Gallatin Undergraduate Rationale: The creation of Artificial Intelligence

Abhi Agarwal

Artificial Intelligence has been a hotly debated topic in academia in the last couple decades, and even more debated in the media in the last couple years. A lot of very large and powerful organizations and universities have bet for the success of this field, and many have claimed that the next age of computation will be powered by Artificial Intelligence techniques. The creation of Artificial Intelligence aims to deal with different aspects around the definition, creation, and impact of creating intelligent machines.

What is Artificial Intelligence?

To explore what Artificial Intelligence is, it is important to study components that would make it up. In the book 'Superintelligence: Paths, Dangers, and Strategies', Bostrom lays out three general features that this system would require. First is the capacity to learn, second the "ability to deal effectively with uncertainty and probabilistic information [and] extracting useful concepts" (Bostrom, 23), and lastly "for leveraging acquired concepts into flexible combinatorial representations for use in logical and intuitive reasoning" (Bostrom, 23). The system he proposes allows it to iteratively improve and develop like the mind of a child. This system is able to learn and make decisions based on logic and reason it develops on its own rather than having it be hard coded into it, which means that it starts off with a blank slate. Russell and Norvig, in their textbook 'Artificial Intelligence: A Modern Approach', write "[learning] allows the agent to operate in initially unknown environments and to become more competent than its initial knowledge alone might allow" (Russell and Norvig, 55).

Translated into fields in Artificial Intelligence, this system would, at first, constitute of the fields: natural language processing (to communicate), knowledge representation, automated reasoning, and machine learning. The commonality between these sub-fields is that they are all trying to build and focus on different aspects of intelligence, and so we can deduce that the key aim of Artificial Intelligence is to build intelligent entities or intelligent systems[1]. Furthermore, a system that exhibits signs of 'intelligence' is said to be artificially intelligent. There are degrees to how artificially intelligent a system is. A system can be weakly intelligent, strongly intelligent, or super intelligent.

[1] Using the word system rather than machine as machines usually either depict either physical servers or shiny mental robots, and system has a more generalist tone to it.

A system that is weakly intelligent specializes in one area. The classic example is Deep Blue, which was a system that could only play chess. It was incredibly good at playing chess, but if one was to ask it the weather, it would not be able to compute that. Weak AI is as good at performing a single task as a human mind. A system that is strongly intelligent is as intelligent as a human mind in every possible way. This system is as socially skilled, able to reason and make judgements, make plans, and be scientifically creative as a human is. Moreover, a system that is super intelligent is "a system that can do all that a human intellect can do, but much[2] faster" (Bostrom, 53). This is the system which most researchers are fearful of (or excited by); this system is able to learn and make decisions orders of magnitudes faster than a single human-mind, but the amount by which it is intelligent can be either just a little smarter or millions of times smarter.

The systems that currently exist are weakly intelligent – also called narrow intelligence.

What is intelligence? How do we define and measure it?

Intelligence has been defined in many different ways, and different definitions of intelligence allow us to quantify or understand it in a different way. The simple act of telling an individual how intelligent she is, is the most basic act of quantifying intelligence, which we do in our day-to-day lives. By stating that a particular individual is intelligent we could mean several different things. For example, we could be commenting on their ability to gain knowledge at speed, their accomplishments, their society/community group, their ability to reason, and much more. These are certain characteristics that we, as a society, think about when making a judgement about an individual's intelligence. The most intelligent individuals partake in this by accepting awards for their high IQ or even accept genius grants; even those who may not be as intelligent participate in this by observing this phenomenon and by discussing it. Therefore, there is an inherent part of our society and the way we perceive of the world that needs to compare or judge intelligence.

Defining and measuring intelligence become more popular when Alfred Binet and Theodore Simon, in France, designed the first wide-used intelligence test known as the Binet-Simon Scale. Binet personally believed that intelligence is too broad a concept to quantify with a single numerical value. However, he did agree that intelligence is influenced by a number of factors, and can be compared if broken down into its parts. Intelligence is a construct that we have created to define certain characteristics that an individual has, but it is a construct that is slowly being supported by scientific evidence.

Intelligence testing & factors of intelligence for intelligent machines

In 1916, the Binet-Simon Scale was brought to Stanford University and researchers adapted it to become the Intelligence Quotient or IQ. The reason for the U.S. to create an intelligence test, as we read in class, was to screen army recruits during World War 1. In addition, IQ tests were

[2] Where the system is multiple orders of magnitude faster.

also used to screen immigrants as they arrived at Ellis Island, and became increasingly useful to governments as the century progressed.

The factors that we will consider to cover intelligence are: problem-solving, knowledge, reasoning, and planning, uncertain knowledge and reasoning, learning, and communicating, perceiving, and acting.

Conscience, desire, rationality, intuition, common sense, and their relation to intelligence

Methods of quantification for a thinking or intelligence machines

Embedding of AI into our social structure

De Waal, in his book ‘Primates and Philosophers’, points out that there are aspects of morality that are unique to human beings. He notes that the ability to weigh, reason and judge two separate moral decisions and choose an outcome is one of them. The other is the ability to be impartial and spectate a situation. In addition, there are also fundamental differences in our society and biologically. We have evolved to a point where we don’t necessarily form social groups in order for survival, but for cultural interests, religious interests, etc. We also differentiate in the ability to communicate through an established language. This is an important aspect that helps us in making our moral decisions. The biggest differentiator between human beings and animals, to me, is the idea that we’re able to express things we’re thinking about in a formal and clear manner. We’re able to have a discourse about the disagreements we have, and debate whether those disagreements are valid or not. These aspects are critical to understanding the evolution of our morality.

