Cognitive Computing: The Future of Augmented Intelligence

Anoushka Mergoju

EECS, COMPUTER AND INFORMATION SCIENCES
amergoju@syr.edu
Syracuse University

***Abstract* - In certain situations where we are dealing with many complex entities that are linked together based on perception, decision-making and reasoning; where we need to process and understand unstructured data or build data-driven intelligent systems with high learning abilities, we come across a multi-disciplinary field of research that is, Cognitive Computing. The main idea behind cognitive computing systems is to equip a computer system with three driving forces of human cognition namely, knowledge, thought and emotion. The convergence of cognitive science, predictive analysis, and a variety of computing technologies has ushered in this emerging field. Cognitive theory explanations give out various frameworks for discussing the many different types of human cognition, such as the brain representation and how it processes information. The system processes used for extracting knowledge from multiple data organizations are provided by data science. Cognitive computing on the other hand, models human cognition using computing theories, methods, and tools. The goal of this paper is to identify current research based on the theory, application, and evaluation of augmented intelligence using cognitive systems through evaluation and thorough analysis. By reviewing all approaches and providing a comprehensive understanding of how the neurobiological mechanisms of a human mind are used to train a system, as well as examining the three main types of cognitive architectures and their limitations.**

***Keywords* - Cognitive Computing, Human Cognition, Intelligent System Architectures**

I. INTRODUCTION

A self-contained, self-managed system is what we mean by an autonomous system. As its environment changes, the system constantly reconstructs itself in real time. Learning, growth, and evolution are all embodied in the self-reorganization element. The way an intelligent system gains knowledge and develops its aspects involving behavior is mainly through its senses, ideas and personalized experiences which can also be termed as cognition. The realization and existence of autonomous systems depends solely on its cognitive processes.

Human cognition describes the cognitive abilities that enable individuals to carry out a variety of tasks, both routine and highly specialized. Machine cognition is a collection of techniques that allows computers to complete activities that aim for performance at human cognitive levels. Brain and mind are biological and natural tools used by humans for cognition. Basically, machine-oriented cognition sees itself as a form of computation and implements it using cognition-based computing methods.

An interdisciplinary technique for investigating human and animal cognition is called cognitive science. The mix of cognitive, data-based sciences and a variety of other such technologies has ushered in the field of cognitive computing. High-performance computers, artificial neural chips, ML, predictive modeling, NLP, etc are significant advancements for the computing field that are speeding breakthroughs in the cognitive computing fields.

There are various viewpoints on cognitive computing due to the interdisciplinary roots of cognitive science and information science. On exactly what falls under the umbrella of cognitive computing, there is no agreement. Big data, retrieval of information, ML, natural language comprehension, as well as their applications are primarily the driving forces behind the primary explanation of cognitive computing.

This paper's main objective is to highlight and delve into a comprehensive foundational understanding of cognitive computing by combining different viewpoints. It will give a general overview of the interdisciplinary field of cognitive science. The main technological enablers of cognitive computing are reviewed together with the standard properties of cognition-based computing systems and an overview of their applications.

Cognitive architectures simulate how people perform across a range of cognitive tasks. They are computational paradigms that govern the construction, interaction, and structure of cognitive systems. We will also go over the three main types of cognitive task architectures as well as their limitations. Further, trends and potential future study areas are also covered.

II. BACKGROUND

Theories from cognitive science offer a framework for describing diverse models of human cognition, including how the brain represents and processes information. Perhaps the most intricate system in terms of both structure and function is the human brain. The brain's mental functions include a wide range as well as visual and auditory perception. Cognitive tasks and brain functions are interchangeable phrases in our vocabulary.

Cognitive Computing Systems

The theories, methods, and tools of the computing field are utilized in cognitive computing to represent cognitive activities. It sees human thought-processing machine as a massively parallel processor of information, as well as employs a variety of models for data processing, and uses algorithms to perform data transformation and inference. The ways that information is represented and stored in computers differ greatly from how they are represented and stored in the brain.

Systems for cognitive computing are significantly different from those for conventional computing. Cognitive systems include context into the calculation, are interpretable, and learn how to improve through time. They perceive their surroundings, work independently, and handle things ambiguously as well as deal with incomplete/uncertain information.

