

The Problems of “Artificial Intelligence” in Modern Philosophy and Science

Yuan Gao

Department of Literature, The National University of Mongolia, Ulan Bator, Mongolia

Abstract: Patriotism, an important component of the Chinese national spirit, has inspired generations of Chinese to strive for national prosperity. Promoting patriotism and implementing patriotic education is an eternal topic. If the youth is robust, the country will be strong. Because college students are the vital force of the country and the hope of the nation, it is especially important to cultivate their patriotism. China is facing new challenges, with profound changes in domestic and foreign situations, rapid technological development and increasingly frequent Internet exchanges. The patriotic education environment has also become more complex under the impact of undesirable Western culture. With the external and internal influences, further patriotic education for college students still faces many challenges. We should face up to the problems in contemporary patriotism education in higher education institutions, explore the solutions and cultivate patriotism among college students in the new era. Therefore, the study of patriotism education in higher education institutions has important theoretical and practical significance. This paper mainly collates literature through literature and historical research method, and the combination of theory and practice, analyzes the problems and causes of patriotism education in higher education institutions with contemporary society, college and family education as well as the characteristics of college students themselves, and puts forward targeted countermeasures for solutions. In the main body, this paper is divided into four parts. First of all, there is an introduction, which mainly includes the background and significance of research, the current status of domestic and international research, research methods, innovations and deficiencies. The framework of the paper was determined to be based on relevant domestic and international studies and the theory was well prepared for the article. The second part mainly elaborates the theories and necessity of patriotism education for college students, mainly including the connotation and characteristics of patriotism education. The third part presents a comprehensive analysis of the problem of patriotic education of college students and its causes from the social environment, patriotic education in higher education institutions, family education and college students themselves. The fourth part is the core of this paper, the practical effect of patriotic education in higher education institutions is ensured, by summarizing the relevant theories and proposing effective technologies and methods against the corresponding problems.

Keywords: Technologies and methods; Patriotic education; Higher education institutions.

1. Introduction

Many famous people have attempted to capture the nature of technology and apply it to society and human experience in technology philosophy. The findings of their study in the first half of the twentieth century primarily revealed a disconnection between technology and human existence. Technology was viewed as a self-sufficient power that annihilated fundamental aspects of humanity. Philosophers seemed to abstract from the influence of concrete technologies by restricting the concept of technology to historical and transcendental presumptions. During the 1980s, a more empiricist view of technology emerged, based on the ideas of American philosophers who considered the impact of concrete developments in their thinking (Achterhuis, 1997). The shared reliance of technology and society is a central theme in this study. This ‘empirical turn’ allowed us to account for the many faces of technology, as well as the various roles it can play in society. This viewpoint was established further by technology theorists, such as those at the University of Twente (see e.g. Verbeek, 1999).

Artificial intelligence

In 1956, artificial intelligence was developed as a research field. It is concerned with computing machines' intelligent behavior. The research objectives can be divided into four categories:

- Systems that think like humans
- Systems that think rationally

- Systems that act like humans
- Systems that act rationally

After years of high hopes for the task's success, issues emerged in the field about how to reflect intelligence that could be useful in applications. This included a lack of context expertise, the computations' intractability, and the knowledge representation structures' limitations (Russell & Norvig, 1995, pp. 20-21). However, the design group was not the only source of issues. Philosophers, who have been preoccupied with intellect and logic since Plato, began to complain as well. They attempted to demonstrate the AI project's inherent shortcomings by using both mathematical objections (based on Turing and Gödel) and more abstract concerns regarding the existence of human intelligence. Hubert Dreyfus was the most well-known of them all.

2. Relevance of Artificial Intelligence (AI) to Philosophy of Science

The consequences of artificial intelligence creation for science philosophy are the subject of this paper (AI). Is AI, on the other hand, really important to science philosophy? To begin, we will argue that it is indeed important. In reality, new AI findings have ramifications for a wide range of issues in science philosophy, as well as general philosophy. However, in this paper, we will concentrate on the significance of AI results for only one category of problems in science philosophy, although a very significant group of problems.

These are the issues that arise when induction, confirmation, and probability are combined. Since Bacon in the 17th century, these issues have been fundamental to science theory. Machine learning, which is a method by which a computer obtains hypotheses or predictions from empirical evidence, is now a fundamental technique of AI. It seems clear, then, that examining effective machine learning systems would likely shed light on the conventional philosophical of science induction problems just mentioned. Indeed, the fact that machine learning has been shown to be efficient has changed many of the previous discussions of induction. A controversy between Popper and Carnap on whether or not an inductive logic existed exemplifies this.

Popper criticized the argument that there was an inductive reasoning in his book *Logic of Scientific Discovery* (1934), concluding (p. 29):

“The numerous difficulties of inductive reasoning here sketched, in my opinion, are insurmountable.”

Carnap, on the other hand, defended the nature of inductive reasoning in his 1950 book *Logical Foundations of Probability*. He writes on page 192:

“By adding a description of c to deductive logic, inductive logic is constructed.”

