**Writing Assigment – The Cambridge Handbook of Artificial Intelligence**

Gustavo Luís Luchetti, 9871969

The Cambridge Handbook of Artificial Intelligence begins by proposing that to better understand the **history behind AI**, we should first investigate its points of contention, that is, things that were controversial when they were first proposed, and still are widely discussed. Personally, I find that fascinating, because an area of study which constantly debates and revises existing frameworks already has a thriving community that looks forward for future and past discoveries. Some of those points are symbolic AI vs neural nets, brain in a vat vs embodied AI, which will be discussed with more detail in the next paragraphs. Major areas in AI are introduced, such as knowledge representation, expert systems, machine vision, machine learning, natural language processing and many others. Also goes in depth about recent trends and directions of AI, like data mining, cognitive computing and science. Finally, it goes back to those initial points of contention and reaffirms that they all still have their place in the discussion of AI, with different degrees of importance. Again, that is surprising, but given the wide range of different point of view working with AI, it is, in a way, expected to be this much attrition between those topics.

'Artificial Intelligence was built on concepts and theories developed by philosophers and logicians'. To put it simply, a program has AI when it solves a task that would necessitate intelligent behavior for it to be solved. Those are fundamental concepts in the roots of AI inside philosophical discussions. I find it really interesting the ongoing comparisons between AI and the way our own mind works. A classical conflict is also mentioned, between viewing AI as a science to better understand the mind and using AI solely as an engineering tool. Two important sections/date are mentioned as well, cognitive revolution, which started in the 1950s, and the theory of computability. The Future of Psychological A.I. is extremely difficult to predict, some philosophers proposed that we would have fully Strong A.I. by now, and some thought that it would never be possible. Personally, I think we are still a few decades away, but I’m a firm believer that it will one day be possible to replicate our own mind with a computer.

As the last chapter explored, philosophy has always been an essential part of A.I. Some authors argue out that there need to be a material connection/existential connection for something to have human reason. The Chinese Room Argument gains a lot of tractions, to disprove the possibility of Strong AI. The argument is explained as such: a man inside a room receives squiggles, the man has a rulebook to which he can translate the squiggles into answers, and then he delivers those answers. Actually, the man was receiving questions written in Chinese and translating them to their respective answers. It is argued that, that man, no matter how long he stays in the room, we’ll never truly learn Chinese, therefore, an agent, even with the capabilities of translation, does not truly know Chinese. Those type of thought exercises are very interesting, although I don’t think they are useful besides that, being exercises. There are a lot of contextual clues that are needed to define what is strong or weak AI, and I don’t think a single argument can fit every scenario.

Good Old-Fashioned Artificial Intelligence,or **GOFAI** for short, is a label for classical, symbolic AI. Its key concepts are heuristic search and planning. Given these notions, it can be said that GOFAI usually attempts to simulate high-level human thought. A branch of GOFAI evolved into evolutionary programming, that is a program that gets random changes in its own rules, just like cells mutating. GOFAI doesn’t come without its issues, it is so great at representing hierarchy and order, that as soon as there is an option that is ‘outside the box’, the system fails to grasp it. Even with all those issues, it should be mentioned that GOFAI hasn't really failed, it has failed if we were to base its success on extremely hopeful promises, but it has birthed innumerous useful technical applications, especially expert systems. Psychologically however, it still has a way to go; it’s extremely precise and sometimes even crass way of thinking still can’t capture all the nuances of the human mind. For me, GOFAI really shines in the fact that despite being a lot simpler than newer frameworks of AI, it has still managed to produce fantastic results, so it has certainly established itself as a pillar in the area.

Moving on, the handbook explores **connectionist models**, models that consist of networks of processing units interconnected through connectivity patterns of various kinds. Some interesting properties of the models are parallelism, adaptivity, graceful degradation, spontaneous generalization, etc. There are 3 types of ways those models can ‘learn’: supervised learning, where the program knows what the desired outcome is, and gets an error signal if it isn’t achieved. Unsupervised however, does not have the same error signal, so it is best used to finding the best model to fit the data. Finally, reinforcement learning doesn't get an error signal, but a 'good/bad' performance, if the output desired was close (or not) to its goal, so the desired behavior can be reinforced through repeated training sessions. Hybrid connectionist models are mentioned, those are the synthesis of connection and traditional symbolism. However, several issues arise, not only are these models much more complex, but there are also a lot of decisions that must be made as to construct a proper hybrid model for a given task. Therefore, hybrid models have currently only given limited results. Even with those challenges in mind, certainly one of the hybrid models of A.I., be it these or the ones mentioned in the next chapter, SEDs, those are the ones that I think will shine when we attempt to build truthfully strong A.I.

Finally, the **SED Framework**, that is, an agent that is always aware of where it is Situated (knows its own environment), of its (Em)body and its physical properties, and Dynamical brain (that is, its own internal dynamic state). All those aspects are extremely important to guide the agent’s possible actions. Situation is important as too narrow down the field of actions an agent can execute. Embodiment on the other hand, has three different advocates: physical, its material properties and capabilities, biological and conceptual, our lived-through experiences and metaphors. I find that very interesting, since this framework really attempts to replicate the different areas of feeling and behavior that we have as humans. Lastly, there is the Dynamic section, which has a few ideas as well, including dynamical systems theory, mostly mathematical theories, dynamical framework and dynamical hypothesis. These are all valid, although, specific claims of how we define what actions we take, but it really shines when all the frameworks are combined; 'behavior is a property of the entire coupled brain-body-environment system'. These frameworks are not immune to criticism however, being reactive means it is intrinsically sensory based, same input resolves in same output. It has been extremely difficult to develop these so-called SED systems, given that we lack the tools necessary to analyze our own's body unimaginably big net of inputs, environment, body, biological functions, etc. Even so, SED already shows great strides in developing what we know about cognitive science, although the framework itself has had much less time to develop than other frameworks in AI (like computational and connectionist, for example.).

**- Most interesting and/or noteworthy insight, what did I learn?**

**- What surprised me? what did I manage to take away from it?**