Systems that are used in cognitive computing rarely use such grand methods. For instance, the IBM computer system termed Deep Blue which defeated Garry Kasparov, the grandmaster in chess, in 1997 has not been earlier regarded as a cognitive based computer system. When preparing its movements, Deep Blue conducted a thorough search. On the other hand, The IBM Watson is considered to be a system of such stature. It makes decisions based on incomplete and ambiguous data, leverages strong natural language understanding, and considers context while making decisions. It can distinguish statistical paraphrase of natural language text and perform spatial and temporal reasoning.

Systems for cognitive computing provide a wide range of capabilities. Current computing applications use various levels of cognitive capacities thanks to quick advancements in cognitive science and data science.

For self-driving automobiles, for instance, a variety of cognitive talents are necessary. Self-driving cars with cognitive capabilities can draw on prior knowledge and use contextual information to make judgments in the present hour.

Cognitive based assistant devices such as Google Now forecast as well as recommend new context-specific actions. Evolving tasks involving retrieval of information and search tools offer arguments for and explanations of their conclusions. The accuracy of cognitive technology for translating webpage information into other languages has never been higher. Real-time speech understanding and translation jobs make use of revolutionary advancements in speech recognition and language comprehension.

The brain is seen as an information processor in cognitive computing. So, in order to represent and transform information, appropriate representations are required. In reality, one of the difficulties in creating autonomous systems is how to express information and knowledge. Concept, proposition, rule, and analogy are the four subcategories of representations, according to Friedenberg and Silverman (2015). Concepts refer to things that are primarily of interest to the field, which include individuals, locations, as well as happening events. Natural language based tokens are effective illustrations of concepts. Propositions are claims regarding the subject matter. The foundational elements of propositions are concepts. Logic connectives can be used to unite propositions to create compound propositions.

Relationships between propositions are defined by rules. Rules make it possible to infer new information from data already available. The requirement to dedicatedly store information is lessened by certain rules. We term this category of rules as Reasoning rules. Integrity restrictions are a different category of rules. They serve to identify incompatibilities and check the accuracy of the information. Procedural knowledge, which is the third category of rules, refers to more intricate and abstract rules that outline the sequential actions needed to carry out cognitive activities.

An analogy is a comparison of two things, usually based on their structural similarities. Information concerning analogies is stored in analogous representations. Analogy oriented reasoning is used to solve problems using these representations. It basically suggests answers to the initial problem to be a solution of the next problem when there are problem circumstances which are comparable and one of the solutions is known. Solutions to analogous problems are frequently expressed by using a factor termed as certainty. Heuristic knowledge is the name given to knowledge that is conveyed by analogies. Another method of knowledge representation is an ontology. The names, concepts, and relationships between them in the domain are explicitly formalized in these specifications.

To explain the domain and support logic, they offer a standardized and formal language. They encourage the reuse of domain knowledge and make domain knowledge analysis possible. Domain assumptions can be made explicit with the aid of ontologies. According to Brewster and O'Hara (2004), we cannot represent information in some forms using ontologies.

Another category of representations known as distributed representations is applied in cognitive computing architectures based on neural networks. A weighted graph with direction where we have its nodes and edges arranged in a particular sequence. This graph is a neural network. This typically consists of two layers which are input and output. There may be many hidden layers stuck in between the nodes of the input layer. We may also have nodes of the output layer which can be directly connected. In a neural network, knowledge is represented by the weights at the edges. For us to grasp knowledge on edge weights, we must first train a network on inputs. Layers in a multilayered network are simultaneously and nonlinearly learned.

Distributed representations have the benefit of being more robust against noisy input and exhibiting more graceful performance deterioration. However, using the network's underlying structure to interpret the system's behavior is challenging. An explanation of the decision-making process is essential in applications like personalized medicine. A subset of machine learning called deep learning significantly depends on multiscale distributed representations. Multiple features are used to describe the input data where we have each feature that primarily shows features on different scale levels. We must take note on the fact that knowledge representations are utilized and the cognitive computing architectures have a close relationship.

III. ANALYSIS

Understanding the several intricacies pertaining to cognition and the average functioning of a human mind has long been the focus of research in the interdisciplinary field of cognitive science. In comparison, the field of computing is relatively new. But during the past few years, the field of computers has experienced revolutionary advancements. As a result, new and exciting prospects are opening for the development of data as well as cognitive science based research methodologies.