Since $c(h, e) = r$ denotes the degree of confirmation of h given e , $c(h, e) = r$ denotes the degree of confirmation of h given e . Carnap considered c to be a probability function that could be interpreted logically, making his inductive logic a form of logical Bayesianism.

As a result, in 1950, Carnap, Popper, and “most theorists and scientists,” according to Carnap, believed that machine learning was impossible. Nonetheless, we now know that machine learning exists and has proven to be extremely effective. Surely, this will change people's minds about induction, particularly the issue of whether or not there can be an inductive logic.

Due to the above reasons, one of the authors of this paper (Donald Gillies) spent time researching AI from the mid-1980s to the end of the 1990s in order to see how new influenced philosophical debates regarding induction results. The findings of this investigation were published in his book *Artificial Intelligence and Scientific Method*, which was published in 1996. The book's fifth chapter continues Popper and Carnap's discussion about inductive logic. It claims that there is an inductive logic, but that it differs from the one proposed by Carnap, based on new findings in machine learning.

3. Why artificial intelligence needs philosophy?

The concept of an intelligent machine is not new, but the stored-program computer is still waiting for serious work on the issue of artificial intelligence, or even a serious understanding of what the problem is. Turing's essay 'Computing Machinery and Intelligence' (Turing 1950) and Shannon's (1950) discussion of how a computer might be programmed to play chess can be considered the beginnings of artificial intelligence. Since then, artificial intelligence has primarily progressed in the following directions. Programs have been written to solve a collection of problems that humans find difficult to solve: Playing chess or checkers, proving mathematical theorems, converting one symbolic expression into another using given rules, combining expressions made up of elementary functions, and

determining chemical compounds based on mass-spectrographic and other data are just a few examples. In the process of developing these systems, intellectual processes of varying degrees of generality are established through introspection, mathematical study, and human subject experiments. Testing the programs will lead to a deeper understanding of the cognitive processes as well as the discovery of new ones. Another option is to start with the cognitive mechanisms (for example, memory, decision-making based on weighted sums of sub-criteria, learning, tree search, and extrapolation) and create problems that exercise these mechanisms. Several attempts have been made to create a general intelligence with the same level of versatility as an individual. This has meant different things to different investigators, but none of them have had much success with it, except in terms of general knowledge used by the investigator in question. Providing sufficient conditions for general intelligence is not difficult. Turing's idea that the computer should be able to fool a knowledgeable observer into thinking it's a person for half an hour would suffice. If we focus our energies on this target, however, we will be distracted by some trivial aspects of human behavior that must be imitated. Any of these were ruled out by Turing's specification that the person to be imitated be at the end of a teletype line, obviating the need to consider accent, appearance, smell, and other factors. Turing did cause himself to be sidetracked by conversations about arithmetic fallibility, laziness, and the ability to use the English language. However, a better understanding of what intelligence is will aid work on artificial intelligence, especially general intelligence. Giving a strictly behavioral or black-box concept is one choice. In this case, we would define intelligence as the ability of a computer to solve certain types of problems that require human intelligence or to thrive in an intellectually challenging environment. This description appears a little hazy; maybe it could be made more concrete without departing from behavioral terminology, but we won't try.

4. AI & Philosophy of Science: The Situation in the 1990s to 2020

Many very technical advances in logic and philosophy of science were indicated by new AI findings in the 1996 book *Artificial Intelligence and Scientific Method*. However, we will not address these more technical findings in this section, instead focusing on the general conclusions drawn about science philosophy. The effectiveness of machine learning programs was a major factor in reaching these conclusions. There were four of them.

(1) Induction is real

Popper wrote (1963, p. 53):

“Induction, i.e. inferences based on many observations, is a myth. It is neither a psychological fact, nor a fact of ordinary life, nor one of scientific procedure.”

Popper was wrong on this issue, as shown by the success of machine learning systems. However, it should be noted that Popper was not entirely right in 1963 when he made the remark. Machine learning did not achieve its first notable successes until the late 1970s.

(2) Testing and falsification are used in machine learning

Popper's account of scientific methods emphasizes testing and falsification, and machine learning here complements rather than contradicts Popper's account. Machine learning algorithms learn from some training data, then validate and

likely reject their assumptions using more test data. As a result, we have a series of conjectures and denials. Machine learning differs from Popper in that conjectures are created mechanically by a program rather than intuitively by humans.

(3) The importance of background knowledge

Gillies (1996) examined machine learning programs that used both context information (K) and empirical proof (e). "Computer inferences really take the form from K&e infer h rather than from e infer h," the authors conclude (p. 70). For example, each of the residues in GOLEM's application to the protein folding problem was defined in terms of properties that scientists in the field recognized as being important to how a protein folded. These included characteristics like "hydrophobicity," "polarity," "aromaticity," and so on. Furthermore, GOLEM was programmed with 9-place predicates that directed the program to look for laws deciding the character of a residue in terms of the 4 residues on either side of it.

