CHAPTER II

REVIEW OF RELATED LITERATURES AND STUDIES

A. RELATED LITERATURE AND STUDIES

I. Foreign Literature and Studies

Neural Networks as defined by Fiona Nielsen on his article what is "Neural Networks – algorithms and applications" (2010) is a field of Artificial Intelligence (AI) where we, by inspiration from the human brain, find data structures and algorithms for learning and classification of data. Many tasks that humans perform naturally fast, such as the recognition of familiar face, proves to be a very complicated task for a computer when conventional programming methods are used. By applying Neural Network techniques, a program can learn by examples, create an internal structure of rules to classify different inputs, such as recognising images.

Based on a review by Yann LeCun, Yoshua Bengio and Geoffrey Hinton about Deep Learning (2015) Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that

are used to compute the representation in each layer from the representation in the previous layer. Deep convolution nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shine light on sequential data such as text and speech.

Based on George Edward Dahl from his article "Deep learning approaches to problems in speech recognition, computational chemistry, and natural language text processing" deep learning approach to machine learning emphasizes high-capacity, scalable models that learn distributed representation of their input. This dissertation demonstrates the efficacy and generality of this approach in a series of diverse case studies in speech recognition, computational chemistry, and natural language processing. For speech recognition, he develops a more accurate acoustic model which uses rectified linear units and dropout, and improves error rates. In natural language processing, he describes a new restricted Boltzmann machine training suitable for text data. Then, he introduces a new neural network generative model of parsed sentences capable of generating reasonable samples and demonstrate a performance advantage for deeper variant of the model.

Comparative Study of Intelligent Personal Assistant by Dewi Agushinta R, Baby Lolita, Dwike Aprilia Setianti, Hardimen Wahyudi, I Putu Partadiyasa user can only communicate with computer by typing and reading. Many can only dream to just have to speak to computer and it will understand the user's

command and the computer give answer by speaking like it is also an intelligent being. These dreams began to appear as an achievable goal with years of advances in computer programming. Natural Language Processing (NLP) is the field in computer science that helps to solve the problem to make computer understand human language.

According to the study about Neural Networks by Anh Nguyen, Jason Yosinski and Jeff Clune (2015) since Deep Neural Networks (DNNs) have recently been achieving state-of-the-art performance on a variety of pattern recognition tasks, most notably visual classification problems. Given that DNNs are now able to classify objects in images with near-human-level performance. A recent study revealed that changing an image (e.g. of a lion) in a way imperceptible to humans can cause a DNN to label the image as something else entirely (e.g.) mislabelling a lion a library). Here they sow a related result: it is easy to produce images that are completely unrecognizable objects with 99.99% confidence (e.g. labelling with certainty that white nose static is a lion).

Based on the article "Lessons in Neural Network Training: Overfitting May be Harder than Expected" by Steve Lawrence, Lee Gies and Ah Chung Tsoi (1997), Neural networks are one of the most popular AI machine learning models, and much has been written about them. A common belief is that the number of parameters in the network should be related to the number of data points and the expressive power of the network. The results in this paper suggest

that the characteristics of the training algorithm should also be considered because Neural Networks and other AI machine learning models are prone to "overfitting"

The study, "Neural Networks and the problem of overfitting" by Nitish Srviastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov states Deep Neural Networks contain multiple non-linear hidden layers and this makes them very expressive models that can learn very complicated relationships between their inputs and outputs. With limited training data, however, many of these complicated relationships will result of sampling noise, so they will exist in the training set but not in real test data even if it is drawin from the same distribution. This leads to overfitting and many methods have been developed for reducing it.

A similar study, "Descriptive examples of the limitations of Artificial Neural Networks applied to the analysis of independent stochastic data" by Henry Navarro and Leonardo Bennun. An Artificial Neural Network (ANN) is a mathematical model inspired by the biological behaviour of neurons and by the structure of the brain, which is used to solve a wide range of problems. A network is reached by connecting several neurons with a specific architecture (Hopfield Networks, Kohonen Networks, Perceptron, and so on), in which neurons learn through a process of self-organization. During the learning process

of the ANN, when a data is introduced into the network, just the neuron that has a positive activity inside the proximity will be activated at the exit stage.

