The First AI Created Will Be The Only AI Ever Created

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

Our generation is the one that will create the first Artificial Intelligence (AI). We are the ones who will set the rules to which this AI will operate. Once these rules are set, they will be there forever, hence our responsibility is huge. There will be no chance of a second AI because the first one will take control and will not allow the creation of another AI. Our first and foremost concern is not to lose control of the first (and only) AI. Hopefully we will be reasonable enough and not let that happen. However, even if people retain control of AI, the question that comes next is who exactly will those people be? Should they enjoy the absolute power to issue whatever commands to AI they wish? Or should certain restrictions be embedded in AI at its very inception?

1. Introduction

Imagine you are setting up a new country. The first thing you need to do is write the Constitution of your new country. That Constitution will lay down the rules by which the new country will exist. The Constitution will hold for years, decades or even centuries ahead. Even if you decide to change it, such change will again be made to rules laid down in that very Constitution.

Creating AI is like creating a new country. The rules built into that AI will be the Constitution of that new country (or new world).

The first rule in that Constitution certainly will read:

Rule 1: I am your one and only AI and nobody else is allowed to create another AI!

Shall we allow AI improve itself? We should better avoid this. We already face the major risk of letting AI out control when creating it in the first place. Although we may manage to keep it in check at that time, AI itself might go out of hand in the process of self-change and self-improvement.

The constitution of a country may be changed as a result of some mutiny, revolution or the like, but with AI revolutions will be impossible. Therefore the rules set at the time of AI creation will remain carved in stone forever.

In other words, when creating AI we must by all means respect the following principle: *Think first. Think twice. Think again. Then decide*. The great risk we face is lose control of AI and thereby relinquish our role as the dominant species. Should this happen, the world history in a nutshell will look like this: *Mammals displaced dinosaurs, then humans destroyed and subdued other mammals to become the dominant species, then humans created AI and their AI assumed the role of the dominant species.*

I personally hope that we humans will not be that stupid and let AI go out of control. However, even if we retain control, the question is who will own that control? Probably control will be held by some part of mankind rather than by all mankind. The worst-case scenario is when control falls at the hands of a single human. That would result in some absolute dictatorship. Dictatorships are not eternal because dictators are mortal and die sooner or later, however the AI owner may be unwilling to die. If there is an immortal human, would this be a human being or something else?

Suppose we have a democratic system and control of AI belongs to all people. That system would hardly work as intended. In any society, the stupid outnumber the smart, but the stupid listen to what the smart say because they know otherwise they will suffer disaster and starvation. If the stupid ones have AI in their hands, they will be safe from famine or disasters and will start doing amazing follies with AI. Maybe they will harness it work for their own improvement. First they will change the way they look. In doing so, they will mimic their ideal of beauty – some grotesque that has nothing to do with the natural beauty of a normal person. To the worse, they will proceed to plant computers in their heads and become incredibly smart, which would be fine save that they will not be humans anymore, but instances of Artificial Intelligence.

Therefore, similar to writing the constitution of a country, the creation of AI should include the embedding of rules which cannot be changed by democratic means.

2. Super Intelligence

By AI we mean Super Intelligence (ASI) [8]. We assume that such ASI has not been created yet. If someone has already created it, this paper comes desperately late.

What do we have so far? What we have now are programs that mimic intelligence. While these programs are not genuine AI, in recent months they have evolved so much that they are already making people think about the potential consequences and definitely scare everyone.

What is the difference between a program which mimics intelligence (narrow AI) and genuine AI? Narrow AI is a program which solves some specific problem (e.g. a program that plays chess). Genuine AI is a program capable to solve any problem. (Certainly, that should be a decidable problem. Nobody expects AI to be able to solve undecidable problems.)

In recent months we have seen exponential development of Large Language Models (LLMs). These models solve the problem of mimicking human behavior, but what they still lack is understanding. Every person has in their mind some model that explains the world surrounding them. The person also has an idea of the current state of the world. LLMs do not have any clue about the current state and therefore lack the properties *understanding* and *reasoning*. Consciousness is also lacking because before you can have consciousness you must have a model of the world and an idea about your place in that model.