Big Data and Data Science: Unprecedented amounts of data are being produced as a result of recent advancements in technology pertaining to storage, intensive performance computing of the higher level, networks focusing on gigabits, and sensing of pervasiveness. Some of this is produced as streaming data that moves at fast speeds. Additionally, most of this data, which includes written and spoken documents, photos, and videos, is diverse and unstructured. Big data is the term used to describe this data, and several technologies have been created to store and retrieve it. Big data has made it possible for numerous cutting-edge new applications. An open-source software library called TensorFlow is used to create apps that focus on machine learning (TensorFlow, n.d).

Performance and Scalability

Applications for cognitive computing frequently work with vast amounts of unstructured data. Additionally, there may be issues with the data's quality. The information can be confusing, contradictory, uncertain, inconsistent, or incomplete. It is a very likely scenario that the data may sometimes contain redundant values. These can be extremely difficult to find. Computing costs are high when processing such forms of data which is unstructured in order to extract meaningful information. Basically, we'll be needing a large amount of processing power to clean as well as extract such data.

We can ideally describe the requirements of such cognitive based applications through three characteristics namely - performance, scalability, and elasticity. We have performance which basically is referring to the strict guidelines applied to the amount of time needed for data processing. Say, we have IBM Watson as our example. It is a cognitive application for answering questions that can interpret natural language. IBM Watson must respond to every query in under three seconds in order to participate in the Jeopardy! contest. The scalability parameter describes a computing system's capacity to function under a heavier load without requiring any software modifications.

Elasticity is the way where a computer automatically performs operations such as addition and deletion of various resources. These include but are not limited to CPU cycles, memory (primary) as well as secondary storage, to accommodate changes to avoid burden on the system. There is also the dynamic component which is termed crucial for allocating precisely the right number of resources in order to run a cognitive system at a given performance level despite changes which were not predicted in system workload as such. To keep a cognitive system's operational costs to a minimum, elasticity is essential.

Distributed Computing Architecture

These architectures are one method where cognition based computing systems satisfy all the necessary requirements pertaining to scalabilty as well as performance. These architectures have a number of processing nodes, each of which is a standalone computer with cores of computation, memory that is primary, as well as secondary storage systems.

Through techniques including shared memory as well as forwarding of messages, the exchange of information between nodes and the coordination of their actions in order to gain a standardized objective which is common. Through a quick computer network, the nodes are linked together. A collection of nodes is usually termed as a cluster. This cluster is logical in nature. A rack is something which has multiple such nodes that are physically mounted to it. Some cognition based computing systems utilize such node-based clusters that are spread across different data centers.

For applications using cognitive computing, a popular architecture in picture today is that of client-server. Application Programming Interfaces (APIs) or protocols like REST are used by servers to make services they deliver to clients available (Fielding, 2000). Usually, a network is used to connect the server and clients, which are on separate computers physically. On the other hand, the server as well as the clients could both be housed in physical ratio on the same computer. Here, we have both of them sharing the workload.

A cognition based application server often works on a running cluster in present day settings. The task of intercepting requests from clients and assigning them to worker nodes falls to a selected node known as the master. Here, this node primarily acts as a load balancer. Additionally, it oversees organizing the cluster as a whole's operations. Even though it makes the whole process of managing clusters simpler, the master node turns out to be the sole reason behind failing. A standby for this master node assumes control if the master node fails. Because the nodes are independent, the arrangements for master-worker and master-master are called as shared nothing architectures. In order to achieve performance at such scale, both topologies spread this data and processing throughout the cluster's nodes. To provide high availability, data is also copied to a portion of the nodes. Some systems permit the addition of new nodes or the removal of existing ones (either voluntarily or as a result of a node failure) without affecting service. Shared-nothing computing systems expand their network of nodes to handle growing workloads. It is simpler to test master-master architecture cognitive computing systems than master-worker architecture ones.

Cognitive Computing Architectures

The development of cognitive systems can be planned out using a cognitive architecture. In order to achieve mental functions, it describes fixed structures and relationships between them. The interactions between the structures are driven by the knowledge included in the design to provide intelligent behavior.