According to the article " I Know Everything About You! The Rise of the Intelligent Personal Assistant" by Gavin Kenny (2015). IPAs have complete access to your electronic life and an ability to undertake tasks autonomously, and the abilities are not within your immediate control. The task of accurately decoding voice and understanding what is being said is 100 times more processing intensive than a simple web search. IPAs send a recording of what you have said to a data center where the application actually resides. As a result, everything you ask of your intelligent personal assistant is stored and processed outside of your immediate control.

Based on "Object Recognition and Detection by Shape and Color Pattern Recognition Utilizing Artificial Neural Networks" by Jerome Paul N. Cruz, Ma. Lourdes Dimaala, Laurene Gaile L. Francisco, Eerica Joanna S. Franco, Argel A. Bandala and Elmer P. Dadios Humans recognize a multitude of objects and images with little effort, despite the fact that theimage of the objects may vary in different viewpoints,sizes or even when they are rotated. For instance, when a human sees a hand regardless whether it's open ornot, he still recognizes it as a hand. Object recognitionhas been man's daily lifestyle. The more often a person sees an object the more he gets familiar with it.In object recognition, it was known that thehuman brain processes visual information in semantic

space mainly, that is, extracting the semantically meaningful features such as line-segments, boundaries, shape and so on. As for computers, it is a challenge than what humans regard as the simple process of recognition. Numerous attempts are done to realize object recognition such as in that used the principle of visual vocabulary. Another approach in is introduced, wherein fuzzy logic is used to detect an object based on its color. Finally, in a combination of artificial neural networks and fuzzy logic is utilized to evaluate an object based on a visual and audio information.

Artificial Intelligence as defined by Alan Copeland on his article what is "Artificial Intelligence?" (2010) is usually defined as the science of making computers do things that require intelligence when done by humans. AI has had some success in limited, or simplified, domains. However, the five decades since the inception of AI have brought only very slow progress, and early optimism concerning the attainment of human-level intelligence has given way to an appreciation of the profound difficulty of the problem. Artificial Intelligence are designed to perceive human mind or cognitive functions. It is designed to think as something stated. It response as if you are communicating with someone. Today, artificial intelligence is widely used in medical diagnosis, robotics, games and etc. It is commonly used because Artificial Intelligence is very useful in their respective fields. Mainstream thinking in psychology regards human intelligence not as a single ability or cognitive process but rather as an array of separate

components. Research in AI has focused chiefly on the following components of intelligence: learning, reasoning, problem-solving, perception, and language-understanding. **Learning.** Learning is distinguished into a number of different forms. The simplest is learning by trial-and-error. For example, a simple program for solving mate-in-one chess problems might try out moves at random until one is found that achieves mate. The program remembers the successful move and next time the computer is given the same problem it is able to produce the answer immediately. **Reasoning.** There has been considerable success in programming computers to draw inferences, especially deductive inferences. However, a program cannot be said to reason simply in virtue of being able to draw inferences. Reasoning involves drawing inferences that are relevant to the task or situation in hand. One of the hardest problems confronting AI is that of giving computers the ability to distinguish the relevant from the irrelevant. **Problem-solving.** Problems have the general form: given such-and-such data, find x. A huge variety of types of problem is addressed in AI. Some examples are: finding winning moves in board games; identifying people from their photographs; and planning series of movements that enable a robot to carry out a given task. **Perception.** In perception the environment is scanned by means of various sense-organs, real or artificial, and processes internal to the perceiver analyse the scene into objects and their features and relationships. Analysis is complicated by the fact that one and the same object may present many

different appearances on different occasions, depending on the angle from which it is viewed, whether or not parts of it are projecting shadows, and so forth. **Language Understanding.** A language is a system of signs having meaning by convention. Traffic signs, for example, form a mini-language, it being a matter of convention that, for example, the hazard-ahead sign means hazard ahead. This meaning-by-convention that is distinctive of language is very different from what is called natural meaning, exemplified in statements like 'Those clouds mean rain' and 'The fall in pressure means the valve is malfunctioning. The Artificial Intelligence has come a long way from the old days. It was with the invention of the computers that the Artificial Intelligence method began to manoeuvre researchers. The technology was finally available and seemed to stimulate intelligent behaviour.