(4) Human interaction with the results of machine learning

Gillies (1996) examined machine learning programs that all generated rules that were understandable to humans. Muggleton's GOLEM, for example, stated an explicit rule relating to protein folding in Gillies (1996), p. 53. This was common during the period when machine learning was mostly used to learn the rules of rule-based systems. As a consequence of this scenario, the following study of human interaction with machine learning results [Gillies (1996) pp. 54-55]:

"Generating rules that are humanly understandable has a lot of benefits because it allows for the following type of human-machine interaction. The human scientist's background information is encoded in a machine learning software. This produces previously unknown but humanly understandable laws that refer to the domain in question. The human scientist will then investigate these rules and possibly gain new knowledge in the field."

We will see if the original findings of Artificial Intelligence and Scientific Method are backed by new developments after reviewing new developments in AI since 1996. In several cases, the approaches and issues addressed are close to those of 1996. However, there have been a number of changes:

- The increasing importance of machine learning relative to other AI techniques
- The increasing use of neural networks
- The use of deeper, more complex, models and larger datasets
- The use of deep neural networks for representation learning.

We can now see how these affect the four conclusions:

(1) Induction is real

This seems to be the case. Machine learning, which is a form of artificial induction, is more critical than ever. In reality, this argument may be bolstered. The main scientific insight when using machine learning is selecting an acceptable mathematical form for the machine learning model, and the automated method is simply setting those parameters, according to one potential counter-argument to machine learning being a form of scientific induction. The determination of the orbits of the planets by Kepler is an analogy. The mathematical structure of the orbit (an ellipse with one focus at the sun) was his main insight, and calculating the model's parameters (the dimensions of the ellipses) could be automated. Machine learning models can be thought of as increasingly complex mathematical forms.

However, as machine learning models become more complex and general-purpose, such as deep neural networks, the mathematical type chosen appears to become less and less scientifically important.

One element of the back-propagation algorithm, however, could be seen as supporting Popper's position. Back-propagation starts with random weights and corrects them with errors in the output. Rather than induction, this use of errors to develop the model may be considered automatic falsification. This does not rule out the original hypothesis, since many machine learning algorithms are explicitly inductivity, but it does imply that machine learning may be used to automate both inference and falsification.

(2) Testing and falsification are used in machine learning

Testing is also an important aspect of machine learning. Machine learning practitioners will split their data into two sets: a training set for the machine learning algorithm and a test set for testing the model after it has been trained. If the model fails to perform correctly on the testing results, it is considered a failure, and the training process must be changed. This procedure is used to ensure that the model can generalize to new data and hasn't 'over fitted,' meaning that it hasn't learned to recognize aspects of the training data that aren't relevant. To improve this phase, a number of specialized training methodologies have been developed. Cross validation, for example, separates the data into training and test data in a variety of ways and conducts several training steps on each split. The results are used to compare various learning algorithm parameters. Multiple testing sets are standard in professional systems. During software development, a development set is used to evaluate the model. A separate test set is used as a final validation after the model tends to perform well on the development set (developers often gather new data at this stage). These findings reinforce the original conclusion that testing is an essential aspect of machine learning methodology.

(3) The importance of background knowledge

This is a lot more difficult. Those who argue that deep learning demands less background information than previous machine learning methods are definitely right. Context experience can be used in a variety of ways. The first is when a machine learning model's mathematical form is chosen. As previously stated, however, this is becoming less scientifically important. Second, in the GOLEM method mentioned above, the option of which features of the data to use as input to the machine learning model was the primary use of context information. One of the most critical aspects of deep learning is that raw data features such as pixels are used to learn intermediate representations of the data. This eliminates the need for a scientist to choose meaningful features, as well as the need for prior expertise.

Other applications of context knowledge, on the other hand, continue to be significant. The creation of the data sets themselves is one example. This entails the selection of data examples that would be useful to machine learning. The data must also be labeled, as each instance must include the appropriate output. The sequence of amino acids, for example, would be the input data for protein folding, but the data set would also need to have an output label: the correct 3D folded structure. This mark explicitly necessitates extensive prior knowledge.

(4) Human interaction with the results of machine learning

The benefits of making human-comprehensible laws or models are a part of the research process that should not be

altered by developments in machine learning methods. However, as we've seen, neural networks are becoming increasingly relevant, despite their reputation for being difficult to understand by humans. The emphasis on interpretability by Muggleton et al. (1992) can be seen as a justification for their approach over neural networks. The fact that neural networks are made up of a large number of very basic units makes them difficult to comprehend. Although a computer can easily process this form of model, humans find it difficult to keep track of all of the interconnected components. Within the machine learning community, the complexity of understanding neural networks, as well as machine learning in general, is increasingly seen as a challenge. O'Neil (2016), for example, has highlighted the social concerns that arise when machine learning algorithms make decisions about health, crime, and other social issues without a human being to double-check their accuracy. Having scientific findings that are incomprehensible to humans seems to be a similar problem. As a result, researchers are looking into machine learning approaches that can justify their findings.

