Rhys Lee
December 4, 2015

Gradational Transition To Post-Artificial-Intelligence Society

Introduction

Before we start, I feel the need to point out that there are enormous amount of books and articles publicly available, of highly variable reliability, attempting to dissect the ideology and methodology behind the subject of Artificial Intelligence(AI), the authors of which usually talk only on the theoretical level and few of them have actual experience in the academic study of AI. Such fact makes any discussion related to the topic of AI require substantial background knowledge and discretion from the readers to not be nonsensical and misleading.

Unfortunately, the threshold to comprehend the state-of-art AI models in academic criteria is rather high and most AI-related fields are volatile and less well defined. Meanwhile, many expert researchers in the field often shy away from general and non-productive discussion about AI. Besides, they have few channels to intellectually and professionally share their very area-specific knowledge about AI with the general audience. This has caused many problems such as the high threshold for general public to participate and contribute in academic AI research. Thus studies and researches regarding AI are mostly confined to the professionals, Master and PHD students, less often to people outside of the academia.

I am not an academically credible author to talk about AI either. My little credibility, if any, to talk about this subject comes from twelve years of programming experience, three years of personal research regarding theoretical intelligence and neural networks, college classes in social science and politics, two years of experience in machine learning, and less than a year experience in building artificial neural network(ANN). The AI topic is easy for chitchat but extremely hard to get into its depth and essence, which is also the exact reason that I choose to share my humble knowledge about it, because for

AI to succeed as a social instrument, rather than a complete disaster to mankind, It is important and crucial to make AI be paid enough attention to and better understood by people inside and outside of the academia, preferably long before it becomes a reality(I will prove this argument later).

At last, please use your discretion and don't directly take any of my words for it(except this sentence).

Background

What is Artificial Intelligence(AI)?

Semantically, AI is any artificial implementation that can manifest the intelligence as we perceive (Barrat). Obviously, the essential obstacle to understand AI and to implement it is to understand intelligence. Noticeably, there is no widely agreeable interpretation of intelligence. Despite many tend to believe that intelligence is a definable self-standing category of certain characteristics that can be established outside of our subjective perception. This might not be true, because the humanly perceived intelligence is profoundly intertwined with our biological desires, survival instincts, and tendencies (which probably best differentiate "intelligence" from "computation"). For example, jumping off a cliff to one's death is perceived by us as unintelligent, in other words, the concept of intelligence is formed and confined in relation to the entity's performance in survival and personal gain. Also there is intelligence within the realization of subjectively objective principles of the perceived world. The Artificial General Intelligence(AGI) is defined in regard to both kinds of intelligence. An AGI is an AI that manifest intelligence indistinguishable from ours (human intelligence is the baseline for AGI). Whereas Artificial Super Intelligence (ASI) is the AI that manifest intelligence with capabilities strictly (usually by a very very long shot) superior than ours.

AI does not necessarily work in a way comparably to the biological brain. The way AI think depends on its architecture, the structure which manifests man-like intelligence. In a sense, our physical brain seems more and more readily comparable to computers, which pose the question whether our intelligence is essentially different from the seemingly different functioning of a metal box full of chips and wires? My understanding of AI is based on the assumption that it's not. After all, think about the Universal Turing

Machine(UTM), our brain is within the physical world inside which UTM can theoretically imitate interactions(computations). There is nothing special enough about our brain to make it outside of the physical world even though the structure of our brain is special enough to breed consciousness. From the other direction, average human intelligence is more than enough to imitate a Turing Machine (actions including reading, making reference, writing down a symbol, moving the pen, and repeating). Thus our brain is at least as powerful as a TM, while TM is at least as powerful as our brain— if both arguments are solid then brain is equivalently powerful as a TM. This makes our intelligence seems much less mysterious: we can study it, analyze it and one day reveal all the secrets behind it to implement models as good as it that will breed AGI.

What are the ways to implement AI

For this paper there is less need to get too technical, but a peek of the current research results in the field of AI might help with the understanding of this paper.

