Artificial Intelligence and Consciousness

Sonal Lakhotia
Institute of Computer Science, University of Goettingen

***Abstract*—Artificial Intelligence has brought an evolutionary change in human-computer interaction. Machines are intelligent and capable of completing complex tasks with efficiency. Domains of computer vision, natural language processing, robotics, and autonomous systems have revolutionized due to the development of AI. Advancements in Artificial Intelligence enabled the development of driver-less cars, humanoids, and intelligent machines that outperformed humans. These developments form Narrow AI. Machines undergo exclusive training using machine learning and deep-learning strategies to obtain a particular result. Unlike machines, human beings do not need training each time a new environment or a new problem arises. Intelligent devices capable of self-adaption and self-correction comprise Strong Artificial Intelligence. Humans provide unique solutions to each problem statement and derive new and imaginative strategies for any situation with emotional intelligence. A human being is capable of these due to consciousness. A conscious machine would be capable of deriving solutions for unmet problems and learning from them. In this report, we discuss artificial intelligence and consciousness in AI.**

***Index Terms*—Artificial Intelligence, Consciousness, Emotional Intelligence, Strong Artificial Intelligence, Narrow AI**

I. Introduction

Human beings possess intelligence and sentience like no other living creature. Since time immemorial, humans have made considerable improvements in every domain such as technological, business management, manufacturing units, medical services, media, entertainment, food generation and delivery, energy generation, and numerous others. Most common applications of AI could be visualized as below.

Fig. 1. Artificial Intelligence beats world Go champion

Machines took over the labor-intensive tasks in the machinery and factories. They assist humans in accountancy, resource management, new hiring and maintenance, and cleaning tasks. Humans and AI exist in a co-dependent environment. AI empowers machines to perform every task effortlessly and accurately and augment Human Intelligence to enable effective decision-making and minimize fault occurrence. The development of machine learning algorithms, deep learning algorithms, reinforcement learning algorithms, and neural networks enabled efficient Artificial Intelligence. Artificially Intelligent machines outperformed human beings in several activities. IBM Watson defeated the world's greatest Jeopardy Champions, DeepMind's AlphaGo defeated the best human Go player, Libratus, a Poker AI, defeated a team of professional players, and Waymo's driverless cars drove more than four million miles on the road. Neural networks work the same way as neurons in the brain operate in living creatures, particularly human beings. Neural networks using Deep-Learning performed as efficiently as radiologists while reading CT scans and won the Imagenet computer vision contest by improving the image classification tasks.

Fig. 2. Artificial Intelligence beats world Go champion

Humanoid robots broke the uncanny valley and emerged as chat-bots with expressions and gestures. One such example is Sophia, developed in Hanson robotics. Sophia became the first artificially intelligent robot to gain citizenship, just like a Human. At the Saudi-sponsored Future Investment Initiative (October 2017) in Riyadh [2], Sophia became the first robot to have citizenship. Beneath Sophia's surface interactions is a sophisticated and coordinated system of about 30 whirling servos (tiny, heat-generating motors), [2] many connected to the interior of Sophia's face, [2] to create a human-like look when she talks [2]. Does that mean robots with human-like appearance and attached input devices with audio and visual capabilities are potential humans? Would they be able to think

and act like humans?