One method, which is especially popular in machine learning for medicine, is to create learning models that are designed to be understandable by humans. Wang and Rudin (2015), for example, worked in medical diagnosis and wanted to create models that could be used in practice. As a result, they used a model that was identical to current doctor check lists. These were dubbed "falling rule lists," which are a set of if-then rules that are checked in sequence before a decision is reached. While these lists are currently created by humans, the authors suggest an algorithm for learning them. However, in comparison to deep neural networks, these models must be simplified, making them less efficient in terms of the models they can learn. As a result, their overall accuracy can suffer. Producing local explanations is one way to combine complex models with human understanding. Although the model is still too complicated for humans to comprehend, each decision it makes can be clarified. Ribeiro et al. (2016), for example, use a local intelligible model to describe individual decisions. This work is similar to the previous work on intelligible models, but it does not try to describe a machine learning system's entire behavior using a simplified model. Instead, it would construct a simpler model (for example, a linear model) to illustrate how a machine learning model behaves on a single data point. The model can quite effectively describe behavior in that data point's immediate neighborhood, but it does not attempt to explain behavior beyond that.

As a consequence, even though recent advances have made learning outcomes less interpretable, we may conclude that machine learning interpretability remains a critical problem in science. As a result, machine learning in science continues to face significant challenges.

5. Development of AI from the early 1970s to the late 1990s

The first digital computers appeared in the late 1940s, and by 1950, many people were considering how to use these new computers to construct artificial intelligence. In his famous 1950 article, *Computing Machinery and Intelligence*, Turing sketches out ideas along these lines. In the 1950s and 1960s, several researchers worked on such plans and developed ideas that would later prove useful. However, by the early 1970s, there had been no notable successes in the emerging field of

artificial intelligence. As a result, the whole AI project was attacked. Hubert Dreyfus' book *What Computers Can't Do: A Critique of Artificial Reason*, published in 1972, is perhaps the most well-known. Dreyfus has expressed serious concerns about computers' ability to play chess with skill, arguing in his book that "further substantial advancement... in Artificial Intelligence is highly unlikely" (p. 197). Dreyfus himself was defeated by a machine at chess, and he appeared irritated by the joy with which the AI community greeted the news of his defeat. Following events revealed that he was completely incorrect about machine chess. In a six-game match played under standard conditions, Garry Kasparov, the world chess champion, was defeated by a computer (Deep Blue) in May 1997. Still, to be fair to Dreyfus, it should be recalled that when he published his book in 1972, the AI project had been in the works for more than two decades with nothing to show for it. Expert systems, on the other hand, provided a breakthrough in AI in the early 1970s. The Stanford heuristic programming community, especially Buchanan, Feigenbaum, and Shortleaf, took the lead in this area. They discovered that the secret to success was to extract from an expert the information that he or she used to perform a specialized task, and then code that knowledge into the machine. They were able to create "expert systems" that performed complex tasks at the level of human experts in this way. DENDRAL was the first method of its kind, inferring a possible molecular structure from a compound's atomic composition and mass spectrogram. Another early expert method (MYCIN) was based on blood infection diagnosis. It was created in the 1970s by Edward Shortliffe and his colleagues in collaboration with the Stanford medical school's infectious diseases department. The area's medical expertise was codified into rules of the form: IF such and such is found, THEN such and such is the most probable inference. Over 400 such rules were collected from medical experts and added to MYCIN's knowledge base. Expert systems, on the other hand, quickly became a challenge. Since these systems were based on law, they were often referred to as "rule-based systems." Initially, the rules were gathered by interviewing experts in the area, but this method had two flaws. To begin with, the experts were capable of performing a mission, but they did not always understand the rules they used to do so. Perhaps these laws were learned subconsciously rather than consciously. Second, interviewing experts was a time-consuming and costly process. Feigenbaum first mentioned this problem in 1977, and it became known as Feigenbaum's bottleneck.

In 1977, Feigenbaum proposed his bottleneck problem, and in the following years, the first attempts to develop machine learning techniques to solve it were made. Meta-DENDRAL, one of the first programs, was created to generate rules for DENDRAL to use. It was effective in accomplishing this goal, with some of the rules even being published in a chemistry journal. [See Buchanan and Feigenbaum (1978) for more information.] Michalski also developed a machine learning method for diagnosing soybean disease at the same time. Several hundred correctly diagnosed examples presented by domain experts were used to infer the rules. [See Michalski and Chilautsky (1980) for more information.]

Gillies (1996, pp. 31-55) examined two current state-of-the-art machine learning systems in depth. They were both based on logic. The first, ID3, was concerned with decision tree induction. Quinlan was the one who came up with this (see his 1986). The second was a device called GOLEM,

which Muggleton and his colleagues developed. This was used to solve the problem of figuring out how proteins fold up. A protein is made up of a linear sequence of amino acid residues that fold up in various ways depending on the existence of the residues. GOLEM did, in fact, generate some protein folding rules that were correct. [See Muggleton, King, and Sternberg (1992) for more information.] It's worth noting that GOLEM used a data set of around 500 proteins in this issue, all of which had their folding patterns determined using X-ray crystallography or NMR (nuclear magnetic resonance) techniques.