According to the study SPIED: Stanford Pattern-based Information Extraction and Diagnostics conducted by Sonal Gupta Christopher D. Manning, Department of Computer Science Stanford University (2012) entity extraction using rules dominates commercial industry, mainly because rules are effective, interpretable by humans, and easy to customize to cope with errors (Chiticariu et al., 2013). Rules, which can be hand crafted or learned by a system, are commonly created by looking at the context around already known entities, such as surface word patterns (Hearst, 1992) and dependency patterns (Yangarber et al., 2000). Building a pattern based learning system is usually a repetitive process,

usually performed by the system developer, of manually examining a system's output to identify improvements or errors introduced by changing the entity or pattern extractor. Interpretability of patterns makes it easier for humans to identify sources of errors by inspecting patterns that extracted incorrect instances or instances that resulted in learning of bad patterns. Parameters range from window size of the context in surface word patterns to thresholds for learning a candidate entity. At present, there is a lack of tools helping a system developer to understand results and to improve results iteratively. Visualizing diagnostic information of a system and contrasting it with another system can make the iterative process easier and more efficient. For example, consider a user trying to decide on the context's window size in surface words patterns. And the user deliberates that part-of-speech (POS) restriction of context words might be required for a reduced window size to avoid extracting erroneous mentions.

Based on the book "Speech and Language Processing." by Daniel Jurafsky & James H. Martin (2014), The HMM is a sequence model. A sequence model or sequence classifier is a model whose job is to assign a label or class to each unit in a sequence, thus mapping a sequence of observations to a sequence of labels. An HMM is a probabilistic sequence model: given a sequence of units (words, letters, morphemes, sentences, whatever), they compute a probability distribution over possible sequences of labels and choose the best label sequence.

The study, "Flite: a small fast run-time synthesis engine" by Alan W Black and Kevin A. Lenzo (2011) states and defines Flite as a small, fast run-time synthesis library suitable for embedded systems and servers. Flite is designed as an alternative run-time synthesis platform for Festival in applications where speed and size are important. Voices built using the FestVox process may be compiled into efficient representations that can be linked against Flite to produce complete text-to speech synthesizers. The Flite library is much faster and much smaller than the equivalent Festival system. This paper describes the motivation and the basic structure of the library, and gives figures of its size and speed. Some intended enhancements are also discussed.

A similar study, Evolutionary Algorithms in Natural Language Processing Lars Bungum, Bjorn Gamback Department of Computer & Information Science Norwegian University of Science and Technology (2010) simply defines natural language processing as the automatic treatment of natural languages (as opposed to constructed) deals with a moving target. The use of language varies across domains (application areas and sublanguages) and environmental factors (who speaks to whom and where), as well as being part of a continuing development and change (we even speak a different language from the one spoken in our childhood). Language evolves. Accordingly, the idea of evolutionary algorithms is a conceptual good fit with language technology as a notion of continuous evolution is in essence a requirement for a fully descriptive account

of language. A stale grammar representing a language at a point in time is useful, but not a full description of how this language functions, as new elements are constantly added to it.

A Short Introduction to Text-to-Speech Synthesis by Thierry Dutoit (2011) discusses that a Text-To-Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. Let us try to be clear. There is a fundamental difference between the system we are about to discuss here and any other talking machine (as a cassette-player for example) in the sense that we are interested in the automatic production of new sentences. This definition still needs some refinements. Systems that simply concatenate isolated words or parts of sentences, denoted as Voice Response Systems, are only applicable when a limited vocabulary is required (typically a few one hundreds of words), and when the sentences to be pronounced respect a very restricted structure, as is the case for the announcement of arrivals in train stations for instance. In the context of TTS synthesis, it is impossible (and luckily useless) to record and store all the words of the language. It is thus more suitable to define Text-To-Speech as the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter.