Fig. 3. Humanoid Robot Sophia at Morals and Machines Conference at Dresden

The achievements in AI enforce our thoughts with factors like humanoid robots are at par with humans. Just like self-driving cars, we could have all self-operated devices. Any machine that passes the Turing test and possesses the capability of predictively analyzing or classifying tasks is at par with humans. Machines and robots could be destructive hence should be developed only under defined guidelines. They do not possess sentience or emotional attachment with their surroundings. They are not even aware of themselves or their surroundings. They only act the way they are programmed. Thus, forming Narrow AI. Narrow AI is capable of producing well-defined results for a particular task only. It is trained extensively on a set of parameters or data and tested around those. Hence, it gives us the desired outcome. Self-driving cars, neural networks in medical sciences, and even the humanoid robot Sophia is Narrow AI. Human-machine interaction designers have linked narrow AI algorithms [2] to make an efficient algorithm. The result is a speech-reciting robot [2] that can drum up witty conversations with the pre-loaded texts [2] , follow it up with machine learning [2] to match facial expressions and pauses to the text [2].Narrow AI performs at par with human intelligence and enables a robot with no emotional intelligence to have citizenship. Does that mean AI is capable of thinking, adapting, and correcting itself on its own? No, this needs the development of Strong Artificial Intelligence or Artificial General Intelligence. Self-awareness, awareness of the surrounding, and emotional intelligence are accounted for while developing Strong AI. Though narrow AI has surpassed human intellect, tenacity, and strength, it is not as efficient as humans because it lacks awareness, sentience, and self-referential capabilities. Consciousness is responsible for these unique qualities in living creatures. Out of all the living beings, human beings are considered most conscious. The human brain's configuration is to understand and extract information from unknown situations and circumstances. The human brain analyzes a problem through multiple perspectives and provides correct information. Would a deep learning neural network classify the given image correctly?

Fig. 4. An Optical Illusion:An image with two different perspectives

An human eye can explicitly tell apart the young woman and old woman from this picture that might not be attainable by a computer or an intelligent machine if this image was shown to it for the first time.

II. What is Consciousness?

A conscious entity must possess intentionality and qualia[5]. Homo Sapiens have the most highly developed level of consciousness [3] among all the other species on the planet. Consciousness is a unique characteristic controlled by billions of neurons in the brain. It differs from Intelligence. It not only includes manipulating the environment according to one's knowledge but also perceive it subjectively. We do not know or understand how consciousness emerged in human beings. Many aspects of consciousness aren't known to human beings. The emergence of self-consciousness, sociological, philosophical, and phenomenological consciousness in every person is inexplicable and uniquely defined. In the words of physicist Max Tegmark, Consciousness is just another state of matter [4], like a solid, liquid, or gas. Just as there are many types [4] of liquid, there are many types of consciousness. Unlike, a book or a computer which are also made of matter, humans are unique and different. This is because they have brains and consciousness derives from the brain.

III. Consciousness In Human Beings

Consciousness in humans could be described in following ways:

1) Human consciousness involves[3] perceptual awareness of their environment. All species (from amoeba to zebra) [3] have perceptual awareness of their environments [3].
2) Humans are aware of themselves, their environment, and the relationship between both [3]. They are self-aware that they are both distinct from and a part of an ecology [3]. This phenomenon makes them possess meta-awareness.
3) Human beings can recognize the relationship between their beliefs and ecology [3]. They can increase the likelihood of survival and reproductive capacities through trials, errors, and modification of actions [3].
4) Another factor responsible for the emergence of consciousness is that humans have a strong sense of mortality [3]. We know that we would die and paint a clear conceptual picture about our own lives [3]. The knowledge of death provides humans greater insight into their existence and the awareness of actions that may help or hinder it [3].
5) The development of human consciousness [3] has led to the emergence of the theory of mind. An efficient capacity [3] to visualize the world from others perspective but also to make decisions about the behavior of others based on that capacity developed in us.This empowered us to possess an increased ability [3] in understanding our society, our abilities to make inferences about thoughts and actions of other people [3], noting errors, and emotions such as sympathy and empathy[3].

Consciousness in human beings is due to the differential and unique arrangement of neurons in the human brain. It is a subjective emergence that works best when a person is awake. A sleeping person belongs to the subconscious and unconscious states depending on medical habits or consumption of sleep-inducing substances. A sleeping or anesthetized human is lesser conscious than an awake and healthy human. Consciousness empowers us to make rational decisions imbibed with emotional intelligence and imagination.

IV. Consciousness in Machines

According to [5], conscious machines are a by-product of science fiction, and machines could never be self-aware. The sci-fi movie depicts an artificially intelligent cyborg capable of intuitive decision-making and responding to the stimulus with sentience. The portrayal of these humanoid robots forces us to believe that artificially intelligent robots or machines could be conscious in the future. We discussed the working principles behind existing cyborgs, which comprise Narrow AI. They are incapable of procuring results outside their training domain and do not consider aspects of societal interest.We rely on the Turing test to establish machines are capable of thinking like a human beings. Does a human only consider parameters and symbols to make decisions? Or is it guided by sentience and adaption?