Machine learning techniques in AI were first used to induce the rules used in expert systems in the 1980s and early 1990s, and they appeared to be applied to data sets containing a few hundred instances. However, there have already been significant achievements in the form of computer chess and the discovery of competitive but previously unknown rules in scientifically significant areas. In the next part, we'll look at Donald Gillies' findings for philosophy of science from his 1996 novel. Before we go any further, it's worth noting that the advancements in AI that we just discussed made extensive use of findings from philosophy of science. Shortliffe and Buchanan's 1975 paper contains 33 references, 14 of which (or over 42%) are to works in philosophy of science dealing with the validation of scientific theories by proof, as well as related issues concerning induction and probability interpretation. Carnap, de Finetti, Hempel, Popper, Ramsey, and Salmon are among the philosophers of science listed. Other AI researchers from the time period cite Sir Francis Bacon's works from the 17th century. This demonstrates the prospect of a fruitful collaboration between AI and science philosophy. The early stages of AI development benefited from ideas from philosophy of science. The outcomes of this research could lead to improvements in science theory, which could be beneficial to future developments in AI and other fields.

6. The Case-Based Reasoning paradigm

The sense of CBR is first sketched in terms of perceptions, motives, and intelligence viewpoints. Then, inside the CBR framework, modeling problems are addressed. The CREEK system, built at the University of Trondheim, is the subject of the third section of the chapter. As a result, case-based reasoning serves as a foundation for the philosophical study that follows. Some philosophical questions are raised as entries into such a study in this chapter.

Assumptions

One of the assumptions of symbolic AI, according to Dreyfus, is that reasoning is interpreted as rule-based symbol manipulation (the psychological assumption). This assertion is not made in the case-based reasoning model. Reasoning is based on a memory of stored cases rather than rules. Reasoning in CBR is dependent on recall. The most important cases are retrieved and adapted to the new situation when solving a problem. Two new assumptions are introduced by CBR (Leake, 1996):

- The universe works in predictable ways: similar problems have predictable solutions.
- The types of issues that an agent encounters are likely to recur.

The first question that arises is how these conclusions relate to Dreyfus' point of view. Symbolic AI, according to Dreyfus,

requires formalizable knowledge and a formalizable truth (epistemological and ontological assumptions, respectively), while case-based reasoning assumes regularity rather than formalizability. CBR requires a normal world as well as information that can be expressed in a regular manner. As a result, cases are guaranteed to be relevant for potential problem solving. The first core characteristic of CBR, in my opinion, is the assumption of regularity.

Case-based reasoning separates itself from the rationalistic tradition of philosophy by assuming regularity rather than formalizability. Instead, there appears to be a connection between CBR and the empiricist tradition. For example, Hume (1886) argued that what we think of as natural laws are really notions based on lifelong observation of cases rather than unbreakable rules.

7. The phenomenological approach

Phenomenology is a scientific movement that focuses on human-reality relationships. In the analysis of intelligent machines, there are two ways to apply the phenomenological approach. For starters, we will discuss how these devices alter human-world relationships. Second, in the same way that we characterize humans as being related to their reality, we may describe computers as being related to their reality. This necessitates seeing the computer as a part of the universe.

Post phenomenology

There has been a renewed interest in the phenomenological movement in philosophy in recent years. Don Ihde and Peter-Paul Verbeek, for example, use phenomenological philosophers' theories to explain the impact of technology on our lives. 'Post phenomenology can be thought of as a phenomenological approach that shares the postmodern aversion to context-independent truths and the radical separation of subject and topic, but does not want to transform this into indifferent relativism.' (Chapter 4 of Verbeek, 2000)

Don Ihde, a philosopher who attempted to apply old phenomenological theories to technology, coined the term "post phenomenology." In building a perspective on the influence of innovations in human life, Peter-Paul Verbeek (Verbeek, 2000) uses the term in a wider context. There are two major requirements for such a viewpoint. To begin, it must avoid, as many early philosophers of technology did, reducing technology to its transcendental presumptions (as explained in chapter 1). Second, it must allow for the assessment of technological influence, which necessitates a departure from constructivism, as in Actor-Network Theory. Since there is no space in the latter methods for assessing the role of artifacts in hermeneutic or existential processes, this would be a rejection of the early philosophers' evaluative perspectives.

Phenomenology

Edmund Husserl (1859-1938), with his popular credo "zu den Sachen selbst," established phenomenology as a philosophical movement. He ushered in the twentieth century with a slogan intended to 'bracket life' and focus attention on the directedness of consciousness toward the universe. Intentionality was the name given to this directedness. The emphasis on the relationship between consciousness and the universe provided an alternative to the classic extremes of idealism and realism, in which the first prioritized consciousness and the second prioritized the world in knowing questions. In other words, phenomenology did not claim that the universe or consciousness were the sources of knowledge, but rather that they were connected and

constituted the foundation of knowledge through intentionality.

