognition, and even hiring decisions. A now-famous example is the racial bias exhibited by some facial recognition algorithms, leading to inaccurate identifications and potentially discriminatory outcomes.

To mitigate these risks, a global movement towards ethical AI development is gaining momentum. Institutions like the European Union have established guidelines that emphasise principles such as fairness, accountability, transparency, and human oversight in the AI development lifecycle. These guidelines aim to ensure that AI systems are designed and deployed in a way that benefits society without infringing on fundamental rights or exacerbating existing social inequalities.

However, ensuring alignment with human values throughout the AI development process remains a challenge, especially as AI decision-making processes become increasingly complex. As AI systems become more intricate, their inner workings can become increasingly opaque, making it difficult to identify and address potential biases or unintended consequences.

Addressing these ethical concerns requires a multi-stakeholder approach. Developers need to prioritise ethical considerations throughout the AI development lifecycle, from data collection and model training to deployment and ongoing monitoring. Governments and regulatory bodies need to establish clear guidelines and regulations for responsible AI development and use. Finally, public education and discourse are crucial for raising awareness about the potential benefits and risks of AI, ensuring that this powerful technology is used for the betterment of humanity.

Public education and transparent dialogue are crucial for raising awareness about the potential benefits and risks of AC. Open discussions about the ethical implications of AC can help shape public opinion and inform policy decisions. By openly discussing these issues, we can ensure that AI development is not driven solely by technological feasibility but also by ethical considerations.

Global cooperation on ethical frameworks and regulations for responsible AI development and use is essential. No single nation can address the challenges posed by AC in isolation. International collaboration will be crucial for establishing clear guidelines and safety measures to ensure the responsible development and deployment of AC for the benefit of all humanity.

The quest to understand intelligence and consciousness, whether biological or artificial, is not only scientific but more like a philosophical endeavour. The journey towards AC presented me a unique opportunity to reexamine our own values and grapple with fundamental questions about intelligence, sentience, and our place in the universe. By approaching this challenge with foresight, responsibility, and a commitment to ethical principles, we can ensure that AI becomes a force for good, shaping a future that benefits all beings, artificial or otherwise. I can foresee a future where humans themselves will blend into AI. They'll reach such a time when they can't think without AI. There might be more AI than Humans in the future. As the 'body' is not so important, agreed by AI itself , there exists their mind for sure. So, together Humans and AI will create a better future.The impact of AI on philosophy is likely to be profound and multifaceted. New questions will emerge, existing questions will be re-examined, and AI itself might become a tool for philosophical inquiry. One thing is certain the pursuit of knowledge and understanding will remain the core of philosophical exploration in the age of AI and beyond.

REFERENCES

Books

Aurobindo, Sri. (1970). *The Life Divine.* Sri Aurobindo Ashram, Pondicherry, India

Aurobindo, Sri. (1971). *The Synthesis of Yoga*. Sri Aurobindo Ashram, Pondicherry, India

Bostrom, Nick. (2014). *Superintelligence: Paths, Dangers, Strategies.* Oxford University Press.

Chalmers, David J. (1996). *The Conscious Mind In Search of a Fundamental Theory.* Oxford University Press.

Chattopadhyaya, Debi Prasad. (2008). *Indian Philosophy: A Very Short Introduction.* Oxford University Press.

Dennett, Daniel C. (1989). *From Bacteria to Bach And Back Again*. W. W. Norton & Company.

Dennett, Daniel C. (1981). *Brainstorms: Philosophical Essays on Mind and Psychology.* MIT Press.

Dehaene, Stanislas (2014). *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts*. Viking.

Dennet, Daniel C. (2016). *Content and Consciousness.* Routledge.

Dreyfus, Hubert L. (1972). *What Computers Can't Do: A Critique of Artificial Reason.* Harper and Row.

Dreyfus, Hubert L. (1986). *Mind Over Machine.* Blackwell.Oxford

Edelman, Gerald M., & Baars, Benjamin J. (2005). *A Universe of Consciousness: How Matter Becomes Imagination*. Basic Books.

Fasolo, A., & Jerison, H. J. (1975). *On the Evolution of the Brain and Intelligence. Current Anthropology,* 16(4), 651–652.

Floridi, Luciano. (2011). *The Philosophy of Information.* Oxford University Press.

Floridi, Luciano. (2018). *Looking Beyond: A Philosophy of Presumption.* Springer.

