

A History of Artificial Agents

by Matthew Louis Mauriello

The aim of Artificial Intelligence is that of creating agents. At the most basic level, these agents should be able to complete assigned tasks and reach goals. However, the larger goal is to create an artifact that can imitate intelligent human behavior. [1] Formal thought and research has been conducted for potentially thousands of years as we have attempted to understand our own intelligence. As such, there is a rich history in terms of artificial agents and artificial intelligence.

Description

The term agent comes from the Latin *agere*, which translates to “to do”. Agents are artifacts that serve a purpose (i.e. a computer program, a thermometer, and a calculator can all be said to be agents). The agents can act upon their environment, which can be either physical or digital as well as anything in between. The agent receives data from environmental inputs that drive an internal program, or apparatus, that results in actions. These actions then impact the environment. Precepts and actions are interpreted by sensors and performed by actuators respectively. A simple reflex agent is displayed in Figure 1. The internal program has three major components: an interpretation of sensor data, a set of logic conditions, and a list of possible resultant actions. Ideally this process happens within a loop [2].

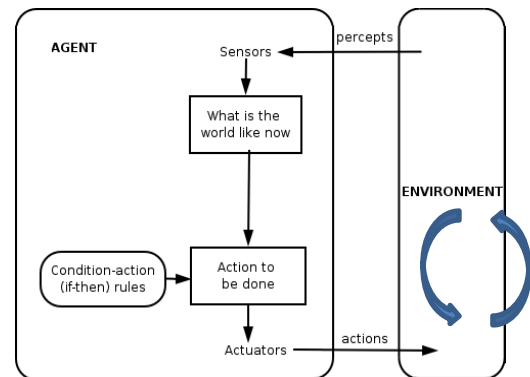


Figure 1: Simple Reflex Agent
Modified from: <http://www.wikipedia.com>

Background

As a species, Man has always thought of creating agents that can do things for us. Homer wrote of the mechanical “Tripods” that served Greek gods. Golems were reasonably intelligent artificial beings from the Jewish tradition. You can also find any number of examples of artificial intelligent agents in fiction [1]. There is a collective human interest in developing artificial beings and it is this interest that has sparked many debates: whether or not it is possible to create a new intelligence, the inner workings of how our own intelligence functions, and a great many ethical questions to consider (e.g. how these intelligences will be treated by human society).

Many of these questions were first asked by early Greek philosophers and mathematicians around 428 B.C. This is the time of Plato and Socrates who concerned themselves with trying to understand human nature, the rules of the physical world, and how formal rules can be logically applied to generate conclusions. The mind-body problem has its roots in this early time period. This problem is the debate over the fact that there is the physical brain, or body, and this construct that we call the mind. The mind is believed to be our soul and it exists separately from the body [2]. Although a body could be built, it would never obtain knowledge since it lacks a mind (or soul) and therefore could never think. It is in this ancient time that the debate of whether or not it was possible to construct an artificial agent originated.

Aristotle began studying how the rational mind behaves and generated a syllogism of logic that was only replaced by propositional logic some 2000 years later [3]. Aristotle proposed that his form of logic could be carried out via a mechanical process if such an artifact could be built [2]. In the 1500s, Leonardo da Vinci designed an early mechanical calculator. Although he did not build the calculator, it was later shown to have worked [2]. Da Vinci

would also go on to use his advanced mechanical knowledge to build one of the first mechanical men (i.e. suits of armor that moved like humans through a pulley and gear system) and even a mechanical lion. Table 1 provides a summary of interesting agents and automatons throughout history.

Century	Description
1300s	Albertus Magnus creates mechanical talking head; many such automata were created as a parlor tricks and destroyed as heresy. Demonstrates early work to recreate human capacities in machines.
1400s	Johannes Gutenberg creates the printing press. Very important machine that follows the agent principle highlighted in figure 1.
1500s	Clocks are made by the first clockmakers. Leonardo da Vinci works as a siege engineer, inventor, and artist. Da Vinci's works includes several mechanical automatons that could replicate natural motion. Juanelo Turriano constructs several human automata, most notably "the Lute Player".
1600s	Pascal and Leibniz are both working on calculators. Descartes proposes Dualism in opposition of Materialism. Christiaan Huygens builds the pendulum and several automata simulating soldiers.
1700s	Mechanic toys and the rise of automation: The Digesting Duck by Jacques de Vaucanson, Henri Maillardet's automaton for drawing/writing, and a (phony) chess playing machine known as "the Turk".
1800s	Charles Babbage develops the Difference and Analytical Engines. Gottlob Frege develops propositional logic.
1900s	An explosion of research is conducted into agents and artificial intelligence after the advent of the electronic computer. John McCarthy's Dartmouth conference on Artificial Intelligence occurs. International Business Machines (IBM) Corporation creates an agent called Deep Blue (Chess). Sojourner, an anonymous robot made by NASA, is deployed to Mars.
Recently	DARPA Grand Challenge won by Stanford's autonomous vehicle. IBM creates an agent called Watson.