8. Reasoning programs and the Missouri program

The philosophical issues that must be resolved will become clearer in relation to a specific type of proposed intelligent program, known as a reasoning program or RP for short. A reasoning program communicates with the environment by input and output devices, some of which may be general sensory and motor organs (such as television cameras, speakers, and artificial arms) and other communication aids (for example, teletypes or keyboard-display consoles). Internally, RP can represent data in a number of different ways. Dot arrays or lists of regions and edges with classifications and adjacency relations, for example, may be used to display images. Scenes can be interpreted as a list of bodies in various positions, shapes, and speeds. Situations could be defined by abstract expressions with transition laws. Utterances can be expressed by digitized time functions, phoneme sequences, and sentence parsing. However, in simpler systems, one representation can play a dominant role and even be the only representation present. This is a representation made up of a series of sentences written in a formal logical language. W-order logic with function symbols, description operators, conditional expressions, sets, and so on, for example. It's debatable if we need to have modal operators with referential opacity. In the following way, this representation is dominant:

1. All other data structures have linguistic explanations that explain how the structures relate to one another and what they reveal about the environment.

2. Each subroutine has a verbal explanation that explains what it does, whether it's manipulating data internally or manipulating the environment externally.

3. Language is used to express the laws that express RP's views about how the environment acts and the implications of strategies.

4. The experimenter's aims, its devised sub goals, and its assessment of its success are all linguistically articulated.

5. We will assume that RP's knowledge is sufficient to solve a problem if it follows logically from any of these sentences that a specific action plan will solve it.

6. RP is a deductive program that seeks out methods of action that it can show can solve a problem and then implements them.

7. Strategies can have sub goals that RP must solve, and part or all of a strategy may be strictly intellectual, i.e., the quest for a strategy, evidence, or any other intellectual item that meets certain requirements.

McCarthy (1959) was the first to mention such a program, which he dubbed the Advice Taker. McCarthy (1963) proposed a preliminary approach to the necessary formalism, which has since been superseded by this paper. This paper is in part a response to Y. Bar-observation Hillel's that the original paper involved certain philosophical presuppositions when it was introduced at the 1958 Symposium on the Mechanization of Thought Processes. Constructing RP entails both the epistemological and heuristic aspects of the artificial intelligence problem: that is, the knowledge in memory must be sufficient to establish a strategy for achieving the target (which may require the acquisition of additional information), and RP must be smart enough to discover the strategy and prove its correctness. Of course, these issues are intertwined,

but since the emphasis of this paper is on the epistemological aspect, we'll mention the Missouri program (MP), which focuses solely on that aspect.

The Missouri program (whose slogan is "Show me") does not seek out methods or evidence that the strategies are effective in achieving an objective. Rather, it helps the experimenter to show proof measures and verify that they are accurate. Furthermore, when it is 'convinced' that it can take action or carry out a plan, it does so. This paper can be thought of as an attempt to construct a Missouri program that can be convinced to achieve its objectives.

9. Representations of the world

The first step in creating an RP or MP is deciding what structure the world should have and how knowledge about it and its laws of change should be interpreted in the machine. It turns out that when one is discussing the expression of general laws or concrete facts influences this decision. As a result, our understanding of gas dynamics is reliant on the representation of a gas as a large number of moving particles in space; this representation is critical in deriving the mechanical, thermal, electrical, and optical properties of gases. The location, velocity, and excitation states of each particle are thought to determine the state of the gas at any given time. We never decide the position, velocity, or excitation of a single molecule, however. The pressure, temperature, and velocity fields, as well as average pressures and temperatures, are used to express our realistic knowledge of a specific sample of gas. From a metaphysical standpoint, this is perfectly natural, and we are not prone to deny the presence of things we cannot see, or to be so anthropocentric as to believe the universe must be built in such a way that we have direct or even indirect access to all.

From the perspective of artificial intelligence, there are three types of adequacy for world representations. If the universe could have the shape without contradicting the facts of the part of truth that concerns us, it is said to be metaphysically adequate. The following are some examples of metaphysically adequate representations for various facets of reality:

1. The representation of the world as a set of particles that interact through forces between pairs of particles.

2. The universe is represented as a massive quantum-mechanical wave function.

3. Representation in the form of a system of interconnected discrete automata. We're going to make use of this illustration.

The primary use of metaphysically adequate representations is in the construction of general theories. A further move is to derive measurable consequences from the theory.

If a representation can be used literally to articulate the facts that one has about a particular aspect of the universe, it is said to be epistemologically adequate for an individual or machine. As a result, none of the above interpretations are sufficient to convey evidence such as "John is at home," "dogs chase cats," or "John's phone number is 321-7580." Ordinary language is clearly sufficient for expressing the information that people talk to one another in ordinary language. For example, expressing what people know about how to identify a specific face is insufficient. The second section of this paper is concerned with a formal representation of common-sense facts of causality, capacity, and information that is epistemologically adequate.

If the reasoning methods actually used to solve a problem

are expressible in the language, the representation is said to be heuristically adequate. In this paper, we will not go into greater detail about this tentatively formulated definition, except to point out later that one particular representation appears epistemologically adequate but not heuristically adequate.

10. Counterfactuals

Of course, there is a substantial literature on this ancient philosophical problem, almost none of which appears to be directly applicable to us. However, Rescher's (1964) theory, which was established recently, may be useful. We will not attempt to describe Rescher's theory here because his book is so well written. The reader should be aware of Sosa's (1967) critical review, which recommends some minor changes.