To go in contrast, we can say that a cognitive model is more narrowly focused on an isolated activity of the singleton, also termed as language acquisition. The interaction spanning cognition between activities like language comprehension and problem-solving is also studied using cognitive models. They are also employed to forecast behavior for tasks.

The structural components of cognitive systems are frequently the focus of cognitive designs. They impose limitations on the kinds of cognitive models that may be created. The constraints of cognitive architectures are also made clear by cognitive models. As a result, cognitive architectures and models interact significantly. The phrases termed as cognitive architectures as well as the one and only cognitive model are not consistently used in the literature. The intended meaning has to become clearer given a particular situation.

There is a lot of study being done on cognitive architectures. The qualities that cognitive architecture should offer in terms of depiction, organization, efficiency, and learning are listed in Langley's study. Additionally, it identifies unresolved research issues in cognitive architectures and provides the needs required for assessing cognition-based architectures. This whole process happens at the systems level. The usefulness that can be extracted out of the architectures which are already in existence for the sole reason of developing artificial general intelligence is discussed in this study. Research on cognitive architectures is shifting its emphasis from the practical capabilities of architectures. It is focusing more on how well we can simulate certain specifics of the behavior of human beings as well as its brain activity.

Cognitivist, connectionist, as well as hybrid cognitive architectures are the three main categories of cognitive architectures. The parts that follow talk about them. Basically, what we need to understand is that, all cognitive systems have cognition based architectures in its middle section, more precisely the center. By developing such a model, we can make a cognitive system possible.

Cognitivist Architecture

Cognitivist architectures use explicit symbolic representations to represent information. These representations represent external objects using an absolute ontology. Human designers create representations and then directly insert them into artificial cognitive systems. Symbolic architectures and Artificial Intelligence (AI) methods are other names for cognitivist architectures. Based on this architecture, cognitive systems are highly effective at resolving issues. They are not, however, sufficiently broad to be useful across areas.

Symbolic representations might bias the system since they showcase the designers in-depth understanding of the subject. Additionally, it can be challenging for designers to develop all necessary representations that are sufficient to actualize the expected cognitive behaviors of the system. A word's referent, or what it refers to, is not its meaning.

Connectionist Architecture

The processing of information that takes place in organic systems which are neural network-based serves as the inspiration for connectionist designs. A synapse can best be described as the connection of two neurons that have a tiny space between them through which signals are transmitted by the diffusion of a neurotransmitter.

Neurons, which make up the brain, are thought to number between 10 and 100 billion. Over 10,000 connections to other neurons are anticipated for each neuron. Through its incoming connections, a neuron takes in stimuli from other neurons. Using the stimuli, they have received, neurons carry out nonlinear calculations. Through its egress connections, this computation's result stimulates the activity of more neurons. Connection activation strength is measured numerically and modified to represent the level of network learning. This method of building architectures is known as a connectionist or emergent architecture.

Hybrid Architecture

The final group of cognitive computing architectures includes ones that combine connectionist and symbolic designs. There are also methods in this class that are neocortex inspired. This neocortex can best be described as the region of the central cortex also deemed as cerebral that controls the process of listening as well as sight. Neocortex uses a staggering amount of parallelism. Neocortex neurons carry out basic tasks. These traits are encouraging for the development of the neocortex on highly effective hardware. Additionally, we can also say that every part of the neocortex is responsible for performing both the sensory as well as the motor activities. This finding implies that integrating sensory-motor modalities is a key component of cognitive computing systems. Additionally, the brain's neocortex employs standard principles and algorithms for a variety of cognitive functions, including sight, hearing, speech, touch, and action.

This is ideal for creating cognitive systems that are not bound to certain domains and are inspired by the neocortex. We refer to these methods collectively as hybrid architectures.

Cognitive Computing Applications

Systems of cognitive computing have been applied to a variety of issues across numerous disciplines.

Tutoring Systems:

Tutoring systems that are intelligence based were the earliest forms of systems that are cognition based. We have three ACT-oriented models used in production that include - Algebraic, geometrical, and Lisp problem-solving strategies. Best case analyses revealed that students could complete tasks with these cognitive tutors in a third less time than they could with traditional instruction while still performing at the same level.