De Waal manages to establish and solidify the idea that there was an evolution of morality, and the roots of morality can be seen in primates. To me, thinking about the evolution of morality is intriguing. We’re attempting to build machines that can think and potentially machines that will walk among us. Soon we will progress into developing artificially intelligent machines. This raises a question, can morality evolve to being implemented digitally? How could morality evolve to thinking machines as it did from primates to us?

There exists a thought problem in the field of Artificial Intelligence that makes this question worth exploring. Hypothetically, I order an AI to build paper clips. The AI would obey its creator and would begin creating paper clips. Eventually, it would turn the entire universe into paper clips. The reasoning behind this is that the AI doesn’t have an understanding of limitations like we do. An AI only knows its task and that it has to keep working on that task until it reaches a goal. I believe that our understanding of limitations stands from our morality, and our knowledge of what is right and wrong. How can we extend this moral behavior into thinking machines?

In the Aristotelian framework ’virtue’ is moral responsibility and represents the bestowment of praise or blame. To qualify for this bestowment of praise or blame must be done voluntarily by the agent. In Nicomachean Ethics, Aristotle writes, “virtue is about feelings and actions.

These receive praise or blame if they are voluntary, but pardon, sometimes even pity, if they are involuntary” (Aristotle, 30). The first condition for our thinking machine must be that it is able to perform actions voluntarily. Through this condition, the thinking machine becomes an agent that is morally praiseworthy. The moral actions of a thinking machine should be indistinguishable from any moral person. The second condition is that thinking machines are able to make intelligent decisions. They use previous knowledge and experiences to make decisions and weigh the outcome of their actions. The last condition is that thinking machines are able to adapt and learn from their experiences - meaning that their ’system’ can be adapted by learning new information. The big assumption here is that there is a motivation for a thinking machine to be moral.

In addition, there is a big discussion in the field of AI regarding how each thinking machine should be ‘born’. There are two distinct routes. The first being that thinking machines should be ‘born’ as children, and then learn and adapt the same way as human children do. The second being that thinking machines should be ‘born’ as adults, and have programmed some knowledge about their purpose. Both are extremely flexible and have their own series of pros and cons. The purpose of being born as an adult is to quickly be able to utilize the skill that the thinking machine is created for. If the thinking machine is created to clean the house, then we don’t necessarily want to have to wait 21 years before it would be able to do so. However, moral rules would have to be programmed within the adults in order for them to have an understanding of right and wrong. Here, we also further our definition of thinking machines as we establish the idea that their decision-making must be stochastic.

Since we could be able to program a computer to make certain deliberations, what should the deliberations be? Should we even make deliberations? There are four possible ways that researchers have thought to be valid options. They are: Direct Specification, Domesticity, Indirect Normativity, and Augmentation. These are all options that consider different approaches to programming certain sets of information into the thinking machines, and consider what their motivation would be if that option was considered. What goals would we program a computer to fulfill? Churchland, in his book ‘Braintrust: What Neuroscience Tells Us about Morality’, suggests that morals are not objectives or transcendent, but sometimes to us they feel as though they are. Depending on the approach we take, in order for us to allow computers to follow these morals or program them, we have to make them objectives or clearly defined checks that they do.

The more optimal choice would be a combination of these options. The option of Indirect Normativity applied with the idea of Domesticity would work well. Having thinking machines be domesticated and being able to derive a standard from that would be ideal. This way we’re able to create thinking machines that are specialized for their tasks.

Applications of Weak Artificial Intelligence

A field that has been and will further be revolutionized quite incredibly by techniques in Artificial Intelligence, and one that I have explored in depth, is news and media.

Impacts of Strong Artificial Intelligence

Lingering questions

How can we extend our moral behavior into thinking machines?

References

Booklist

Ancient, Medieval and Renaissance Classics
At least seven works produced before the mid-1600s.

- Sidereus Nuncius – Galileo Galilei
- The Discourse on Method – Rene Descartes
- The Republic – Plato
- Bhagavad Gita – Sage Vyasa
- Utopia – Thomas More

Modernity – The Humanities
At least four works, produced after the mid-1600s, in Humanities disciplines such as Literature, Philosophy, History, the Arts, Critical Theory and Religion.

- Treatise of Human Nature – David Hume
- Godel, Escher, Bach – Douglas R. Hofstadter
- Superintelligence: Paths, Dangers, and Strategies – Nick Bostrom

Modernity – The Social and Natural Sciences
At least four nonfiction works, produced after the mid-1600s, in the Natural Sciences and Social Science disciplines such as Political Science, Economics, Psychology, Anthropology, and Sociology.

- An Enquiry of Human Understanding – David Hume
- On the Origins of Species – Charles Darwin
- The Descent of Man – Charles Darwin
- On Natural Selection – Charles Darwin
- Computing Machinery and Intelligence – A. M. Turing
- Hackers – Steven Levy
- The Bell Curve – Richard J. Herrnstein and Charles Murray

Area of Concentration
At least five additional works representing the student’s area or areas of concentration; students whose area of concentration already appears among the above categories may simply choose five additional works from these categories.

- The Human Brain – Susan Greenfield
- On Intelligence – Jeff Hawkins
- Artificial Intelligence: A Modern Approach – Stuart Russell and Peter Norvig
- The Big Test: The Secret History of the American Meritocracy – Nicholas Lemann
- Dataclysm: Who We Are – Christian Rudder
- The Singularity Is Near – Ray Kurzweil
- Darwin's Devices – John Long

Optional influences

These are books I have not fully read or embedded into my rationale, but I have used parts of these books to influence my arguments.

- The Structure of Scientific Revolutions – Thomas S. Kuhn
- Flatland: A romance of many dimensions – Edwin A. Abbott
- I am a strange loop – Douglas R. Hofstadter