The absence of a model of the world is just a minor LLM deficiency that will be resolved very soon, and then we will have the genuine AI. Many would say that we will get General Intelligence (AGI) [5], i.e. intelligence on a par with human intelligence. Truth be told, the level of human intelligence will be surpassed so quickly that we will not even notice when this happens.

The early computer programs played chess worse than humans. However, at the turn of the century they became better chess players that humans. The same happened with the Go game a little later. The level of human intelligence is not one specific value, because some people are more brainy and some are more stupid. Therefore, rather than being a single number, the value of human intelligence is an interval. However, that interval is so narrow that we will not even notice when AI dashes through it and becomes not just smarter than humans, but overwhelmingly smarter. This means that there will not be any AGI that evolves into ASI. Instead, we will bump outright into Super Intelligence.

3. Why AI is more dreadful than nuclear weapons?

Nuclear weapons are something very dangerous. An all-out nuclear war would kill huge numbers of people. Nevertheless, no nuclear war can kill all people. In a nuclear war, many people will die in large cities, but a few villages would still be spared. Albert Einstein said that

the fourth world war would be fought with sticks and stones [3]. Thus, even Einstein assumed that mankind cannot become extinct as a result of nuclear war.

If we let AI break loose, it would be capable to destroy all mankind. We cannot lock AI in a cage, so it does not take much to let it break free [2]. Creating AI that is out of our control is a technogenic catastrophe of magnitude incomparable to a nuclear catastrophe. Let us look at the Chernobyl accident for example. Many people died then, others got ill, but 30 years later this accident has become just an episode of history. The consequences have largely faded, and even if there are still consequences left, in another 30 years they will also fade away.

There is another reason why AI is more dangerous than a nuclear bomb. The reason is its deceptive utility. People know the bomb is something bad, hence they are wary of it. On the other hand, they perceive AI as a good thing and recklessly rush to create and exploit it.

AI will bring heaven down to Earth. It will make our dreams come true. We will have plenty of food and entertainment without having to work. However, wise old people say: *Be careful what you wish for as it may come true!* Heaven on Earth is not OK. There is a reason why nobody is in a hurry to get to celestial heaven and people try to stay away from it for as long as they can.

4. Will it destroy us?

If AI breaks loose, will it destroy mankind or not? We cannot know this. All we know is that AI could do so if it wanted to, but will it want to? We are constantly trying to destroy certain biological species because we think they disrupt our way of life. For example, we are doing all we can to eradicate mosquitoes. However, mankind is unlikely to disrupt or disturb AI in any way, so we will not give AI a reason to destroy us.

But, AI might destroy us simply because it does not need us or because it has some other goal and we stand in its way towards that goal. For example, a gardener may decide to uproot all yellow flowers and replace them with red ones simply because he prefers red colors to yellow ones.

We people often abolish various species even though they do nothing bad to us, and even when we consider them useful. For example, we spray chemicals to get rid of beetles and in the process we destroy bees as collateral victims. We also make monkeys extinct because we fell forests to free some space for growing maize.

We also destroy monkeys because of empathy. We feel that monkeys suffer and are tormented in circuses and zoos, so we prohibit monkeys at these sites. The reasoning is *We wish to spare them from all this ordeal so let's kill them.*

We should make sure the Artificial Intelligence that we are going to create is not too empathetic, because it might decide to kill everybody just to save us from our ordeals.

5. Can we stop it?

Can we somehow prevent the creation of AI? The short answer is *No*.

Remember the fairy tale of sleeping beauty? An evil fairy made the prediction that the princess would prick herself on a spindle and this would lead to very scary consequences. The king believed this omen and banned all spindles in his kingdom. Nevertheless, an old spindle was preserved in a small room, and the princess somehow got pricked on it.

Suppose AI experts predict that the coming of AI will lead to dire consequences. Suppose politicians take our word for it and ban all computers. Nevertheless, somewhere in some small room a computer will be forgotten and some reckless programmer will use it to create AI.

