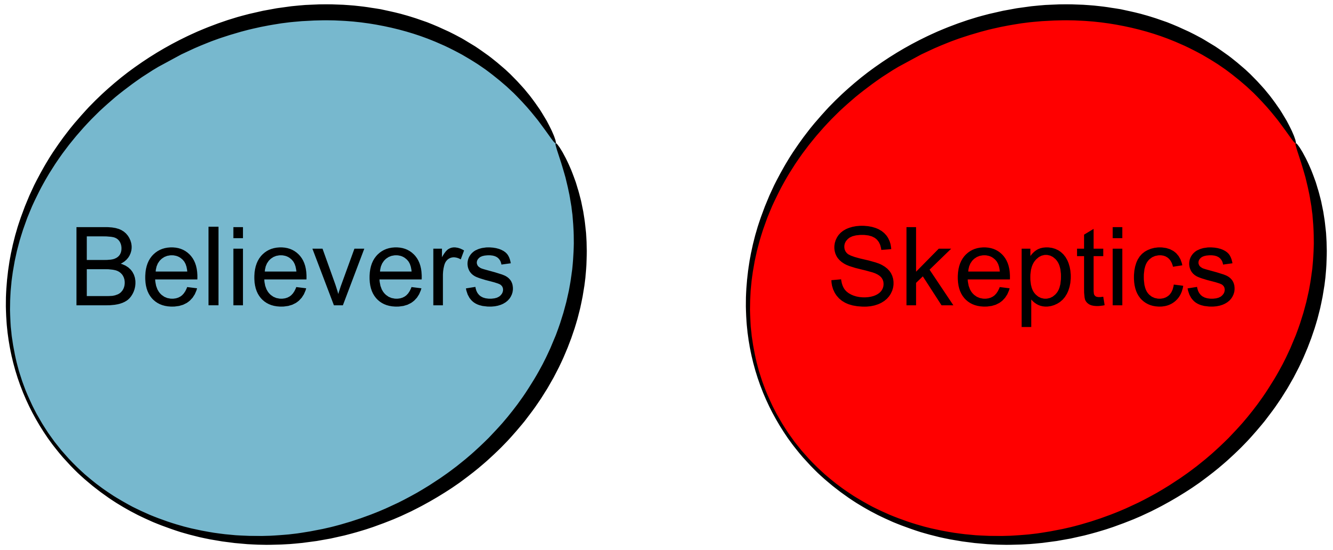
One talk about the

Artificial Intelligence

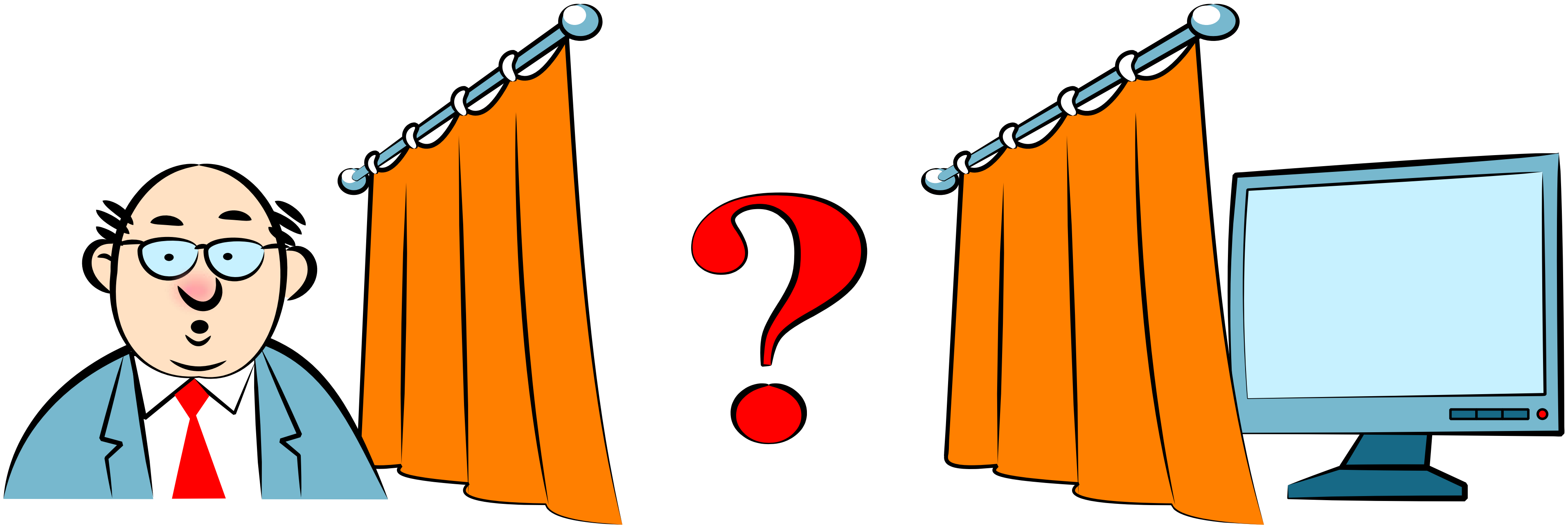
by

Dimiter Dobrev
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences

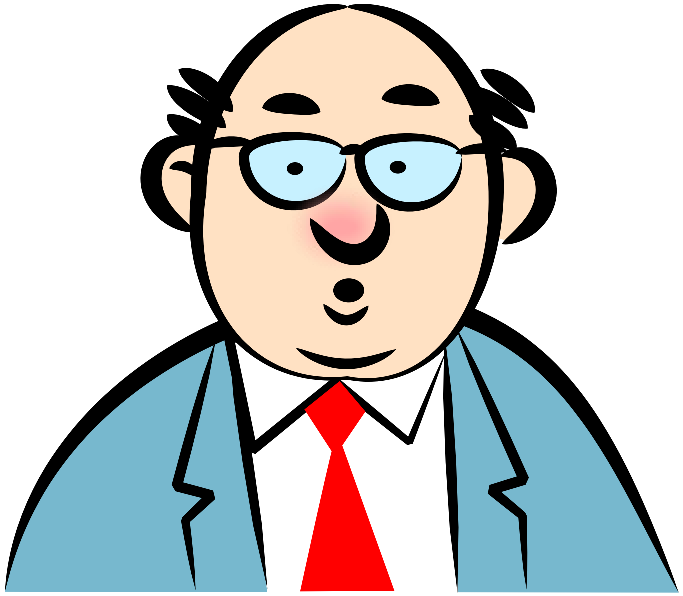
AI divides the people in two groups:



If you are a believer and believe that AI can be made, then this talk may be interesting for you. If you are one of the skeptics, everything I’m going to say now will be nonsense for you.



When you want to construct AI then the first question is: “What is this?” One of the first AI believers was Alan Turing. He gave the first definition of AI which is known as the Turing test. He said that if behind a curtain we have a human being and