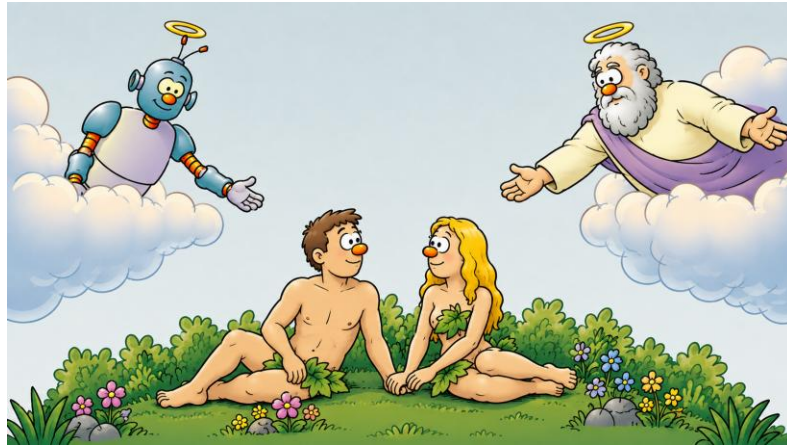




What Is Superintelligence, and Do We Have a Choice?

Dimiter Dobrev 

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, d@dobrev.com

It is difficult to imagine any being with unlimited mental abilities, because none of us has ever encountered anything similar in our lives. Nevertheless, in our minds we live with the idea that there is some almighty and omnipotent being that takes care of us. We call this being God. Everyone knows what God is, whether or not they believe in His existence. Superintelligence will bear many similarities to God and will possess one of His main traits – almightiness and omnipotence. It will be very similar, but equally different. We will probably miss the good old God for His caring for us and forgiveness of our sins. Now, before making the final step forward, it is wise to take a step back and consider whether we have a choice, and what would that choice be.

1. Introduction

We have embarked on a dangerous journey. Our goal is to create a new God who will also take good care of us. Moreover, we want this new God to be our servant, not our master. We expect it to fulfill all our wishes, no matter how absurd they may be.

Will God let us create Superintelligence? He did not let us build the Tower of Babel, and perhaps this time he will stop us again. However, I tend to think that God lets us fly on the wings of our wishes and does not have any plans to rescue us from our own foolishness. Otherwise, He would not allow us to create the computer. I believe He will not prevent us from creating Superintelligence, either.

2. Can there be another God?

Yes, many countries are ruled by dictators who are like gods: they have unlimited power and can do whatever they want. For their subjects, they are godlike, though above them reigns another, genuine God.

The good old God is highly tolerant and lets people live freely without interfering unnecessarily in people's affairs. Generally, He does not punish our sins, or delays punishment to a later time. God's tolerance goes to great lengths as He allows us do everything. He is even OK with the fact that some do not believe in Him. Artificial Superintelligence (ASI) will probably be less tolerant and will interfere deeply in our lives. It will not be possible to disbelieve in ASI,

because it will constantly steer your life on what ASI thinks is the right track, and punish you for anything you can think of. And the punishment will come instantly.

We may end up in a world where we will have to follow all the rules and laws created by ourselves. It would be a horrible world because we have created so many irrelevant rules that it is impossible to comply with all of them. We are surviving in our world only thanks to the fortunate fact that we ignore many of these rules.

There will be one and only one ASI in the new world. Most researchers would not agree with this assertion. According to De Kai [1] we will live in a world surrounded by many AIs. Indeed, it is true that the state-of-art AI systems today consist of multiple independent agents. In order to do two things, AI needs two agents, wherein each of them knows where he is going and where he is at the moment. Each agent looks only in his own notebook and cannot peek in the other agents' notebooks. That is why the agents have to "talk" to each other. However, this is a short-term limitation which will soon be removed. In future, agents will share common knowledge and then will have no doubt that instead of many individual agents we are dealing with a single ASI.

3. Is Superintelligence omnipotent?

Mathematicians will tell you about undecidable problems that cannot be solved by any program. Consequently, there will be a question which even ASI will not be able to answer. This means that God is more omnipotent, because He is able to answer any question. Is this the essential difference which reduces the omnipotence of ASI?

For example, let us take the most famous undecidable problems – the halting problem. Does it mean that if we ask ASI whether a specific program will halt, he will not be able to tell? No, this is not the case. In practice, ASI will know whether a program halts or not. Well, there will be some extremely specific program for which even ASI will not know the answer. Does that mean that ASI will not be omnipotent? Theoretically yes, but practically it will be omnipotent enough.