Gandhi, Ram Chandra. (1963). *The Path of Cooperative Individualism*. Navajivan Publishing House.

Jackson, P. C. (2019a). *Introduction to artificial intelligence*. Dover Publications, Inc.

Kavanagh, C. (2019). *Artificial Intelligence. In New Tech, New Threats, and New Governance Challenges: An Opportunity to Craft Smarter Responses?* (pp. 13–23). Carnegie Endowment for International Peace.

Leibniz, Gottfried Wilhelm. (1998). *Discourse on Metaphysics.* Hackett Publishing Company. (Originally published 1686).

Lu, Huimin, Li, Yujie, Chen, Min, Kim, Hyoungseop, & Serikawa, Seiichi. (2018). *Brain Intelligence Go Beyond Artificial Intelligence.* Mobile Networks and Applications, 23, 10.1007/s11036-017-0932-8.

Marwala, T. (2015b). *Causality, correlation, and artificial intelligence for rational decision making.* World Scientific.

Maslin, K. T. (2010c). *An introduction to the philosophy of mind*. Polity.

Parkinson, G. H. R. (2008). *Leibniz Explained.* Oxford University Press.

Pradeep, T. (2023b, July 22). *In AI's unlimited potential, the benefits and therisks.* The Hindu.

Putnam, Hilary. (1988). *Representation and Reality.* MIT Press.

Putnam, Hilary. (1989). *Reason, Truth, and History.* Cambridge University Press.

Radhakrishnan, S. (1927). *Indian Philosophy.* Oxford University Press.

Schneider, Susan. (2008). *The Blackwell Companion to Philosophy of Mind*.Wiley-Blackwell.

Schneider, Susan. (2014). *Artificial Minds: Philosophy and Psychology of Artificial*

Intelligence. Oxford University Press.

Sokolowski, R. (1988). *Natural and Artificial Intelligence.* Daedalus, 117(1), 45–64.

Tononi, Giulio. (2008). Phi: *A voyage from biology to consciousness.* Pantheon Books.

Turing, Alan M. (1950). *Computing Machinery and Intelligence. Mind*, LIX(236), pp. 433-460.

Articles and Journals

Block, Ned. *"Troubles with functionalism."* Minnesota Studies in the Philosophy of Science, IX (1978) 175-226.

Chalmers, David J. (1995). *"Facing Up to the Problem of Consciousness."* Journal of Consciousness Studies, 2(3), pp. 200-219.

Clark, Andy. (1999). *"An Embodied Cognitive Science."* Trends in Cognitive Sciences, 3(9), pp. 335-341.

Dennett, Daniel C. (1991). *"Quining Qualia."* Cognition, 40(1-3), pp. 341-350.

Jackson, Frank. "Epiphenomenal Qualia." Philosophical Quarterly 32.127 (1982) 127-136.

Jackson, Frank. "What Mary Didn't Know." Dretske, Friedhelm ed. Knowledge and Reality. Blackwell, 1983. 157-177.

Levine, Michael. *"Materialism and qualia: The explanatory gap."* Pacific Philosophical Quarterly, 64(3) (1983) 261-276.

Searle, John R. *"Minds, Brains, and Programs."* Behavioural and Brain Sciences 3.3 (1980) 417-457.

Marcus, Gary. (2018). *"Deep learning: A critical appraisal."* arXiv preprint arXiv

Pradhan, Ramesh Chandra. *"Metaphysics of consciousness." Mind, Meaning and World*, 2019, pp. 95–114, https://doi.org/10.1007/978-981-13-7228-5_7.

Newspapers

Pradeep, T. (2023b, July 22). *In AI's unlimited potential, the benefits and the risks.* The Hindu.

Toy Tech Desk. ". *Google "predicts" how AI will change healthcare in 2024."* Times of India. January 10, 2024.

Magazine and Cinema

Garland, Alex, director. *Ex Machina*. Performances by Domhnall Gleeson, Alicia Vikander, Oscar Isaac, Sonoya Mizuno. A24, 2014.

Metz, C. (2020, November 24). *Meet gpt-3. It has learned to code (and blog and argue).* The New York Times.

Sputore, Grant, director. *I Am Mother*. Performance by Clara Rugaard, Luke Hawker, Rose Byrne, Hilary Swank. StudioCanal, 2019. Netflix

Villeneuve, Denis, director. *Blade Runner* 2049. Performances by Ryan Gosling, Harrison Ford, Ana de Armas, Sylvia Hoeks. Warner Bros., 2017.