a computer which pretends to be a human and if we cannot distinguish a man from the machine this means that the machine is AI.

This definition has several shortcomings. First, it cannot distinguish the intelligence from the knowledge. This means that this definition defines a trained intelligence but what we need is just an intelligence because we know how to train it if we have someone to train.

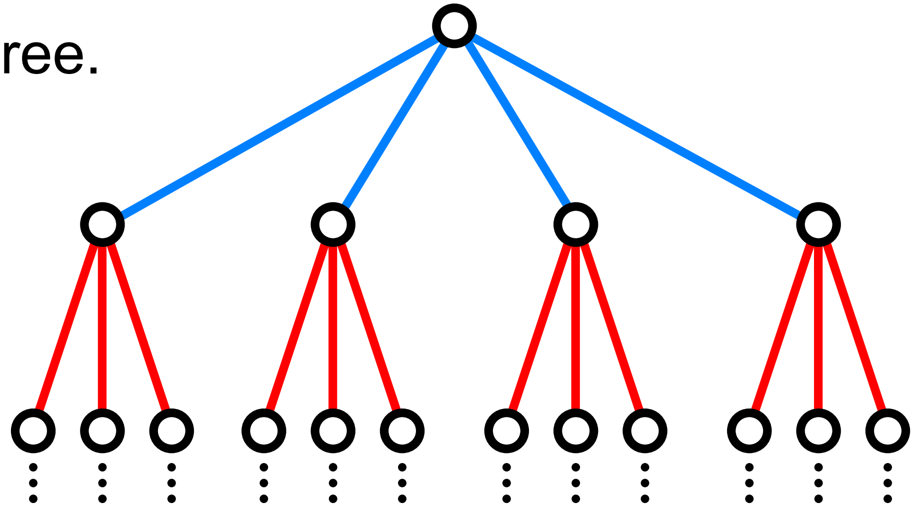
Let’s take for example an university professor and a baby. Which one of them is an intelligence? For the professor, everybody will agree that he is an intelligence. We will assume that the baby is also an intelligence because we know how a professor can be made from an average baby.