The significance of this hypothesis for us is that it implies a different solution to the problem we've dubbed the frame problem. The following is a summary of the situation. As a matter of practice (or possibly inference), one believes that when decisions are taken, all propositional fluent that applied in the previous situation still apply in the current situation. This will often result in an incoherent collection of claims about the new situation; Rescher's theory offers a method for restoring continuity in a logical manner, as well as providing as a by-product those fluent who's meaning changes as a result of performing the operation. However, we haven't looked into it further.

11. Discussion of Literature

The strategy outlined in this paper for achieving a generally intelligent program would undoubtedly be difficult to implement. As a result, it's only logical to wonder whether a simpler scheme will work, and we'll spend this section critiquing some of the simpler schemes that have been proposed.

L. Fogel (1966) proposes that intelligent automata evolve by changing their state transition diagrams to improve their performance on increasingly complex tasks. Fogel's experiments require the evolution of machines with less than ten states to predict the next symbol in a relatively simple series. We doubt that this method would yield interesting results because it appears to be restricted to automata with a small number of states, say less than 100, while computer programs that are considered automata have 21Q5 to 21Q States. This is due to the fact that, while the representation of behaviors by finite automata is metaphysically adequate - in theory, any behavior that a person or computer may perform can be represented - it is not epistemologically adequate; that is, we can't easily articulate constraints we want to put on a behavior or what we've learned from an experience as changes in an automaton's state diagram.

Several researchers (Galanter 1956, Pivar and Finkelstein 1964) have proposed that intelligence can be described as the ability to predict the future of a sequence based on its past observations. The idea is presumably that a person's past can be viewed as a series of discrete events and that intelligent people can forecast the future. Writing programs to predict sequences developed according to some basic class of laws is then used to study artificial intelligence (sometimes probabilistic laws). Again, the model is satisfactory in terms of metaphysics but not in terms of epistemology.

Friedberg (1958, 1959) experimented with using a computer program to describe behavior and evolving a

program to perform a task via random mutations. Desired changes in behavior are often not represented by minor changes in the machine language type of the program, indicating the representation's epistemological inadequacy. Learning a new truth, in particular, has an unrepresentable impact on a reasoning program.

For several years, Newell and Simon experimented with a program called the General Problem Solver (Newell et al 1959, Newell and Simon 1961). The task of converting one symbolic expression into another using a collection of transformation rules is represented by this program. They were able to describe a wide range of problems in this format, but the representation for a lot of them was awkward enough that GPS could only do small examples. The task of enhancing GPS was investigated as a GPS task, but it seems to have been abandoned. The name General Problem Solver implies that the authors once assumed that most problems could be expressed in its terms, but their more recent publications have shown that they hold different views.

Their division of the problem solver into an input program that transforms problems into internal representation and the problem solver itself roughly corresponds to our division of the artificial intelligence problem into epistemological and heuristic components. The distinction is that we are more concerned with the internal representation's suitability. Newell (1965) raises the question of how to obtain "heuristically sufficient" representations of problems, and Simon (1966) addresses the idea of "can" in a way that can be contrasted to the current approach.

12. Logics and theories of actions

Von Wright's action logic, as outlined in his book *Norm and Action*, is the most fully developed theory in this field (1963). Von Wright's logic is based on an unconventional tense-logic of his own invention. Since the basis is a binary modal connective, $p \text{ Tq}$, where p and q are propositions, denotes 'p, then q.' As an example, the action of opening a window is: (the window is closed)T (the window is open). The above-mentioned book went a long way toward formalizing the calculus, but there were still some interpretation issues, as Castaneda points out in his study (1965). Von Wright (1967) has modified and expanded his formalism in response to these and other critiques, as well as providing a sort of semantic theory based on the concept of a life-tree in a more recent article.

We are unaware of any other attempts to establish a single theory of behavior that have progressed to this stage, but there are many discussions of difficulties and surveys that appear to be important. Rescher (1967) covers a lot of ground, and Davidson (1967) makes some good points as well. The key theory of Davidson is that, in order to translate statements concerning acts into the predicate calculus, it appears that actions must be treated as values of bound variables, or as actual individuals (according to Quine's test). Of necessity, the situation calculus fits this advice in that it allows quantification of tactics with actions as a special case. Simon's papers on command-logics (1965, 1967) are also noteworthy. Simon's main point is to demonstrate that a special logic of commands is unnecessary, with ordinary logic serving as the sole deductive mechanism; however, we won't dwell on that for now. He makes many observations, the most important of which is that agents do not perform acts most of the time, and that they only do so when compelled to by external forces. He uses the example of a serial processor operating in a parallel-

demand environment, and the need for interrupts those results. Action logics like von Wright's and ours don't differentiate between action and inaction, and we're not aware of any action logic that has progressed far enough to satisfy Simon's implied critique.