AI is not a really new field. Research of AI has been around for a long time and has been applied to computers since the invention of the first electrical computing device. Many believe that AI can be tackled as a software issue, and most current researches reflects such belief. Researchers have designed many math-heavy models to tackle tasks which used to require human intelligence exclusively to be solved.

The recent advance in AI mostly benefits from the breakthrough and rapid development in Artificial Neural Networks(ANN). ANN is inspired by biological brain, such as the cortex where layers of neurons interconnect and transmit signals. The ANN consists of units that are modeled after the neurons, which can be activated to communicate with other adjacent units. In some field, ANN's performance became unparalleled, surpassing many traditional hand-coded models on various tasks. For example, a very basic model of Convolutional Neural Network (2 conv layer, 2 pooling layers, 2 fully connected layer, 1 softmax-with-loss layer) trained on MNIST dataset for less than 10 minutes can recognize hand written digits (some handwritings are really terrible) with accuracy over 99.2%. There are countless more complex Neural Networks architectures designed for harder tasks and can achieve very good performance (sometimes even better than human in both speed and accuracy).

Research of AI is very rewarding as you can imagine. Except the possibility to have AI that performs better than human on certain tasks, there are even more important incentives to implement AI to replace human: Computers can perform much larger amount of tasks than humans in a shorter time. Jobs that take a human ten years to finish might take an AI less than a minute. Besides, computers don't eat or rest, nor do they quit or demand a rise when given crushingly tiring work.

The AI you want

It's not hard to imagine that if one day you create AGI agents that can think for themselves and are just intelligent as us, no matter how friendly you make their faces look, they gonna induce huge disturbance to our social and political structure, our ethics, our beliefs, our jobs, our market, and etc. Such turbulence is hardly beneficial for any person. Thus, to avoid such social disaster and to minimize the disturbance to our society, people can instead pursue a more stable version of AI agents, which are not implemented to be self-aware. Let's denote them as "narrow AI" in this paper. Such narrow AI agent is, after all, a more autonomous instrument, rather than an entity comparable to biological life forms.

Most likely, there will still be strong opposition to such invention, just like every other time we invented something revolutionary. But if we prepare for it and control it properly, such a tempered disturbance to our society are more likely to transform our society to a more productive equilibrium which will benefit most of us without destroying our society first.

To sum it up, Narrow AI is the AI model without counterpart implementation to our self-consciousness. Besides, a revolutionary AI model will hardly be complete without a system that can probe its environment without manual intervention and autonomously improve its own performance accordingly. Such implementation might take several generations' effort to accomplish.

By now, it should be clear that the meaningful and essential question we should ask is how to prepare ourselves and our society for the sooner-or-later co-evolution with AI technologies. To find an answer, I believe the most intuitive way is to reveal the potential

problems by developing a hypothetical scenario based on the most realistic assumptions we can come up with.

Hypothetical scenario

Arrival - the dawn of a new era

Birth in a private laboratory

Through major adjustments after previous failed tests, the prototype with virtually no manually engineered code finally stabilizes and indicates intelligence beyond the test thresholds (The AI we talked about in this article will always refer to the Narrow AI).

Immediately after several validation tests are passed, a couple of phone calls are made, to those who now wield the power to change the world once and for all:

Here is a little more explanation: the most well-founded and well-staffed research labs on AI are currently funded and run by private corporations, who are driven by the tremendous possible profits therein.

The human factor

Corporates game

An emergency meeting is then summoned, among investors, business managers, research leaders and some of the most brilliant and trustworthy councils from all sectors.

After days of intense close-off debate/negotiation, many decisions were made, but on some crucial questions no agreements can be reached, despite everyone was confident about the decades long preparation dedicated for this very moment.

One thing now appears certain is that nobody is really prepared. The conflict of interest between individuals is magnified by such unprecedented technology, and no one

is willing to back off from guaranteeing one's own interest. This is why we need the general population representative, likely the government, to overtake such decision making and guarantee no individual can abuse such power for oneself.

That's why for our own goods the government must intervene, and forcibly remove corporative interests out of the equation, so AI will go onto the table of political game.