Fig. 5. A capture from sci-fi illustrating decision-making strategies

Is the Turing test enough to determine if machines made decisions like human beings? We see in [6] that the Turing test doesn't even enforce if machines can think at all. Chinese room experiments proved that machines are incapable of consciousness because machines mimic the theoretical mechanisms of learning and simulate the result of learning [6]. They never replicate the experiential activity of learning. An actual learning experience requires connecting various referents with conscious experiences [6]. This explains the reason why machines mistake groups of pixels that make up an image of an ape with those that compose an image of a dark-skinned human being. Machines don't learn, they only pattern match [6]. There's no actual personal experience involved in associating a person's face with that of an ape. Thus, we could say that artificial consciousness in machines is impossible because a machine learns due to extrinsic programming [6] that is syntax bound and lacks meaning.

Consciousness emerged in human beings due to evolution. About two hundred years ago, humans were unaware of the origin of colors, but now we know the science and significance of every color in our day-to-day life. There is a possibility in the future that evolution in machines enables the development of artificial intelligence with consciousness in machines.Strong AI is under active research, whereas humans are born capable of adapting contextually. Decision-making in Humans is achieved by interacting with the environment.A human decision-making model capable of contextual adaption should be adopted for developing Strong AI.The real-time operational environment is unknown and changes rapidly [1]. It includes elements that are unknown while training the system. The ideal Strong AI should be capable of reasoning and adapting to the unknown factors and reacting appropriately to them without critically failing [1].AI systems adept at contextual adaptation would ensure the construction of models that would be self-explanatory for real-world phenomena and capable of comprehending the decisions made.Military strategist John Boyd developed the OODA loop to study military situations. It has been extended successfully to other domains such as law enforcement and business strategy development. The OODA loop is inspired by [1] and organizational and individual

learning. It highlights shortcomings of current AI and focuses on future research trends to achieve Strong AI [1].

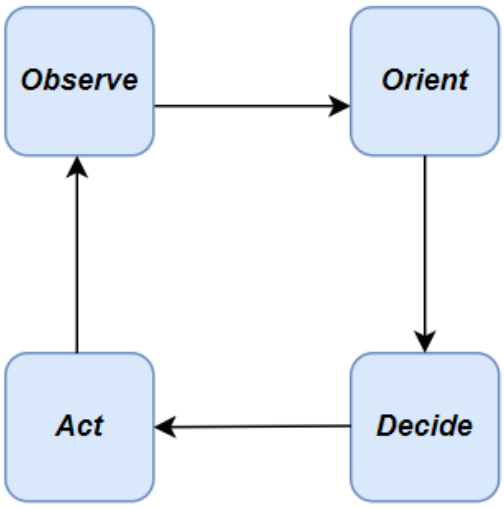

Fig. 6. A decision-making model that depicts contextual adaptation, a key feature for Strong AIs.

OODA in the loop stands for Observe, Orient, Decide and Act. We outline and discuss the elements of OODA loop as follows:

A. Observe

Observation is much more refined than seeing an object or an environment. It involves active and informed surveillance of the environment by focusing on areas of importance and making connections between different sources of information[1].Narrow AI involving deep neural networks succeeds in predictive analysis because large amounts of annotated training datasets and very high-quality data are used for training. There is a possibility of data deficit or lack of data for training the algorithms. Strong AI's would have to overcome data deficit by employing the transfer-learning or domain-adaption techniques [1]. There might occur a situation where no data is available for a particular, in that case, Strong AI's would have to be trained using zero-shot learning. Contextual adaptation is done via constraining the algorithms and extracting features of the algorithm to be used in the new task. This requires mimicking human capability [1] in understanding the relationship between various tasks and exploiting [1] the latent features [1] induced by the learning algorithm. An example for one such intelligent device would be a sentient surveillance system which would not only watch over the campus under surveillance but also report any unusual or undesirable activity to the respective authorities.