The second disadvantage is that this is a functional definition which describes one program through its input and output. If we want to create an Artificial Human, then we have to describe his entire life, not only a moment of it. This means that just a single frame is not enough, we need the entire movie.

Turing was aware of the disadvantages of his definition but he died mysteriously before he managed to improve it.

Before we make an Artificial Human we have to do some digitalization. Let us represent the human life as a digital object. First, we will assume that time is discrete. When you watch a movie you actually see 24 frames per second but you do not understand that the movie is discrete. In the same way you cannot tell the difference between an analog and a digital movie. So we can assume that we can digitalize your entire life.

Let’s represent your life as a tree.



Here the root is the moment of birth. At any step you have finitely many inputs and finitely many outputs. Actually, this is the tree of all possible lives which you could have. Your current life is the path between the root and the current moment.

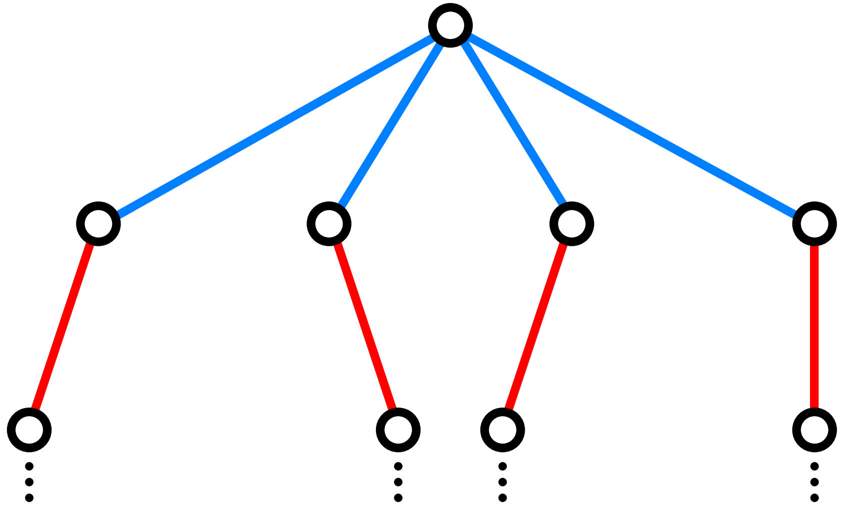
Now we are ready for our first theorem:

Theorem 1. Artificial Intelligence exists.

Now we are ready for our first theorem:

Theorem 1. Artificial Intelligence exists.

After the digitalization we can represent a human being as a strategy in the tree of all possible lives. Strategy is a subtree where only one action is allowed at any situation. The only question is: “Is this strategy computable?” Let’s make the tree finite. To do this, we will restrict the life to one hundred years. Then the tree is finite and all strategies in it will be finite. So, any such strategy is computable.



This result is trivial and absolutely useless. In the same way I can say that if we take the set of all possible number combinations for the lottery, then one of these combinations will win the jackpot. Actually, you don’t want to know that the combination exists. You want to know which this combination is.

I cannot help you determine which the winning combination for the lottery is but I can give you a definition of AI, which is the worthier of the two.

Of course, you do not want to have the AI strategy as a digital object because this object is so big that it cannot be stored in the computer memory. You will say: “One hundred years, 24 frames per second, this is not so much information.” Actually, this is nothing compared to the entire strategy. The problem is that the strategy is a tree and the life is only one path in this tree. So, the tree is huge.

Instead of having an AI strategy computed in advance, you would prefer to have a program which can calculate this strategy.

Before we make this program, we need to give the definition of AI.

First idea for definition:

Let AI be the best strategy.

This definition is OK but it raises the bar too high and it will be extremely difficult to make a program which can calculate this strategy in a reasonable time.

Here is a better idea for definition:

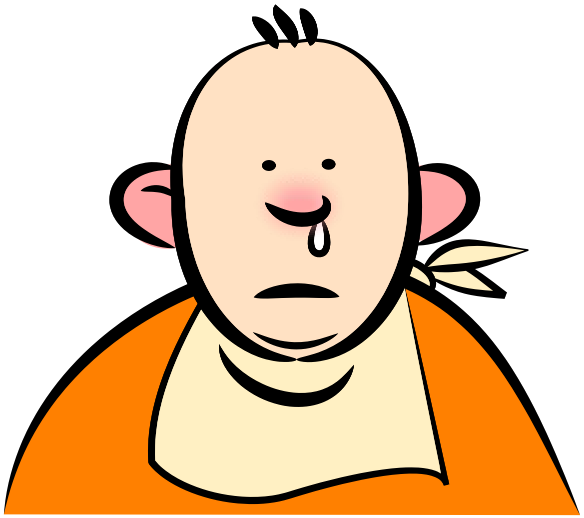
Let AI be any strategy which is not too much worse than the best one.