II. Local Literature and Studies

According to Philippine Star: "Artificial Intelligence by Boo Chanco (2015) the fast pace of technological change is the hallmark of the world we live in today. Cell phone models are changing faster than we can learn how to use them. Indeed, you can ask Siri almost anything and Siri has an answer. Siri isn't a person but a built-in "intelligent assistant" that enables users of Apple iPhones and iPads to voice commands in a normal way of speaking to get information, etc. We have put a lot of faith in the computer and mostly, the trust is deserved. Computers have taken over many boring or repetitive tasks that are subject to human error. But is the computer infallible? Not at all. Computers fail often enough and with devastating

Additionally, based on the article published by Manila Bulletin, "The Amazing Reach of Artificial Intelligence" by Beth Day Romulo this 2016. Since IBM's Watson computer was born in 2011, interest has been aroused, not only in the tech community, but also from the general public. Venture capital investors have poured money into Artificial Intelligence (AI) start-ups and large corporations, such as Google, Microsoft, Apple, and Facebook have been buying AI companies. The investment in AI companies reached $8.5 billion last year. The team AI has been used in science fiction, about machines that are able to think for themselves. The automated voice on your smart phone is a type of AI. So are

features in Google's search engine. Some Silicon Valley experts say people overestimated what can be done with AI in three years, but under-estimated what it will be able to do in ten years. Today's software can recognize images, translate languages, and understand speech.

Based on the study "FILIPINO- MARANAO BI-DIRECTIONAL LANGUAGE TRANSLATOR WITH TEXT-TO SPEECH SYNTHESIZER" by Ayson Q. Fallan, Mohammad Ammar T. Rasuman, April V. Sacil and Illuminada Vivien R. Domingo, DBA (POLYTECHNIC UNIVERSITY OF THE PHILIPPINES) (2013) indicates that early attempts to machine translation to Philippine Languages such as IsaWika! established initial computational linguistic resources for subsequent efforts. The formal grammar established from IsaWika! Indeed, became a stepping stone, as claimed by the authors of said project, to later efforts to machine translation or language technology, in general, using Philippine Languages (specifically, Tagalog). The main distinction of MT systems is in terms of overall strategy: whether translation from SL to TL takes place in a single stage (direct translation), in two stages (via an _interlingua '), or via the _transfer 'approach, where translation proceeds in three stages. Machine translation, using the transfer approach, generally follows different phases: morphology, syntax, and semantics. Morphology refers to the study of the structure of words or how words are formed. Syntax deals with how words can be combined together to

make larger phrases, such as, sentences. Semantics deals with real-world knowledge or the meaning of the sentence. Apelado, et al. stated that Computational linguistics in the Philippines is currently focused on Tagalog using the LFG framework. Their study showed that not much has been done on other Philippine languages with respect to the computational aspects of these languages towards a multi-lingual machine translation system.

The study "Emotional Techy Basyang: An Automated Filipino Narrative Storyteller" by John Christopher P. Gonzaga, Jemimah A. Seguerra, Jhonnel A. Turingan, Mel Patrick A. Ulit, and Ria A. Sagum (2014), a similar study that implements speech synthesis states that Concatenative synthesis was the algorithm used in the TTS process wherein every speech audio that represents each syllables will be concatenated to each other with some pauses for speech turns and delimiters. Also, the researchers used N-Gram in order to syllabicate every word in the story input. This study is aimed to determine its acceptability to the users and its accuracy in terms of its audio modification. The accuracy of the said application is measured by precision and its error rate.

B. SYNTHESIS

The reviewed literature studies are related to the present study that artificial intelligence and machine learning is now becoming a staple of the modern computing and development and is considered as one of the solutions in breaking the technology barrier and to promote a new and effective way of personal computing through natural human interactions and learning without any restrictions.