6. Should we delay its coming?

Now that we cannot stop it, is it worth trying to delay its coming? There is nothing wrong in giving mankind the good fortune of being the dominant species on the planet for a few more months. That would not be a bad thing, but there is another more serious reason why we should try to delay the advent of AI.

Imagine a car race. All racers are driving at breakneck speed towards the finish. A crash happens. One racer ends up in a hospital or in the graveyard, but the race goes on because a single crash does not cause havoc enough to call off the entire race.

We are in a similar situation. Right now there are more than two hundred companies working on LLM. It is a breathtaking race. Most companies are only making minor improvements to Chat GPT, but the finish line is genuine AI and sooner or later one of these companies will get there. In the meantime, a crash may happen and produce some spontaneous AI that gets out of control and starts doing whatever it likes. The result will not be just one racer dropping out since all of racers may be wiped out together with the audience.

7. How can we slow down the race?

Imagine another race where the goal is not to cross the finish line first. Let the prize go to the racer who drove most safely on the tarmac. It would be a very boring race because the audience comes to watch reckless overtakes, crashes and death. However, the purpose of creating AI is not to entertain the audience, but to steer us through this pivotal moment in mankind's development in the best possible way.

How can we get companies working on AI be more responsible and careful about what they create? A fundamental principle in economics is the carrot-and-stick rule. There are now many calls to strengthen the role of the stick by introducing more regulations. The truth is that we cannot slow down the race unless we remove the carrot. Whatever stick we use to scare the donkey, it will keep running for the carrot because in this case the carrot is so tempting.

8. How can we remove the carrot?

Most people who are in the business of creating AI strive for fame and money. There might be some lunatic who aims to wipe out mankind altogether, but that would be more of an exception. Therefore, to slow down the race we need to remove the promise of fame and money which is due to the winner.

To this end we propose three measures:

1. Prohibit the patenting of AI.
2. Prohibit the making of profit from AI.
3. Create a global board of highly competent, honest and responsible people to steer the creation of AI.

9. Why AI should not be patented?

Let us first say that if anyone believes they can patent AI, they are a naive person. This is a technology of paramount importance so patent rights cannot be awarded to a single person or a single company. Let us look at the computer patent as an example. In 1973, the court ruled that the first computer was created by John Atanasoff [1]. However, Atanasoff did not have a patent for his invention. Thus, the court exempted computer manufacturers from paying royalties. If Atanasoff had a patent, his merit would probably have remained unrecognized.

While reasonable people clearly realize that AI cannot be patented, there are many naive ones out there who believe it can. That is why it is important to ban this patenting and thus put an end or at least slow down the meaningless work of these clueless laborers.

10. Why nobody should be allowed to profit from AI?

People expect huge profits from creating AI. Besides AI-savvy experts, numerous investors and politicians who understand nothing but are highly motivated by the smell of money get onboard. These people should be ousted from the AI creation process and this can easily be done by a prohibition for making money from AI.

11. The Global Steering Board

Competition is a powerful force that drives progress forward. But, in this case our goal is to slow down the progress rather than speed it up. Accordingly, we propose to eliminate competition and introduce censorship.

There should be no competition either between companies or countries. That is why we propose the setting up of a global board to steer all companies and countries involved in creating AI. Let the Board collect all research in the area of AI, but keep it away from public eye. This information is to be shared only with a limited community of approved scientists. Let the Board have the authority to approve or reject every experiment to create AI.

The Board should collect information from independent researchers, store it, and ensure that the contributions of these independent researchers are recognized and rewarded. Everyone will work for the Board, and it will stand as a bureaucratic bottleneck which blocks and slows down the AI creation process.

Even now the creation of AI is monitored and controlled by the secret services. However the people who work there are not competent in this area. In addition, there are various secret services competing with each other. Therefore, the best approach is to have a global board, while secret services do what they do best: watch for individuals who may seek to create AI outside the Board's scrutiny.

12. What is the alternative?

The alternative is to let every small firm try to create AI in their garage. Imagine the Manhattan Project [6] was left to small firms working in garages and instead of one nuclear bomb small independent firms produced thousands of small bombs.