This definition is better, but we do not know how good the best strategy is, so we cannot say for one strategy whether it is too much worse than the best one or not.

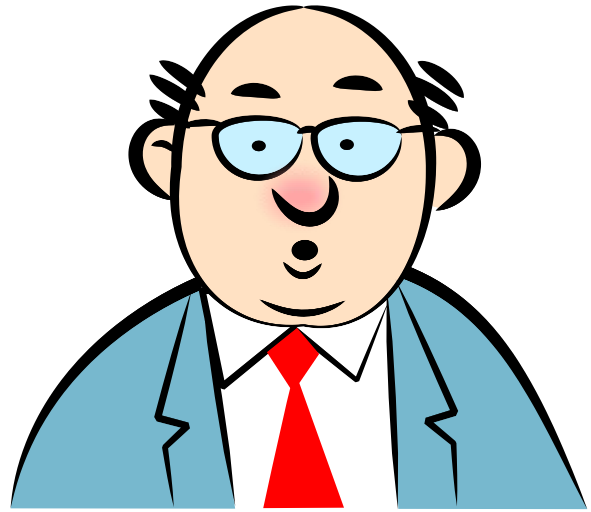
Next idea for definition:

Let AI be any strategy which is good enough.

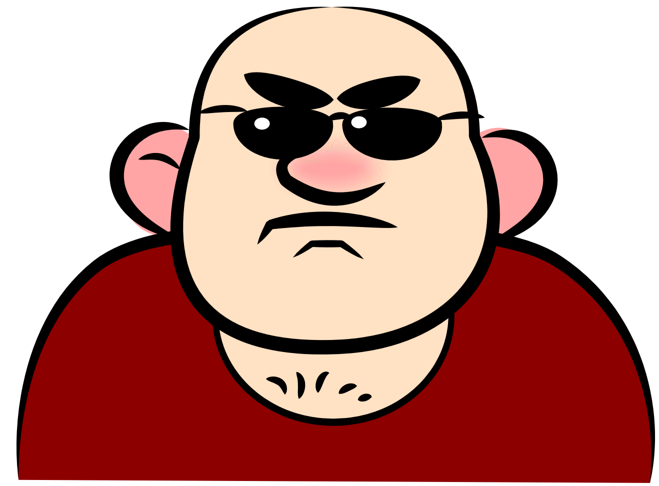
For “good enough” we will understand that it is better than a human being.



Vegetable



Professor



Godfather

For the notion of good and bad strategy to have a meaning, first we have to say which life is better. This means that we have to introduce the notion of meaning of life.

Let’s take three persons.

A Vegetable, an university professor and a godfather of mafia.

Which life was better. We will assume that one’s life is better if more good things and less bad things happen. Let’s assume that a good thing is to see a Swiss resort and a bad thing is to work in a coal-mine. In this case the life of the godfather is worse because at the end of his life he is in prison and he works in a coal-mine. The university professor sees a Swiss resort only when the Logic Colloquium is in Switzerland. The best is the life of the vegetable because it lives in a hospital for vegetables which is in a Swiss resort. Of course, this resulted from the way we chose the meaning of life. If our criteria was who makes more money then the order would have been the other way about.

There is yet another problem. We are looking for a strategy for the real world, but this world is too complicated and we cannot emulate it by a computer program. Often the general case is simpler than the special case and that's why we will replace the real world by an arbitrary one.

Last idea for definition:

Let AI is any strategy which is good enough in arbitrary world.

Several questions arise here, such as: "Which is the set of all possible worlds?" and "Which world is more important and which – less important, i.e. what weight will each of these worlds participate with?"

We will assume that the set of arbitrary worlds consists of all computable worlds. Such world will be any program which inputs the last information which the device saw and the last action which the device made at each single step. On the basis of this input and its internal state, the program outputs a forecast for what the device will see on the next step.

We can assume that the forecast is a concrete action but it is almost impossible to find an absolute model which gives a perfect forecast each time. That is why, we will assume that the forecast is a set of actions with different possibility for each action.