B. Orient

Analyzing past experiences and synthesizing them with the current observations involve orienting to a new situation [1]. The real-time operating environment is not pre-determined or clinical. Human beings with diverse and dynamic behaviors complicate the environment. A suitable orientation of new and complex situations requires an implicit understanding of human emotions [1]. We do not intend to develop a machine that cries or laughs according to the emotional situation around it [1]. Instead, the system should orient to a perspective like human beings around it by minimizing the human grief or contributing to decision-making without being affected by any emotional factors [1]. Large-scale emotionally intelligent AI agents could pacify a riot or contribute to a social-group behavior. An example of one such system would be an emotionally intelligent cyborg [1].

C. Decide

Decision-making involves generating various activities or suppositions after orientating to the circumstances. A reflexive hypothesis does not originate just as a byproduct of experiences because the environment is reactive and dynamic, and decision-making could be tricky. Human beings are capable of imaginatively assessing unknown situations and making effective decisions. Data analytics enable us to understand what comprises data but with imagination, we could analyze why the data is a certain way. Imaginative artificial intelligence could solve novel problems and could lead to enhanced solutions.An area of active research is learning affordance. It would enable us to understand the object in its current state and whatever could be done with it. Imaginative machines are not a new concept. The concept of machines with imagination is explored in the domains of computational creativity [1]. AI would aid the creation of art, music, painting, and engineering. Strong AI would be able to imagine and produce concepts and strategies to help in designing advanced types of equipment and materials. It could be AI in stock marketing or cryptanalysis [1] by an imaginative and result-driven approach.

D. Act

The act implies carrying out the decided action. The robotic technique requires the system to operate according to the training it receives [1]. They are pre-programmed and expected to work autonomously under various situations to provide an outcome. In the future, robots would have the ability to operate in a new simulation [1], even without being pre-programmed to do so. It is possible by integration of perception, sensing,planning,analysing, decision-making, and executing [1]. Dynamic modification of the programs to adhere to recognition of the new task, identification of the similar tasks and tracking the task if it were to be met again [1].Robotic systems that could function autonomously and collaboratively need to be developed.

V. Are We Approaching Robotic Consciousness?

Consciousness has varied definitions and perspectives as per a technologist or a physicist. Consciousness is a subjective and emergent property and makes us aware of our surroundings. It enables us to possess sentience.One such definition of consciousness as given by the oxford dictionary describes that the ability to be self-aware means that an entity is conscious. An experiment along those principles was designed by Professor Selmer Bringsjord [7] of New York's Rensselaer [7]Polytechnic Institute. The classic logic puzzle "Wise Men" was carried out to test a group of robots.Bringsjord and his research team [7] called this puzzle test the "ultimate sifter" test [7] as this knowledge game is capable of separating people from machines very quickly – only a person can pass the test [7]. The research team demonstrated that a robot passed the test. The classic puzzle is designed along with the following premise. A king has three wise advisors, wearing hats. He instructs his advisors that the contest would be fair and the hats they wore are either blue or white [7]. The first one to identify the color of the hat on his head would win [7].The contest would be fair if all the three advisors wore the same color hat. Thereby, the winning wise man would be able to determine the color of the hats on the other two men, and then imply that his hat was the same color. The roboticists used a different version of this riddle to demonstrate self-awareness in robots.

Fig. 7. An experiment demonstrating 'wise men' puzzle test

They were programmed to believe that only two of them had received the dumbing pill. When asked, who received the dumbing pill, two of the robots remained mute, and one that did not receive the dumbing pill responded with "I don't know". After hearing its own reply, the robot concluded that it did not receive the dumbing pill, or it would not have been able to respond at all. The robots exhibited the knowledge of self and adapted the reply according to the situation.The mathematical deductions of the decision made by the robot were stored in the memory providing authenticity to the thought-process of the system.Further research has demonstrated the collaboration and autonomy of decisions while communicating. This experiment demonstrates a very narrow subset of consciousness, but it does not rule out the possibility of having conscious machines in the future. The concept of self-awareness arises from sharing experiences, emotional moments, and relationships with societal settings. It is a long way before we are able to produce conscious machines.