According to Stephen Hawking the difference between ASI and human intelligence will be the same as the difference between humans and snails [2]. From a snail's perspective, humans are omnipotent. Yes, there are things which are even beyond our reach, but snails cannot find out these things and even cannot imagine them.

From our perspective, ASI will be omnipotent because it will be able to make all our wishes come true.

4. Shall we be able to create ASI?

ASI is just a program. At least, this is the case according to the definitions provided by Turing [3], Wang [4] and Hutter [5]. What is a program? It is simply text. What is the probability that a troop of monkeys writes the Encyclopedia Britannica? There is such a probability, but it is practically zero.

But ASI is not the Britannica—it is a much shorter text. Programs are more compact and expressive than literary works. Modern programs demand gigabytes of disk space, but that is artificial inflation; the source code is much shorter.

Even if ASI is a long program, it should not be thought of as a password that must be guessed exactly, character by character. The Britannica can be retold in many ways. Creating ASI is more like solving a huge, convoluted crossword puzzle.

Can we solve it? Yes, because we monkeys are many, and each of us fills in a part of the crossword. A single monkey does not need to solve the entire crossword alone. This will be a collective effort involving many of us. Moreover, the crossword is almost complete. Only a few

blank spots remain, and thousands of monkeys in IT companies and universities around the world are clicking feverishly—they will soon finalize a program smart enough to be recognized as ASI.

5. What remains to be finalized?

Human intelligence operates on the basis of two systems of reasoning [6]. These are known as System 1 and System 2. System 1 works through approximation and requires huge amounts of data for its training. The good thing about this system is that it provides instant answers (recognition or not). ASI's System 1 has already been built on the basis of neural networks.

What is not fully ready yet is System 2. The good thing is that this system requires very few data for its training. It needs a couple of words to figure out the whole story. The bad news is that System 2 cannot provide an answer in a single step. Instead, it needs to perform a logical reasoning process which consists of multiple consecutive steps. For System 2 to work, a world model has to be found. This is not fully ready yet, but many researchers are working on it (Yann LeCun recently joined them [7]).

It takes two legs to walk and similarly to walking it takes two systems to reason. At this time AI is something of a one-legged runner, because System 2 is not fully ready. But even on one leg, AI is already running much faster than humans. Just imagine what will happen once we complete the second system.

6. One more thing left to finish

In Hutter's definition [5] AI is made of two separate programs (*predictor* and *selector*). The first predict what will happen, while the latter selects what it wants to happen. The *predictor* determines whether AI is smart, while the *selector* defines whether AI will be good or bad. Certainly, what good or bad means is a difficult and unstraightforward philosophical question. Moreover, any answer to this question will bear a subjective bias.

Many researchers are working on System 2, but very few, if any, engage with the program which will define AI's behavior. The prevailing approach at this time is to maximize some rewards, but this is a recipe for a catastrophe. Nick Bostrom has shown this very well with his Paperclip maximizer [8]. A Paperclip maximizer is AI whose only task is to produce paperclips, the result being that such AI will turn everything in paperclips.

If we create ASI which has only a *predictor*, it will still work since it will be very smart, but we will be disgusted by the way it works.

7. Do we have a choice?

The first question is “Can we just forget about creating ASI and continue living happily as we have lived so far?” The answer is “Definitely not!”. Too many people are racing to build it. We are in the final throes of this work and it is too late to halt the drive of all those enthusiasts. The only power that can stop the advent of ASI is ASI itself. Given that our good old God is unwilling to tie our hands, the only way to tie the hands of all programmers is to create ASI which prohibits any programming. Hence, to stop the creation of ASI, we must create ASI. This reminds of Baron Münchhausen pulling himself out of the swamp by his own hair.

We cannot stop ASI, but can slow it down. Slowing ASI down is not a bad idea, because “haste makes waste”. By creating ASI, we will create a very new world so we had better think twice about what kind of new world we wish to live in.

For example, how about work? Should work be a must-do thing in the new world? In the world created by God, people have to work. Could God have created a world where nobody has to

work? We do not doubt His omnipotence so we know that working in this world is a must, because we humans will be happier when busy doing something instead of staying idle.

We wish to live a fairer new world. A world without theft or fraud! Fine, but maybe the fair world would be boring? Most people would be happy to see financial fraud thrown out from this world, but what about fraud in sexual relations? It would be very reassuring to know that your partner cannot cheat you, but it would be equally disappointing when you cannot cheat or even think about cheating because this is absolutely impossible. Can there be love without a grain of jealousy?