Why do we assume that all worlds are computable? In any case, we cannot use a model which is not computable so we have to model the real world only with computable models. What will the weight of the different worlds be? We will assume that all programs have equal weight. Yes, but we want to use the Ockham's Razor principle which says that the simpler models are the more probable ones. Actually, this will be true because the simpler models are calculated by more programs.

Now we are ready to show a program which satisfies the AI definition.

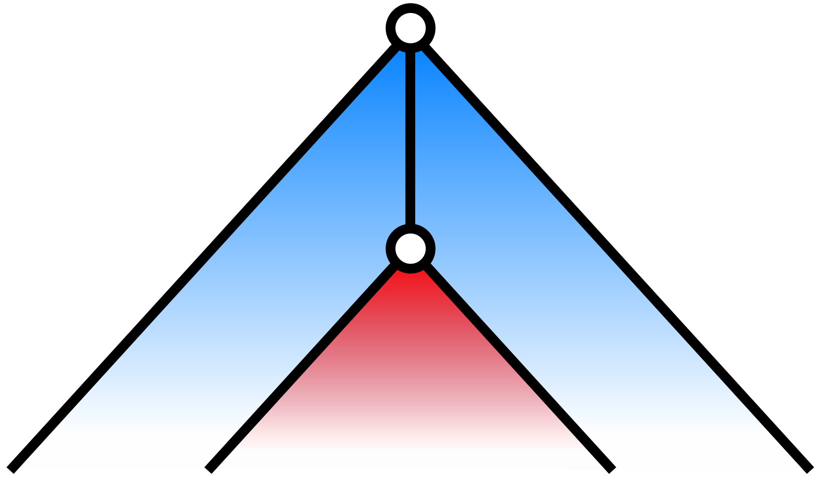
What result do we need – a practical or a theoretic one? Of course, we prefer to have a practical result because we want to sell robots and make money but for the moment I can show you only a theoretic result. Anyway, at the end of this talk I will give some ideal how this theoretic result can be improved and how to obtain a practical result.

Here we have one problem. We want to make an AI which is good enough but we do not know how good it should be. So we will introduce one parameter which we will call “level of intelligence”. If this parameter is higher, then our AI will be more intelligent. So we know that for some level of intelligence our AI is good enough but we do not know which this level it is.

Actually, the level of intelligence will consist of two numbers: k and n. The first one will be the number of computable models which we will use. The second one will be the number of steps which we will look in the future.

How will our program work?

Let us assume that we are at some current moment and we have to choose what the next action of our device will be.



We will take the first k computable worlds and for each of them we will compute the possibility for this world to be a correct model. For this we have only to multiply all forecasts which this model gave for the future. Of course, for this multiplication we will use the possibility of exactly that forecast which actually fulfill. So, the result of this multiplication will be some very small number, usually zero.

After this we will go n steps in the future by the Max-Average algorithm and for each model and for each possible action we will calculate the expected success if this action is made. These numbers have to be multiplied by possibility for the world to be a correct model and we have to make a sum over all k models.

In this way we will determine which action gives the maximal expectation and this is the action which our AI program will choose.

Therefore, here you have a description of a program which satisfies the definition for AI.

This statement sounds so strong that it will be immediately ignored by the audience.

People use to hear very strong statements of the type:
We produce the best washing powder.

Everybody ignores such statements because they assume that they are not true.

In order to make you believe that this statement is true, I have to convince you that it is not as strong as it looks.

Really, this program is AI but this is only a theoretical result. In practice this program is useless because it leads to a combinatory explosion. From the point of view of the theory it is interesting that we have such a program but it is useless in practice.

In order to understand this, look at the following example:

We want to find the telephone number of our friend Peter. He lives in Sofia and his telephone number has seven digits. We can dial 10 million telephone numbers and ask: “Peter, is that you?” This means that from theoretical point of view we can find this number but in practice this algorithm is useless.

How can we improve this algorithm in order to make it applicable in practice?

First, in order to predict the future we have not to use the Max-Average algorithm. This algorithm is a generalization of the Min-Max algorithm, which works fine in the chess playing programs but when we have to look further in the future this algorithm is not applicable.

Second, we have to change the set of models in which we will search for the model of the real world. The set of all programs is a good idea only for theoretical purposes.

One good idea for the set of possible models is to use multi-agent worlds instead of single-agent worlds. For more information about this I will recommend my last paper:

www.dobrev.com/AI