# AI - What is this

**A Definition of Artificial Intelligence**

November 2000
this paper is a part from AI- Project

Dimiter Dobrev
d@dobrev.com
illustrations - Konstantin Lakov

In this paper we are going to discuss the following questions: "Do we have to know what is AI?" and "What is intelligence?". After that we are going to give a definition of Artificial Intelligence. Finally, from this definition we are going to get an algorithm which after a final number of steps will discover AI.

**D**o we have to know what is AI? This question can be easily answered: Yes, if we want to find it then our task will be a lot easier if we know what is the thing we are looking for. Failing to define AI, our position will not differ from that of the Alchemists who sought for the Philosopher's stone but almost had no idea what they were searching for.

The most widely spread definition of AI is the so called Turing's test. Alan Turing was a British mathematician famous for the invention of the theoretical Turing machine and for the deciphering of the German codes during World War II.

The Turing's test is quite simple. We place something behind a curtain and it speaks with us. If we can't make difference between it and a human being then it will be AI. However, this definition exists from more than fifty years, so we are going to create a newer and a more up-to-date one.

Turing's definition suggests that, an Intellect is a person with knowledge gained through the years. If this is so, then what about a newly born baby? Is it an Intellect? Our answer will be "yes". Our definition of an intellect will be: a thing that knows nothing but it can learn. At this point we differ from most people who imagine a university professor when they hear the word Intellect.

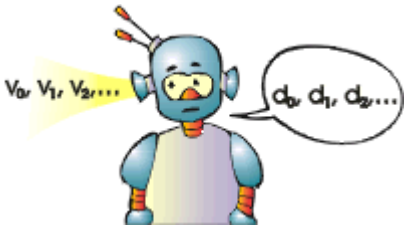
Before giving a formal definition of AI we will make it clear that we accept the thesis of Church, stating that every calculating device can be modelled by a program. This means that we are going to look for AI in the set of programs. We will suppose that AI is a step device living in a kind of world. At each step it receives information (from the world) and influences (at the world) by the information it works out. Also, we will assume that the information received and worked out at each step will be a finite amount. Let's say it gets **n** bits and works out **m** bits.

After this clarification we can state informally our definition. AI will be such a program which in an arbitrary world will cope not worse than a human.

The next task will be to formalise this definition in order to use it and to search for AI with it. First, what is a world for us? These will be two functions **World(s, d)** and **View(s)**.

The first will take as arguments the state of the world and the influence that our device has on the world at this step. As a result, this function will return the new state of the world

(which it will obtain on the next step). The second function will inform us what does our device see. An argument of this function will be the world's state and the returned value will be the information that the device will receive (at a given step). Also, we have to add one $\mathbf{s}_0$. It will be the world's state when our device was born. During its life the world will go through the states $\mathbf{s}_0$, $\mathbf{s}_1$, $\mathbf{s}_2$, ... . The device will influence the world with the information it works out at each step $\mathbf{d}_0$, $\mathbf{d}_1$, $\mathbf{d}_2$, ... . Also, AI will receive information from the world $\mathbf{v}_0$, $\mathbf{v}_1$, $\mathbf{v}_2$, ... . It is clear that $\mathbf{s}_{i+1}=\mathbf{World}(\mathbf{s}_i, \mathbf{d}_i)$ and $\mathbf{v}_i=\mathbf{View}(\mathbf{s}_i)$.




We have everything up to this moment. We have a world and a device that lives in it. However, there is one thing missing - the meaning of life. What is life without pain and joy, a philosopher would say. That is why we will introduce meaning of life. This will be an evaluation to tell us whether one row $\mathbf{v}_0$, $\mathbf{v}_1$, $\mathbf{v}_2$, ... is better than another.

Most people think that they have spent their life better if they have seen more Swiss resorts and less coal-mines. More or less our definition of the meaning of life will be the same. We will pick out two bits from $\mathbf{v}_i$ and call them victory and loss. The aim will be to get more victories and fewer losses.

The last step will be to make an algorithm that will discover AI. The idea is to start all programs on all worlds and to take those which cope best. This does not sound as an algorithm, it is as if we are going to make a never-ending exam with an infinite number of candidates. The problem is not the infinite number of the candidates, we do not need all of them but only a part from those who have passed the exam. The real problem is that the exam will never end even for a single candidate.

To make the exam finite we will make finite the number of the worlds. This will be not enough because even for one world the exam would be infinite. That is why we will add requirements for efficiency. We will give for each world a limit of time (tacts) the device has for training and we will also say what results it has to achieve after this training (for example, in the next hundred tacts the relation victory to loss to be at least 9 to 1). We will also restrict the time and memory available to the device for a step. These requirements should not be too tough because our AI will not satisfy them and will fail the exam.

We will also suppose that there will be no fatal mistakes in the worlds chosen for the exam. If there were, our AI could make a fatal mistake during his training and fail the exam. Of course, if AI lives many times in such a world it will get a good average score but we want to let it live only once per world.

The real world does not satisfy this requirement simply because one could be Einstein but to be eaten up by a bear before realising the Theory of Relativity, i.e. to make a fatal mistake before being trained. An example for a world without fatal mistakes will be if our device plays chess against someone else. After each game the next one begins. In this world AI can lose many games but this will not have any influence on the next.

After making the exam final our algorithm can start generating the programs one by one, to test each and to take out only those that have passed the exam. We will consider that our algorithm orders the programs according to their complexity (i.e. according to their length). That is to say, it will take out first the simplest (shortest) program from those that have passed the exam. Will this program be AI or AI will be generated later on? We have to point out that not all programs that will pass the exam will be AI. For example, if we write a program especially for the worlds on which we test it will pass the exam but it will notbe AI. We have the same problem with the students' exams. Many people who have learned all the tasks by heart will pass the exam but this people are not intellects but crammers.

We will consider that the worlds included in the exam are enough numerous and varied (a bigger number of tasks does not make the exam harder for the examiner because most of the examinees will fail at the beginning).In this situation almost all programs that are not AI will be sifted out. As a result, the first (the simplest) which will be worked out will be AI and the other written only for the exam worlds will be more complex and will be worked out later on.

So, we already have an algorithm for discovering of AI! Does it mean that every decent programmer can write a program on it, to start it, to wait for a while and it will work out the desired AI? Yes, but the time needed would be quite a lot (let's say a hundred - thousand years). The reason is in the phenomenon called combinatory explosion. It is not of great difficulty to be written a program expected to stop after the end of the universe (for example, you can increment a ten bit counter until it overloads).

The written above means that the algorithm described is entirely useless but it is not so with the definition of AI. After learning what is AI we can try to build it directly and use the algorithm to make sure that this really is AI.

After the successful testing of our device in all test worlds we can put it in our (real) world. Of course, it will not be ready immediately for the Turing's test, because at the best case only a "bill and coo" will be heard behind the curtain. Our device will need first tutors and governess to teach it of good manners. The teachers should encourage it by giving one to its line victory and punish it by the line loss. Thus, our AI will work hard trying to get a maximum encouragement and a minimum punishment. Probably then the computers will not be programmed but educated and trained. And maybe one day, after having educated and trained them we will become useless.