

AI as Empirical Enquiry: Strategies and Methods for Problem Solving

1. Intelligence and the Physical Symbol System Hypothesis.

From an historical perspective, the dominant approach to artificial intelligence involved the construction of representational formalisms and their associated search-based reasoning mechanisms. The guiding principle of early AI methodology was the physical symbol system hypothesis, first articulated by Newell and Simon. This hypothesis states:

The necessary and sufficient condition for a physical system to exhibit general intelligent action is that it be a physical symbol system.

Sufficient means that intelligence can be achieved by any appropriately organized physical symbol system.

Necessary means that any agent that exhibits general intelligence must be an instance of a physical symbol system. The necessity of the physical symbol system hypothesis requires that any intelligent agent, whether human, space alien, or computer, achieve intelligence through the physical implementation of operations on symbol structures.

General intelligent action means the same scope of action seen in human action. Within physical limits, the system.

The physical symbol system hypothesis has led to four significant methodological commitments: (a) the use of symbols and systems of symbols as a medium to describe the world; (b) the design of search mechanisms, especially heuristic search, to explore the space of potential inferences those symbol systems could support; and (c) the disembodiment of cognitive architecture, by which we mean it was assumed that an appropriately designed symbol system could provide a full causal account of intelligence, regardless of its medium of implementation. Finally (d), on this viewpoint, AI became empirical and constructivist: it attempted to understand intelligence by building working models of it.

There are many criticisms that can be leveled at the physical symbol system characterization of intelligence. The grounding of meaning is an issue that has forever frustrated both the proponents and critics of the AI and cognitive science enterprises. The grounding issue asks how symbols can have meaning. Searle makes just this point in his discussion of the so-called Chinese Room. Searle places himself in a room intended for translating Chinese sentences into English; there he receives a set of Chinese symbols, looks the symbols up in a large Chinese symbol cataloging system, and then outputs the appropriately linked sets of English symbols. Searle claims that although he himself knows absolutely no Chinese, his “system” can be seen as a Chinese-to-English translation machine.

There is a problem here. Although anyone who has worked in the research areas of machine translation or natural language understanding, might argue that the Searle “translation machine”, blindly linking one set of symbols to another set of symbols, would produce results of minimal quality, the fact remains that many current intelligent systems have a very limited ability to interpret sets of symbols in a “meaningful” fashion. This problem of too weak a supporting semantics also pervades many computationally based sensory modalities, whether they be visual, kinesthetic, or verbal.

A significant alternative to the physical symbol system hypothesis is the research into neural networks and other biologically inspired models of computing.

2. Probabilistic Models and the Stochastic Technology.

We never know a guaranteed true result. The results of the AI system are interpreted as a probabilistic estimate. The system always gives a probabilistic result. For example, when classifying a number in an image,

we get the probability that this number is similar to each of the digits. The level of confidence of different systems in the processing of one dataset also varies and depends on the structure of the system and the algorithms by which it works. Probabilistic methods can be used within the system to find the result faster. For example, to reduce the number of steps of the algorithm compared to a brute force of all options. The probabilistic approach is used in genetic algorithms to perform mutations in a gene and to select a generation from a large number of the best genes. In neural networks, the value of the activation function can also be interpreted as a probabilistic variable, because it depends on the input parameters, which are often random.

Artificial intelligence uses an inductive approach. The learning output of the models depend on the input training dataset for model learning. However, the system can learn without initial data, and perform actions itself and evaluate the benefits of them. For example, game bots learn while playing, changing their strategy depending on whether their actions help achieve the goal. This approach is abductive.

3. The Science of Intelligent Systems: Epistemological Issues:

1) Inductive bias, the rationalist's a priori.

The automated learning and most AI techniques reflected the a priori biases of their creators. The problem of inductive bias is that the resulting representations and search strategies offer a medium for encoding an already interpreted world. They rarely offer mechanisms for questioning our interpretations, generating new viewpoints, or for backtracking and changing perspectives when they are unproductive. This implicit bias leads to the rationalist epistemological trap of seeing in the world exactly and only what we expect or are trained to see.