Political game

The completion of the prototype was well kept secret until government intelligence gathered enough evidence and became alert enough to send a covert operation team secretly seizing the laboratory and all relevant personnels (don't bother to ask if it's legal, they can always tell you it's classified under national security).

All research record was forced to be handed to the government controlled/funded laboratory which has been pursuing the same goal secretly since decades ago. The corporations now have to resign to their fate.

Then the president was briefed before summoning an emergent meeting among president's cabinet councils, major funders, military officials, and the government top scientists/research leaders.

Although as previously said, it's good to have a general population representative to takeover the decision, however in reality, no governmental system (e.g., legislation, court and administration of the USA) is the ideal general population representative, even for "democratic" countries, which is why "democratic" country needs three branches to balance each other's power. In such reality, it's possible that the government will conduct behaviors even worse than the corporates could ever do. So we need to find out who will be the real players in this game, and how to refrain them from abusing the technology.

Real players emerged

The AI is still kept secret for a while, and during this period of time, the major funders of the office, which are also some big corporations, start to redistribute and deploy their capital before the public announcement. The president issued several presidential decrees, bestowing the AI ethic committee (AIEC) funded decades ago

temporary power to supervise the research result, to design a new government agency specific for AI technology supervision, to draft regulations, and to draft initiatives for the congress to pass.

Although trying to downplay the event, the president can't avoid to eventually make a public announcement about the research result, and at the same time the AIEC propose their initiative to the congress to pass. The initiatives passed very fast and smoothly in both houses, and became laws in no time. Here I just have simplified how AIEC could keep their integrity and how congress could green light them so easily, because that's totally another essay, and that's politics.

The laws classify the research and its results as top secrets, forbid civilian access, forbid unauthorized distribution, modification, and deletion.

The law also requires all records and copies of the prototype to be kept under a certain small number and each of them under constant supervision. Besides, any future research must be conducted in a supervised, classified federal laboratory, where all the key researchers in the original research are transferred to.

I will explain why such laws are crucial in later analysis.

After some time, the government authorizes some organizations (e.g. military weapon contractors, national academic laboratory, industrial equipment manufactories) partial access to utilize some limited/encrypted versions of the prototype,(so that the only entity who can study, recreate and modify the AI is still the government supervised laboratory)

The weapon makers first train the AI agent knowledges regarding physics, chemistry, electronics, explosives, biology, aerodynamics and etc. Then they utilize AI's intelligence to improve their current manually designed weapons to maximize their power, to design completely new weapons, and to integrate AI programs into some most advanced weapons(e.g. killer robots). But fortunately such actions are limited and supervised by the AIEC.

The Academia teach the AI agent knowledges from the basics of every discipline to some cutting-edge research results. Then they utilize the AI's intelligence to search for contradictions/errors in modern science, to interrelate different sciences to find patterns and knowledge that were overlooked, to analyze all the existing data to discover new knowledge, and to design and conduct new experiments (e.g., OED) to explore the truth behind the observation of the Universe.

The industrial equipment manufactories utilize the AI to develop the next generation manufactory equipment (e.g. industrial 3D printer, nano robot constructor, and etc). And then provide these revolutionary production equipment to other factories throughout all industries.

One thing should be noted here, the law must forbid any manufactory using the AI to develop advanced/complicated computer hardware with architecture too different from the old ones for human to comprehend(we already had tremendous difficulties understanding the old ones, almost like they are not created by human), because future programs, most of which are potentially developed by AI agents, might be run on such hardware. If these programs get out of control on hardwares less understood by anyone, by that time, human would be too intellectually incapable to discover the problem, let alone stopping it.

Now the entities that have control/access to the AI, besides the administration, include: members of the AIEC, leaders of the federal research lab, the head of some private corporations close enough to the administration to gain clearance, and some academic scholars.