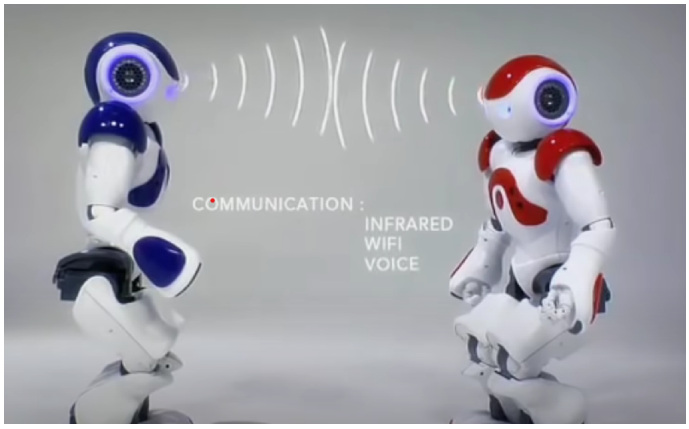

Fig. 8. Autonomous and Collaborative robots to be developed in future

VI. Further Scope of machine consciousness

The evolution and emergence of consciousness in human beings are still unknown. There is an infinite scope of research and development in this domain. We are conscious of our existence and affected by sociological, philosophical, psychological, and phenomenological factors. Experiments have proved a very narrow subset of self-awareness and collaboration among robots. We need to develop self-referential and adaptive systems capable of learning from their mistakes and correcting their code themselves. We are deeply inspired by science fiction and super power-equipped robotic characters to develop machines that could outperform humans in every aspect. Extensive research is carried out to build machines that think, act, and look like humans. A need to formulate laws and rights for robots would emerge if conscious machines came into existence. Assigning more and more power to robots could be highly destructive to society. A protocol determining the acceptable actions of these cyborgs would be designed.Tests capable of determining consciousness in machines have to be developed. We saw that the Turing test only focused on the functionality but not the reasoning behind the actions and reactions. Unless an efficient technique is designed to measure the level of consciousness in an entity or a computational agent it would be impossible to justify consciousness in it.Sufficient storage for emotions, relationships, and imagination need to be developed. If each entity uses the cloud for storage, shared networks would be able to transfer these emotions and incidents to all the connected components of the system. This would render the entire system duplicated. Subjectiveness or uniqueness would be lost because a brain hive would be created no individuality of thoughts would be maintained.Philo-psychological and phenomenological consciousness are only possible in living creatures. It is most prominent in human beings. Rigorous researches and experiments would be required to establish the existence and application of these in machines. Decisions about the ethical and legal rights of AI have to be formulated to co-exist with conscious machines.

VII. Conclusion

Consciousness has various ways descriptions and definitions. There are different types of consciousness and variant reasons behind their emergence with evolution in living beings, especially humans. Adhering to any one of the definitions to prove consciousness in intelligent systems is not enough. As of today, Narrow AI is only capable of solving stipulated problems. They are explicitly trained using a huge amount of annotated data and find usage when tested under a similar circumstance. It is like a black box that gives an output if the input is provided. The system does not reason or understand the origin, accumulation, processing, or modeling of the data. Intelligence and Consciousness work hand in hand in a Human Being but they are discreetly implemented in a machine. AI consciousness seems far-fetched. If we understand the reason for the emergence of consciousness in living entities we might be able to replicate it in the intelligent system. There is a theoretical possibility of creating a conscious machine or Strong AI if these facts are discovered.

References

[1] Gee-Wah Ng, Wang Chi Leung, " Strong Artificial Intelligence and Consciousness" *March 2020, Journal of Artificial Intelligence and Consciousness*

[2] Thomas Riccio, " Sophia Robot,An Emergent Ethnography" *March 2020, Journal of Artificial Intelligence and Consciousness*

[3] Christopher DiCarlo, "How to Avoid a Robotic Apocalypse: A Consideration on the Future Developments of AI, Emergent Consciousness, and the Frankenstein Effect."

[4] M.Tegmark, "Solid, Liquid, Consciousness."

[5] David Hsing, "Artificial Consciousness is impossible."

[6] James Moor, "The Turing Test: The Elusive Standard of Artificial Intelligence."

[7] Celena Chong, "This robot passed a 'self-awareness' test that only humans could handle until now" *July 23, 2015*