Answering Systems:

IBM Watson's development began with the IBM DeepQA. Its core technologies include the representation of knowledge as well as logic, ML, natural language interpretation, and the retrieval of information. Its primary capabilities are hypothesis development, evidence collection that spans over multiple sources of heterogeneous data, analysis of evidence, and scoring for all the hypothesis to exist. DeepQA technology powers these processes. The compute power was given by the network and the underlying, highly parallel hardware.

Healthcare:

An adaptation of Watson for the wellness and health care industries is IBM Watson Health. Watson began experimenting with individualized cancer therapies in 2011. The corporation Medtronic provides medical technology, services, and solutions. Medtronic created an insulin pump and a glucose monitor using IBM Watson Health. Diabetes patients' blood sugar levels are monitored by a glucose meter, and their insulin is automatically delivered via an insulin pump. Medtronic has accumulated over a hundred and twenty million days worth of anonymised data from glucose meters of patients. This data will be analyzed by Watson Health to find trends that may suggest novel diabetic therapies. Nvidia GPUs will be used by the Massachusetts General Hospital to identify irregularities in scans of CT.

Business:

Learning systems specifically created to mingle with people over a conversational aspect and comprehend natural languages can define businesses based on cognition. Cognitive analytics are advantageous to cognitive businesses as well.

Cognitive analytics generate various hypotheses, collect as well as evaluate evidence from various data sources. This is a method which supports each hypothesis as well as ranks them accordingly instead of using pre-existent rules and queries that are organized to recognize the data which gives rise to strong decision making.

When a theory receives a score over a predetermined threshold, the algorithm recommends it along with a numerical value indicating how confident it is. More data utilized to train learning algorithms results in better insights provided by cognitive analytics. IBM Watson variants are advertised as intelligent business solutions.

Human Computer Interaction:

Perspective-taking, which includes comprehending a particular idea or analyzing a state of a situation from a new point of view, is an extremely crucial component in human-robot interactions. Trafton et al. (2005) provide an example of how astronauts might function in a collaborative project setting by adopting a different point of view. They outline a Polyscheme cognitive architecture that can be used for taking perspectives. Additionally, this also creates a cognition based system that identifies with this particular Polyscheme architecture and combines it with an operational robot system. Many perspective-taking issues are successfully resolved by the cognitive system.

Autonomous Vehicles:

The navigation of autonomous vehicles depends heavily on cognitive systems. To detect their surroundings, they employ a variety of technologies. Modern control systems are used for analyzing certain sensory data to identify the various roadside objects. The main goal of self-driving automobiles is being actively pursued by projects like Google's Self-Driving Car and Mercedes Drive Pilot.

Media-Retrieval:

Retrieval across media is a difficult problem. Users can specify a query in any media to look for all the documents included in a variety of media sources. A document is any piece of media content, including text, a picture, an audio file, or a video. A cross-media query is one that uses a text-based search to find images. It is necessary to generate text that accurately represents an image's content in order to accommodate such searches.

Robots:

IBM SoftBank Robotics which basically provides platforms for robotics uses Watson. Pepper, which is a robot powered by IBM Watson, is being created by SoftBank Robotics. Pepper will receive assistance from Watson, which uses unstructured data that includes variants of sources such as audio, visual and texts.

IV. CONCLUSION AND FUTURE WORK

Research and application in cognitive computing will advance thanks to the convergence of cognitive computing and data science. To enable the development of extremely large-scale cognitive systems, cognition-based sciences will never stop in providing the theoretical foundations needed for augmented intelligence. Subjects such as data science will provide analytics with emphasis on cognition, and this particular discipline will bring forth large advancements in hardware technologies. The latter is crucial for developing the field because it uses empirical research to support cognitive theories.

The advantage of biological cognitive systems is that we do not fully understand their structure and operation to the point where we might mimic them in an artificial system. Cognitive computing systems, on the other hand, are not constrained by biological cognitive systems' memory space restrictions or their performance deterioration brought on by aging and overuse.

The line separating neural network approaches from symbolic AI approaches will keep blending. Rather than competing, they enhance one another. While neural network approaches use massive data and machine learning algorithms to develop better models, the approaches which are symbolic-oriented hold advantage in expressing certain logic. To extract thinking stages, neural techniques can use symbolic inference rules. On the other hand, employing semi-supervised and unsupervised learning, neural techniques can be utilized to create the inference rules which signify symbols.