The government could almost always benefit from downplaying AI related issues, since these issues could be extremely controversial and troublesome to most voters, whereas the technology brings enormous political capital. They would spare no effort to advocate how safe and under control the AI is, keep many of their actions in secrecy, and whatever necessary for the "greater good". Thus, the administration with its friends in private sectors would get more and more powerful, less and less transparent…

Now we see that the administration is bounded to get out of control in this scenario, we need to fix this before it happens. One way to do this is to induce transparency early on. This can be achieved with a legislation by the congress, which includes laws demanding all supervisory channels, up-to-date detailed status of each AI agent be permanently and constantly available to the public, thus nullifying any covert attempt to use AI agent for personal or political benefits. Don't neglect the power of a written law, for your reference, Russia could have become an actually democratic country if only its first constitution weren't written in a way to bully the parliament.

Yes, such legislation might create inconvenience for certain geopolitical maneuvers; this might also limit the areas that AI can be applied to (you don't want to broadcast how military contractors make their new weapons to everyone on the internet). But between all these inconveniences and having a dystopia overpowered by a tyrant administration with angry AI robots killers running the street, the choice is so easy to make, right?

After sufficient legislation is deployed to guarantee the transparency of the technology's administration and to enforce heavy punishments to those who attempt to abuse it, the technology's largest disturbance to the society will gradually tend to stabilize.

The less mentioned key point here is to not make many copies of the AI agent, as the more they are, the more likely some of them will get out of control. Some may worry that this will confine the application of the technology, but this worry is largely dismissible: for most situations, such as domestic robots or assembly line robots, to accomplish these jobs doesn't actually require the AI agent at all. Instead, the AI agent can write very intelligent programs for these robots or machines to handle more situations than they would ever encounter. Thus keeping the AI copies in a small number is a very necessary security precaution without much noticeable impact on the application of the technology.

As AI's impact on the political system now tends to stabilize, let's shift our focus to their impacts on other parts of the society.

Sustain and grow

What does Narrow AI bring us

AI brings us an extremely low cost (compare to their human counterparts) but superior intellectual workforce that works ceaselessly at peak sufficiency with continuously improving performance, and is readily integrated with the entirety of digital information we have for it to explore the world beyond our horizon.

Without the implementation of self consciousness, it is an unprecedented instrument that gives enormous advantage to those who wield it, in strategy making, data analysis, system control, analytical prediction and etc.

On individual level, the AI can free people from most well-defined repetitious jobs, thus people can be free to put their productivity into innovations, philosophy, arts, entertainment industry, education, politics, sports and etc.

The redistribution of employment will be inevitable, however the we can ensure gradual transition by banning any commercial application of AI in the first few years,

then unbanning AI in different fields step by step after well informing the population of the agenda to allow them enough time to prepare.

Just like the time when we invented steam engines, the electricity, and the internet, some jobs will be obsolete and some new jobs will be created; some people will leave their original jobs, some people will capture the new opportunity to realize their ambitions; some people will get richer, some people will get poorer… Regrettably, the government is not responsible for every individual's life quality. Survival of the fittest always applies to any society. The loss or reluctance of some should not be the excuse to halt the marching of the whole mankind towards enlightenment.

Those who worry the economy will collapse due to AI is ludicrous. The AI is a tremendous boost to the productivity and the economy. Bubbles may be created like the internet era, but on the higher level the economy will be much stronger in terms of growth. Logically, the commodity price will adjust to the lowering costs, the currency will hold more value, things will cost less, people will be wealthier, however in reality that will be true only if the government forget to turn on the money printer. The current financial system is effectively exploiting the general population but that's not a problem simply fixable by advanced technology.

We can't expect AI to solve all of our problems, but they might. A market supervised and adjusted by an AI program mastered in economics and financial policies will most certainly do a much neater job than the current AFR. The legal system can be ultimately just, that is no one could be favored or empathized by an emotionless machine mastered in front of written laws and constitution. Such program will also conveniently reveal most of the potential contradictions and loop holes in the laws for us to fix before someone exploit the legal system…

Thus, as long as AI is gradually induced to areas of our society to reduce the stress of change, though some unpleasant things always happen, eventually there is an overwhelming chance that our society will be much better and more perfect than ever before.

However, everything comes with a price. With all the benefits AI could benefits us, there will be problems generated thereby. The first problem is dependency.