Table 1: History of Automata [8]

In the 1600s, the philosophical debate of the mind-body problem continued with Rene Descartes' concept of dualism. Dualism argued that we are more than the sum of our parts and that there was a part of us that exists outside of the physical world. Animals and the human body are simply biological machines operating at a very basic level. Humans, Descartes believed, are endowed with a mind of consciousness that cannot be imitated. Though, Descartes seems to have believed that the mind could survive without the body. Descartes also believed in "free will", which he says can only exist within the mind and cannot rely strictly on physical rules to operate [3]. Descartes logic argues the impossibility of intelligent artificial agents, because the physical machine can not possess a mind. Materialism, a rival to Descartes' thinking, believed that all of consciousness and of "free will" could be explained mechanically.

Gottfried Wilhelm Leibniz also worked in the 1600s. Leibniz rejects the ideas of materialists and dualism believing that we are one unified being more than the sum of our parts. Leibniz focused on perception and language and he believed that a language could be created that uses symbols as stand-ins for concepts¹ and ideas. This language would have logical rules that could be manipulated to produce valid conclusions. He believed that this language could perfectly imitate intelligent reasoning and some say that Leibniz was anticipating the possibility of artificial intelligence [6]. This is an important marked change in thinking. Leibniz allows us to believe that consciousness was the result of unity of both body and mind in one physical being. Leibniz proposed that intelligent reasoning could be fully represented through language and that consciousness results from our ability to perceive the world. Leibniz's ideas encourage the development of artificial agent.

¹ Leibniz developed a mechanical calculator that instead of operating on numbers operated on concepts [2]. Leibniz would also develop binary number system that provides the basis for modern day computing.

Moving closer to modern day, Charles Babbage in the 1800s designed two early mechanical computers known as the Difference Engine² and the Analytical engine. Both computers would have read in data from punch cards, performed some internal calculations, and printed out an answer in the same way that the agent of Figure 1 highlights. These designs are considered the forerunner of the modern electronic computer like the ENIAC (developed in 1946).

Alan Turing and John McCarthy are the fathers of modern day artificial intelligence as a discipline in computer science. Turing is famous for developing the Turing test, giving us a somewhat controversial measure of how to validate our success in developing artificial agents that are indistinguishable from humans. McCarthy and his colleagues decided to break the study of the mind into various sub-problems (i.e. Natural Language Processing, Computer Vision, Information Retrieval, etc.). McCarthy thought that reassembling the work of these sub-problems would yield better results than working on problem in its entirety [7].

Recent Successes



Figure 2: Watson's Avatar
Source: <http://www.ibm.com/watson>

International Business Machines (IBM) Corporation has been pushing the envelope in terms of agent capabilities. Watson, a product of IBM's DeepQA project, was a Jeopardy opponent that defeated the best players in the world in 2011. Watson is an amazing example of a successful agent that combined Natural Language Processing, Information Retrieval, Parallel Computing, and a number of other disciplines making it a sophisticated computer system. Watson was also able to bring to cutting edge capabilities of Artificial Intelligence to the fore front of public attention and imagination [4].

Watson also represents significant progress towards a goal that was once believed to be impossible. It is important to think back on our literary traditions and the associated mythos connected to the "Golem of Prague". In these stories a Golem is created out of artificial components to perform a task. However, golem stories are often about hubris. The creators of Golems lose control of their creations, which results in tragedy and death [5]. There is a hidden warning to us in these stories, to be careful when trying to create artificial beings. In closing, beware of Golems, Skynet, and Cylons!

References

1. Wooldridge, M., & Jennings, N. R. (1995), *Agent Theories, Architectures, and Languages: a Survey*. Intelligent Agents, Berlin: Springer-Verlag, 1-22
2. Russell, S. J., & Norvig, P. (2010). *Artificial intelligence, a modern approach*. (Third ed.). New Jersey: Pearson College Div.
3. (2001) *Descartes*, The Internet Encyclopedia of Philosophy, retrieved 2012-09-12. Online: <http://www.iep.utm.edu/>
4. Brodtkin, Jon (2010-02-10), *IBM's Jeopardy-playing machine can now beat human contestants*, Network World, retrieved 2012-09-12. Online: <http://www.networkworld.com/news/2010/021010-ibm-jeopardy-game.html>
5. Franchi, S., & Guzeldere, G. (2005), *Mechanical bodies, computational minds: artificial intelligence from automata to cyborgs*. Massachusetts: MIT Press.
6. (1997) *Leibniz's Philosophy of Mind*. Stanford Encyclopedia of Philosophy. Retrieved 2012-09-12: <http://plato.stanford.edu/entries/leibniz-mind/>
7. Perlis, D. (2012), *Class Notes on 2012-8-30*. Human Level AI and Computational Cognitive Neuroscience. University of Maryland.
8. Buchanan, B. G. (2012). *Timeline: a brief history of Artificial Intelligence*. Association for the Advancement of Artificial Intelligence, retrieved 2012-09-16. Online: <http://aitopics.net/BriefHistory>

² Ada Lovelace was the first "Programmer" writing programs for the engine. The Ada language is named for her.