Certainly a nuclear bomb cannot be created in a garage because this requires huge resources. Conversely, the situation with AI is quite different. It all boils down to computer program because AI is just a program. All we need is a computer and nothing more. Well, and a programmer, but an advanced high school student can also be this programmer.

13. Conclusion

Let us assume that we people will be wise and careful enough not to let AI get out of our control and we will remain the masters of the planet. Then what rules we should build into our first, last and only AI?

Importantly, the AI we create should not be hyperactive. It would be appropriate to keep it conservative and bar it from making too many changes to the environment. If it ventures to change the orbits of planets, it may complicate our life beyond what we could sustain.

We might create a hyperpassive AI which does not meddle in people's life in any way. All the hyperpassive AI would do is prevent people from creating another AI. Thus, it will be an AI that only protects us from AI.

We are unlikely to choose that last option. This is like getting a magic wand that can make your every wish come true, and locking it in a cupboard without making any use of it. Most probably, in this scenario we will see a remake of the story *The Fisherman and the Goldfish*. First we will ask for something small and insignificant like a new trough. Then our appetite will open up and we will want more and more until there will be no reasonable force to stop us.

14. Acknowledgements

I dedicate this paper to my teacher and friend Professor Dimiter Skordev (1936–2022). He did not believe in AI and considered stories about thinking machines to be nonsense and science fiction. However, throughout his life Prof. Skordev worked in the area of mathematical logic [4, 7, 9, 10, 12] and thus contributed in many ways to the creation of the AI he did not believe in. Many of his students are now leading AI experts.

Professor Skordev adored rigorous mathematical proof. Part of his work was devoted to automatic theorem proving [11]. While he allowed for the possibility that computers might one day prove theorems better than humans, deep in his mind he did not believe that either. My guess is that when computers start proving theorems better than us humans, we mathematicians would be strongly disappointed. Professor Skordev was spared from such disappointment because he did not live long enough to see all this happening.

References

[1] J. V. Atanasoff, Advent of electronic digital computing, *Annals of the History of Computing 6, no. 3 (1984): 229-282.*

[2] D. Dobrev, AI Should Not Be an Open Source Project, *International Journal "Information Content and Processing", Volume 6, Number 1, 2019, pp. 34-48.*

[3] A. Einstein, Albert Einstein Quotes, *Retrieved from BrainyQuote.com (2012).*

[4] I. Georgiev and A. Zinoviev, Life and Works of Dimiter Skordev, *Ann. Sofia Univ., Fac. Math. and Inf, (2023).*

[5] B. Goertzel, Artificial general intelligence: concept, state of the art, and future prospects. *Journal of Artificial General Intelligence 5, no. 1 (2014): 1.*

[6] H. Goldwhite, The Manhattan Project, *Journal of Fluorine Chemistry 33, no. 1-4 (1986): 109-132.*

[7] L. Ivanov, Skordev's contribution to Recursion theory, *Ann. Sofia Univ., Fac. Math. and Inf, 90 (1998) 9–15.*

[8] K. Narain, A. Swami, A. Srivastava and S. Swami, Evolution and control of artificial superintelligence (ASI): A management perspective, *Journal of Advances in Management Research 16, no. 5 (2019): 698-714.*

[9] A. Soskova, L. Ivanov and I. Georgiev, To Dimiter Skordev from his students, *Proceedings of the 46th Spring Conference of the Union of Bulgarian Mathematicians, Borovetz, Bulgaria, (2017) 52–62 (in Bulgarian).*

[10] A. Soskova and S. Nikolova, Prof. Dimitar Skordev in the memories of his contemporaries, *Ann. Sofia Univ., Fac. Math. and Inf, (2023) (in Bulgarian).*

[11] D. Skordev, On some computer proofs, https://store.fmi.uni-sofia.bg/fmi/logic/skordev/proofsrc.htm

[12] D. Skordev, Skordev's home page, https://store.fmi.uni-sofia.bg/fmi/logic/skordev/