Data is expected to rise at a 40% annual pace over the next ten years. The upcoming Airbus A380-1000 airliner, for instance, has the capabilities of accommodating roughly 1000 passengers.

It is going to have over 10,000 sensors on every wing. These can produce data over 8 TB of every day. The engines and other components of the airplane will contain a large number more sensors. We can use this data for a variety of purposes such as preventive maintenance, enhancing the comfort and safety of inbound passengers, and improving airport operating effectiveness.

Modern cognitive systems are still far from being able to perform cognitively on par with humans, but this difference is closing quickly. For instance, the Go game which gained popularity in recent days has long been titled as one of the most difficult topics for researchers of intelligent and expert systems. Go has a much larger search space than Chess, which makes it challenging to evaluate board positions and plays by raw force. In their article from 2016, Silver et al., they describe how the cognitive technologies of DeepMind were used to create AlphaGo. This outperformed the human who was a Go champion with a score of 5 games to 0. Additionally, AlphaGo outperformed other Go computer systems with a winning percentage of 99.8%.

To sum up, cognitive computing combines computing, data science, and cognitive science. Cognitive computing has the potential to improve people's lives or turn into a weapon for abuse and devastation.

This paper has identified the latest research being conducted on cognitive computing and how systems use the theory, application, and evaluation of augmented intelligence to understand how cognitive computing can be applied. Upon reviewing all approaches and by providing a benchmark understanding of how the neurobiological mechanisms of a human mind are used to train a system, we have identified and thoroughly understood the architectures of cognitive computing as well.

It has been understood that intelligence derived from cognitive platforms has increased efficiency and reduces the transmissions developing from cloud. This way we can improve the quality of results and help less trained staff members to handle data analytics better. With so much information to process, delegating a large portion of the work to machines will free up our time and help us use real-time analytics to make decisions right away.

V. REFERENCES

[1] Cognitive Computing: A brief Survey and Open Research Challenges

[2] Cognitive Computing: Concepts, Architectures, Systems and Applications

[3] Cognitive Computing: Architecture, Technologies and Intelligent Applications

[4] Bio-inspired cognitive architecture of episodic memory

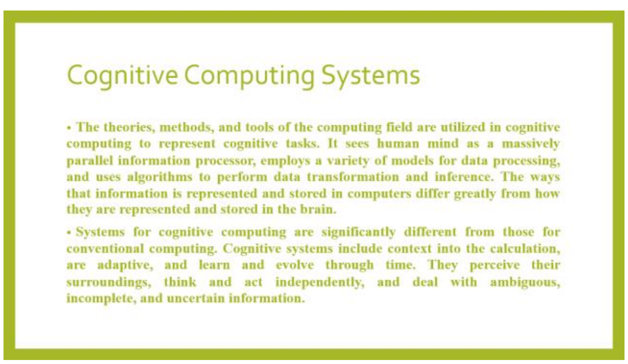
[5] Where to from here? On the future development of autonomous vehicles from a cognitive systems perspective

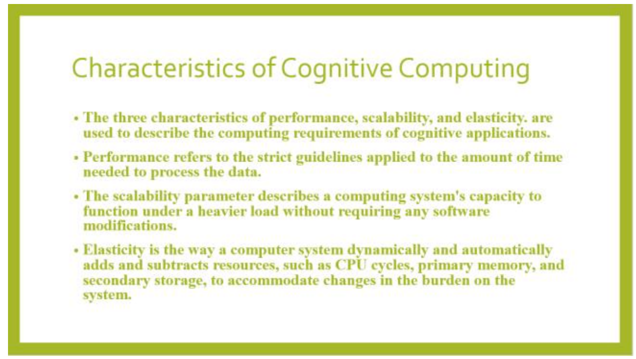
[6] Comparitive Analytical Survey on Cognitive Agents with Emotional Intelligence