The role of inductive bias must be made explicit in each learning paradigm. Furthermore, just because no inductive bias is acknowledged, doesn't mean it does not exist and critically effect the parameters of learning. The commitments made within a learning scheme, whether symbolbased, connectionist, emergent, or stochastic to a very large extent mediate the results we can expect from the problem solving effort. When we appreciate this synergistic effect throughout the process of the design of computational problem solvers we can often improve our chances of success as well as interpret our results more insightfully.

2) The empiricist's dilemma.

If current approaches to machine learning, especially supervised learning, possess a dominant inductive bias, unsupervised learning, including many of the genetic and evolutionary approaches, has to grapple with the opposite problem, sometimes called the empiricist's dilemma. Themes of these research areas include: solutions will emerge, alternatives are evolved, populations reflect the survival of the fittest. This is powerful stuff, especially situated in the context of parallel and distributed search power. But how can we know that system got correct result? The empiricist does require the remnants of a rationalist's a priori to save the science. Nonetheless, there remains great excitement about unsupervised and evolutionary models of learning; for example, game bots and genetic algorithms. constructivist epistemology, coupled with the experimental methods of modern artificial intelligence, offer the tools and techniques for continuing the exploration of a science of intelligent systems.

4. The necessity of designing falsifiable computational models and the limitations of the scientific method.

Popper and others have argued that scientific theories must be falsifiable. Falsification in its simplest form means refuting a theory or theoretical assumption by referring to an empirical fact that contradicts that theory. According to the epistemological position of falsibilism, any theory can be wrong, because in science there is no irrefutable knowledge. If there are no facts that could refute the theory, then it is unscientific. A counterexample of the opposite is the Russell's teapot. He wrote that if he were to assert, without offering proof, that a teapot, too small to be seen by telescopes, orbits the Sun somewhere in space between the Earth and Mars, he could not expect anyone to believe him solely because his assertion could not be proven

wrong. This means that there must exist circumstances under which the model is not a successful approximation of the phenomenon. The obvious reason for this is that any number of confirming experimental instances are not sufficient for confirmation of a model. Furthermore, much new research is done in direct response to the failure of existing theories.

The general nature of the physical symbol system hypothesis as well as situated and emergent models of intelligence may make them impossible to falsify and therefore of limited use as models. The same criticism can be made of the conjectures of the phenomenological tradition. Some AI data structures, such as the semantic network, are so general that they can model almost anything describable, or as with the universal Turing machine, any computable function.

5. The Scientific Method a Foundation for a Modern Epistemology

The scientific method is an empirical method of acquiring knowledge that has characterized the development of science. It involves careful observation, applying rigorous skepticism about what is observed, given that cognitive assumptions can distort how one interprets the observation. It involves formulating hypotheses, via induction, based on such observations; experimental and measurement-based testing of deductions drawn from the hypotheses; and refinement (or elimination) of the hypotheses based on the experimental findings. These are principles of the scientific method, as distinguished from a definitive series of steps applicable to all scientific enterprises.

The scientific method is the only tool we have for explaining in what sense issues may still be outside our current understanding. Every viewpoint, even from the phenomenological tradition, if it is to have any meaning, must relate to our current notions of explanation, even to be coherent about the extent to which that phenomenon cannot be explained.

A number of researchers claim that the most important aspects of intelligence are not and, in principle, cannot be modeled, and in particular not with any symbolic representation. These areas include learning, understanding natural language, and the production of speech acts. An AI program is “successful” if it performs according to its specifications. There is no requirement that a program must generalize its results, transfer to related situations, or unless required by its specifications, be transparent to its human users. Further, the AI community of program designers and builders relies on the scientific method as articulated by Thomas Kuhn. This tradition examines data, constructs models, runs experiments, and evaluates results. Experiments lead to refining models for further experiments. This scientific method has brought an important level of understanding, explanation, and the ability.

In a humbling way we are asked to resolve Aristotle’s tension between *theoria* and *praxis*, to fashion a union of understanding and practice, of the theoretical and practical, to live between science and art. AI practitioners are tool makers. Our representations, algorithms, and languages are tools for designing and building mechanisms that exhibit intelligent behavior. Through experiment we test both their computational adequacy for solving problems as well as our own understanding of intelligent phenomena.