ARTIFICIAL INTELLIGENCE

CS 541

WINTER 2022

A REFLECTION

ON

ARTIFICIAL INTELLIGNCE: A GUIDE FOR THINKING HUMANS

(Author – Melanie Mitchell)

By

Ajay Babu Gorantla

This book *"Artificial Intelligence: A Guide for Thinking Humans"* authored by *Melanie Mitchell* is an excellent source for to know about the history of AI, the current phase we are in the progress of AI, what the future prospective of AI is, and what the goal is in broad. Reading this book has been truly a great and thought-provoking experience. This report summarizes the main ideas of the book and an honest reflection of my thoughts that I had subconsciously while reading the book.

The foundations of AI can be traced back to a small workshop in 1956 at Dartmouth College organized by a young mathematician named John McCarthy. In fact, the dream of creating an intelligent machine that is as smart or smarter than humans, is a centuries old common desire that became part of the modern science with the rise of digital computers. Over a brief period(years), there were some developments in the field of AI broadly in sense of the approaches toward AI. There were broadly two approaches that came up namely: Symbolic AI and Subsymbolic AI. As these were in their nascent stages during the early 1960s, people from across the academia followed the motto of "because we don't deeply understand intelligence or know how to produce general AI, rather than cutting off any avenues of exploration, to truly make progress we should embrace AI's 'anarchy of methods'".

As mentioned in the book, A symbolic AI program's knowledge consists of words or phrases (the "symbols"), typically understandable to a human, along with rules by which the program can combine and process these symbols to perform its assigned task. One of the early examples of Symbolic AI was General Problem Solver (shortly known as GPS). Unlike the symbolic AI, the subsymbolic AI is inspired from neuroscience and tries to capture the underlying unconscious thought processes. An early example of this approach is the brain

inspired AI program called Perceptron invented by Frank Rosenblatt. This invention of the perceptron algorithm is one of the founding stones of the current day deep neural networks. However, the book published by Minsky and Papert proved that Rosenblatt's version of perceptron algorithm can solve only those problems that are linearly separable. They went onto provide a remedy for that i.e., by adding an additional layer (called "hidden layer") in between the input and output layer calling it a "multilayer neural network" and but they couldn't come up with a learning algorithm for this type of network. Their speculation on "Multilayer Neural Networks" was in a way, a final nail in the coffin for Rosenblatt's perceptron model. This later, led to a brief period of "Winter" in AI famously called "AI Winter" due to lack of a learning algorithm until the late 1970s when the "backpropagation algorithm" was discovered. This algorithm is a lifeline and proved to be a major propeller of current day deep neural networks.

These algorithms or broadly the subsymbolic AI has led to major advancement in field of Image Recognition. One of the typical use cases is image classification for example the classification of handwritten digits. These tasks are done using the multi-layer neural network and there is significant success in terms of accuracy. The efficiency and performance of these networks greatly relies on hyperparameters like learning rate, number of hidden units in each hidden layer, and number of hidden layers. Finding the optimal set of hyperparameters is itself a separate challenge. All this looks good, but there are some questions that arise when looked comprehensively. It was observed that the system trained on some labels classifies a picture of random noise to be one of the labels, which is weird and can be concerning in grand scheme of things. It raises us a question of what exactly these algorithms are learning. Games have been in existence since the computers with visual

displays came into existence. Even though they are meant to be for the purpose of entertainment, it relates to a fundamental part of a human's thought process. Hence, applying the developments of AI to the games has been an obvious choice to better understand or demonstrate the performance of AI. Reinforcement learning is a typical go to method for training an AI to play games. It broadly works on the principle of rewarding the AI for good behavior (good choices and winning) and penalizing the AI for bad behavior (bad choices and losing). Q-Learning is a prominent method in implementing reinforcement learning. One of the moments in AI that is considered as a breakthrough was when DeepMind came up with AlphaGo that defeated the world champion Lee Sedol in the one of the most challenging board games 'Go' in 2016. AlphaGo relies on a combination of Monte Carlo Tree Search, reinforcement leaning, deep convolutional neural networks. Although this sounds great, definitely not downplaying the achievement, we must understand that this algorithm only relates to the game of 'Go' and can't be used for another application. The concept of transfer learning can't be used with this algorithm because not all problems can be represented in states as done for the 'Go' game.

Humans communicate using a language. One of the main goals for AGI (Artificial General Intelligence) is that computers understand human language. The study of languages combined with computers is nothing but the field of Natural Language Processing (NLP). There has been a lot of development in the field of NLP. Many problems language translation, sentiment analysis etc. have seen a great progress. This progress has been made through usage of statistical models and different ways of data representation. However, one issue that remains is that current AI systems don't truly understand or sense the language as we humans do.

While reading the book I had many thoughts running subconsciously. I have recently watched the documentary of AlphaGo, and I felt a myriad of emotions happy, sad, and anxious. As mentioned in the *Prologue* by the author, during her interaction with Hofstadter after the meeting at Google, the latter said, "If such minds of infinite subtlety and complexity and emotional depth could be trivialized by a small chip, it would destroy my sense of what humanity is about." Although I was happy about the progress of AI seeing the documentary, I had and have the same fear what Hofstadter had. Although Kurzweil predicts that Singularity will be achieved by 2029, reading this book in a sense gave me a hope that the singularity is not that near as he predicted. I believe that the human intelligence is lot more than what we ourselves as humans understand or perceive. I agree without doubt that there has been significant progress in the applications of AI in various fields but I would say each of these developments are narrow in those respective fields.

The Ultimate goal is to create an "Artificial General Intelligence" (AGI) that can feel and sense things around it as we humans do and understand and sense the result of its actions as we humans do. The day this is achieved, would be the day Singularity is achieved and the day we humans will no longer be the most intelligent (in all senses) species in the world. I in a way am still not able to collect my thoughts together if whether the AGI would be a 'good' to this world, and seeing the progress, the patterns, I believe that singularity can be achieved in many decades from now, if not by 2029 as predicted by Kurzweil with his "exponential progress" theory. This leaves us with the big responsibility of deciding/regulating the usage of AI for the greater good of human civilization.