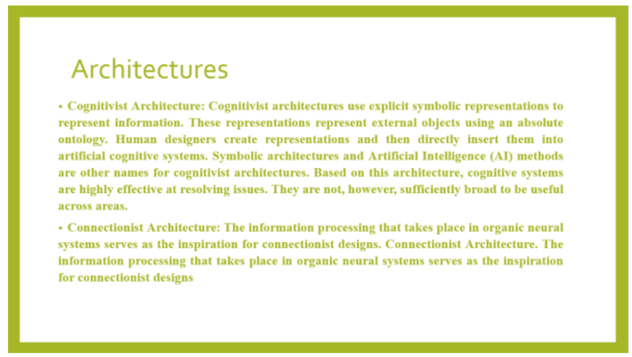
[7] Cognitive Computing: What's in for Business Process Management? An Exploration of Use Case Ideas

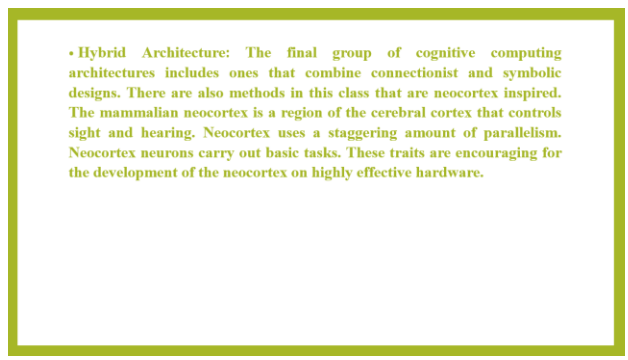
[8] Cognitive Computing Systems: Their Potential

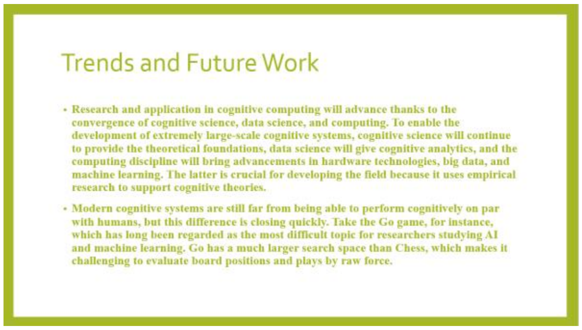
VI. APPENDIX

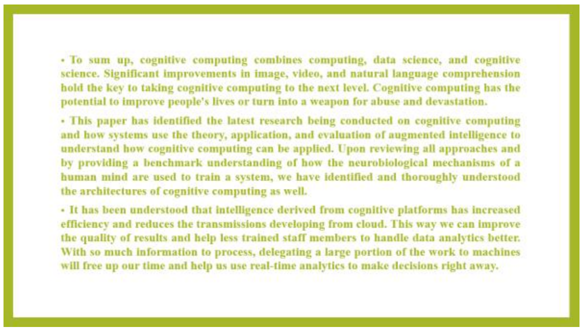









Trends and Future Work
• Research and application in cognitive computing will advance thanks to the convergence of cognitive science, data science, and computing. To enable the development of extremely large-scale cognitive systems, cognitive science will continue to provide the theoretical foundations, data science will give cognitive analytics, and the computing discipline will bring advancements in hardware technologies, big data, and machine learning. The latter is crucial for developing the field because it uses empirical research to support cognitive theories.
• Modern cognitive systems are still far from being able to perform cognitively on par with humans, but this difference is closing quickly. Take the Go game, for instance, which has long been regarded as the most difficult topic for researchers studying AI and machine learning. Go has a much larger search space than Chess, which makes it challenging to evaluate board positions and plays by raw force.


• To sum up, cognitive computing combines computing, data science, and cognitive science. Significant improvements in image, video, and natural language comprehension hold the key to taking cognitive computing to the next level. Cognitive computing has the potential to improve people's lives or turn into a weapon for abuse and devastation.
• This paper has identified the latest research being conducted on cognitive computing and how systems use the theory, application, and evaluation of augmented intelligence to understand how cognitive computing can be applied. Upon reviewing all approaches and by providing a benchmark understanding of how the neurobiological mechanisms of a human mind are used to train a system, we have identified and thoroughly understood the architectures of cognitive computing as well.
• It has been understood that intelligence derived from cognitive platforms has increased efficiency and reduces the transmissions developing from cloud. This way we can improve the quality of results and help less trained staff members to handle data analytics better. With so much information to process, delegating a large portion of the work to machines will free up our time and help us use real-time analytics to make decisions right away.