There is a vast body of purely philosophical writings on behavior, time, determinism, and other topics, the majority of which is meaningless for the purposes of this discussion. However, we'll list two that have recently surfaced and appear to be intriguing: Chisholm (1967) and Evans (1967) wrote papers summarizing recent debates on the distinctions between states, performances, and activities.

13. Conclusion

With a few caveats, this discussion has shown that our four hypotheses hold up in light of recent AI advances. Machine learning demonstrates the feasibility of automatic induction; though testing and falsification are also important aspects of machine learning. In science, the use of machine learning does not negate the need for human experts. It necessitates human context information, and the effects of machine learning must also be interpretable by humans. Deep Learning, on the other hand, has made understanding more difficult by reducing the need for prior information.

These findings call into question Popper's claim that induction is a myth. It does not, however, alter his belief that induction has no place in human science. Instead, it appears that AI works in a somewhat different way than humans when it comes to science. This is backed up by the complexity of understanding machine learning results: mathematical models created by machines vary significantly from those created by humans. This raises critical questions about science's future.

This paper has made many comparisons between AI's use in science and its more well-known applications in gaming, especially Deep Blue's well-publicized defeat of Kasparov. What Kasparov did after that is less well-known. In response to his defeat, he created Advanced Chess (also known as Centaur Chess), a new type of chess game in which the players were humans collaborating with computers (Kasparov and Greengard 2017). The world's best humans and machines operating alone were defeated by these joint teams. Humans and machines, interestingly, continued to play different positions. Computers are better at the game's detailed strategies (which champions like Kasparov excelled at), whereas humans are better at the game's large-scale strategy. This demonstrates that computers would not be able to fully replace human scientists. Rather, computers and humans perform science in different ways, and humans working with computers can combine their various abilities to perform science in a more advanced manner (which is certainly the case today). Understanding these various skills and developing machine learning systems that humans can easily communicate with are likely to open up a world of possibilities for science. If these efforts succeed, the next few decades should see massive new scientific breakthroughs that would not have been possible without AI's new technologies.

References

- [1] Donald Gillies (12 March, 2020), *Artificial Intelligence and Philosophy of Science from the 1990s to 2020*.
- [2] Bengio, Y., Courville, A. and Vincent, P. (2013) Representation learning: A review and new perspectives, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8), pp. 1798–1828.
- [3] Bishop, C. M. (ed.) (1995) *Neural networks for pattern recognition*, Oxford University Press.
- [4] Buchanan, B.G., and Feigenbaum, E.A. (1978) DENDRAL and Meta-DENDRAL: Their Applications Dimension, *Artificial Intelligence*, 11, pp. 5-24.
- [5] Davis, R., Buchanan, B.G., and Shortliffe, E.H. (1977) Production Systems as a Representation for a Knowledge-Based Consultation Program', *Artificial Intelligence*, 8, pp. 15-45.
- [6] Gillies, D. (1996) *Artificial Intelligence and Scientific Method*, Oxford University Press.
- [7] Kasparov, G. K. and Greengard, M. (2017) *Deep thinking: where machine intelligence ends and human creativity begins*, John Murray.
- [8] Anderson, A.R. (1956) *The formal analysis of normative systems*. Reprinted in *The Logic of decision and action* (ed. Rescher, N.). Pittsburgh: University of Pittsburgh Press.
- [9] Aqvist, L. (1965) *A new approach to the logical theory of interrogatives, part I*. Uppsala: Uppsala Philosophical Association.
- [10] Church, A. (1956) *Introduction to Mathematical Logic*. Princeton: Princeton University Press.
- [11] Fogel, L. J., Owens, A. J. & Walsh, M. J. (1966) *Artificial Intelligence through simulated evolution*. New York: John Wiley
- [12] Green, C. (1969) Theorem-proving by resolution as a basis for question-answering systems. *Machine Intelligence 4*, pp. 183-205 (eds Meltzer, B. & Michie, D.). Edinburgh: Edinburgh University Press.
- [13] Hintikka, J. (1967a) A program and a set of concepts for philosophical logic. *The Monist*, 51, 69-72.
- [14] Kripke, S. (1963a) Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16, 83-94.
- [15] Manna, Z. (1968b) Formalization of properties of programs. *Stanford Artificial Intelligence Report: Project Memo AI-64*.
- [16] J. McCarthy, P. J. Hayes; *Some Philosophical Problems from the Standpoint of Artificial Intelligence*.
- [17] Aamodt, Agnar, and Plaza, Enric, 'Case-based reasoning: Foundational issues, methodological variations, and system approaches'. In: *AI Communications 7* (1994), pp. 39-59.
- [18] Aamodt, Agnar, and Nygård, Mads, 'Different roles and mutual dependencies of data, information, and knowledge – an AI perspective on integration'. In: *Data and Knowledge Engineering 16* (1995), pp. 191-122.
- [19] Brey, Philip, 'Hubert Dreyfus: Humans versus computers'. To appear in: Achterhuis, H. (ed.), *American philosophy of technology: The empirical turn*, Indiana University Press, 2001.
- [20] Russell, Stuart, and Norvig, Peter, *Artificial Intelligence: A modern approach*, New Jersey: Prentice Hall, 1995.