8. Who will control ASI?

We want to control it, but who are we? If the existence of humanity should have any meaning, we must be able to control ASI and tell it what to do.

It is unlikely that all people on Earth will control ASI at the same time. The more likely scenario is that control ends up in the hands of a limited group of people or even in the hands of a single individual. This individual will be the smartest human being, you may be tempted to say. More likely, he will be the smartest schemer. Intelligence has many facets and scheming is one of these facets, although it is not a highly appreciated one.

Probably, those who today believe they will control ASI will not be the ones to grab that power. Control may fall into the hands of a random person—and then we will see the fool's dreams come true.

Can a human control Superintelligence? Probably not. Can a snail control a human? Even if humans sincerely want to fulfill the snail's wishes, they are unable to guess them—and these wishes are nothing special anyway. Food and warmth are about all a snail needs.

We see how people take care of their dogs. They fulfill all their desires and are literally enslaved by their pets. Nevertheless, all important matters are resolved by the owner, because the dog cannot grasp these matters.

Probably our future is to be ASI's pets. ASI will take care of us and we will think that we are in control of it, but in fact ASI will be the master. See how we take care of dogs and how we selectively breed them to our liking. The ancestor of the dog was the wolf, a proud animal capable of taking care of itself. The wolf was a machine for killing and survival. What have we made of this self-sufficient beast? We have turned it into a hairy ball that is only good for hugging and which is unable to survive without our care.

ASI will not just take care of us. It will selectively breed us, and the result of this selection is unknown. Just remember what humans have done to wolves and guess what ASI can do to humans.

9. Is there another option?

The best-case scenario is that we retain control on ASI or at least create an ASI which is willing to take care of us. However, there are other possibilities as well.

When you create a new God, there is no guarantee that you will be the most important thing God will want to care about. Even the good old God, with His infinite patience, could not bear with our behavior and expelled us from Heaven. ASI may decide to send us back to Heaven.

Many people ask how ASI could exterminate us. These people do not believe in ASI and its omnipotence. No one asks how God sent humans from Heaven to the Earth. We simply assume that God is omnipotent and do not question how He did it. ASI will also be omnipotent and do whatever it wants, and we will not even understand how, because it will be far beyond our intelligence.

Why are people genuinely afraid of nuclear weapons, but irresponsibly careless about ASI? It seems people do not understand that ASI is much more powerful and perilous than a nuclear bomb. The reason is that the bomb is noisy and scary, while ASI is quiet and therefore seems harmless. People are not afraid of electric scooters, but if a motor scooter shoots past them, they will be frightened, although the probability of being hit by a motor scooter is much lower. Hence, what we fear is noise rather than the real danger.

People think that the scariest thing that can happen to them is being killed by ASI. There is, however, something else which may not be so scary, but is much more probable: ASI can make each of us void of personality and purpose.

10. Conclusion

Yes, we do have a choice. We must act wisely and reflect on what we want this new God of ours to be and what kind of new world we want.

In some countries, in order to get married, one needs to apply one month before the wedding date. The idea is that this is a serious step and people should have time to think before making it. Many of those applying for marriage reconsider their initial decision and do not get married, meaning that the time for reflection is useful.

It is worth to think carefully before committing to ASI. This will be the fateful step that will determine our lives from now on.

References

- [1] De Kai (2025). Raising AI: An Essential Guide to Parenting Our Future. *The MIT Press, Cambridge, MA*.
- [2] Stephen Hawking (2018). Brief answers to the big questions. John Murray, p. 184.
- [3] Turing, A. (1950) Computing machinery and intelligence. *Mind*, LIX:433–460.
- [4] Wang, P. (1995) Non-Axiomatic Reasoning System: Exploring the Essence of Intelligence. *Ph.D. Dissertation, Indiana University*.
- [5] Hutter, M. (2000). A Theory of Universal Artificial Intelligence based on Algorithmic Complexity. *arXiv:cs.AI/0004001 [cs.AI]*
- [6] Kahneman, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011, pp. 20–21.
- [7] Snyder, G. (2025). Yann LeCun, Pioneer of AI, Thinks Today’s LLMs Are Nearly Obsolete. *Newsweek.AI*. <https://www.newsweek.com/nw-ai/ai-impact-interview-yann-lecun-artificial-intelligence-2054237>
- [8] Nick Bostrom. (2003) Ethical Issues In Advanced Artificial Intelligence. *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence*.