AI is, by definition, capable to solve all the problems that humans used to have to maintain a basic life. In our generation, there are already quite a population live on their families or other people. In the future, when AI with robotics can make the cost of most merchandise and daily necessities marginal, people would have less and less problem to

maintain a life quality as we now live. However, the decreasing urgency to work for one's life will inevitably discourage people from hard working and pursuing hard tasks.

Such decrease in productivity in turn will contribute to more dependency on technologies, which becomes a self-reinforcing loop results in greater and greater dependency on such technologies.

Another significant problem will be inequality. Instead of balancing the scale, AI might catastrophically adds to the asymmetrical power distribution existed in the society.

So, how will AI increase inequality?

Suppose an imaginary scenario where everyone has the equal access to copies of the exact same AI powered programs, then even everyone has the exactly equivalent intellectual capability and nobody can use these programs slightly better than anyone else(which sounds impossible), AI still gives unbalanceable advantage to whoever gives the command first in any competitive setting, because all copies of AI are exactly the same and provably in no way could a later start gives better perform the same system that started working earlier.

Now remove the unrealistic assumption that everyone will have equal access to AI. Since the AI is by definition superior in intellectual capabilities than that of human's, thus most competitions between human and AI will deterministically ends with AI's victory. That means those who don't have access to AI (who can't use AI to assist them in a competition) would bear enormous disadvantage to compete in any situation where AI can participate and contribute, which is almost every situation since AGI, by definition, is the AI that manifest intelligence indistinguishable from human's.

Now to make it more realistic, there will be AI with different capabilities, since AI's intellectual capabilities are in every way much more superior than human's, thus in most competitions between two AI the difference between their human operators are negligible, and in most cases a little edge over another AI will be enough to guarantee that the winner always be the better AI. Thus in a more realistic world, the inequality derived from the AI will be worse.

Besides, for system that can self-improve, an early start would, logically, give one system non-reversible performance advantage over an exactly same system that started late.

So giving everyone access the Narrow AI agent is likely to be a very terrible idea.

This is why we need laws to prohibit general populations from accessing the AI agent, and also the reasons why one of the most important precautions is to guarantee a transparent administration where no one can use AI for their personal gain.

There are other AI-derived problems such as the ambiguity to determine the responsibility of an AI-related accidents, and how to pursue the responsibility. Most of these problems already exist currently with some other complex instruments we invented. Exploring solutions to those problems is an ongoing process, thus there is less need to talk about them specifically in this paper.

Conclusion

In summary, The further we look into the future, the more uncertainty it involves, the less reliable our prediction will be, and the less value therein for this discussion. We have revealed many problems that may cause tremendous damage to our society, and we have gained some insights about some basic guide lines.

Here are some most important precautions we need to prevent disastrous consequences:

1. If we allow corporations to develop AI and use AI as they wish for their profits, then corporations will not take the benefits of the general public into their consideration, and cause huge disturbance to the society without any precautions taken for the public population. Therefore we need general public representative to take over any AI implementation.
2. If the AI administration is not transparent, then it defeats the purpose of having the government to take over AI agents from any private companies. Therefore we need to design and pass sufficient legislation to guarantee the transparency of the AI administration.
3. If we allow public access to the AI agent, then no effective supervision could be enforced to keep the AI under control. Thus we need to pass laws to prohibit general public to access the AI agent, and we need to keep the AI agent in small number of copies each under heavy supervision.
4. We need to minimize the disturbance of AI technology to the market and employment by prohibiting commercial application in the first few years and

gradually unbanning it so that we allow people and the market to adapt to the revolution not destroyed by it.

Guidelines:

5. Do not implement consciousness for the AI, otherwise it will cause too much disturbance to the society and might get out of control.
6. Restricting and regulating access to the technology actually benefits everyone.
7. Legislative guidance is of crucial importance to make sure AI technology will be used under control and benefits the general population.
8. Establish Artificial Intelligence Ethics Committee (AIEC) to allow quick and professional response in emergent AI-related situations.

The real problem is still: ourselves.

The biggest unsolvable problem with AI technology or other technologies such as nuclear power is always how to prevent people from misusing them.

Either we are going to heaven or we are marching to hell, technologies won't alter our direction much, but it might help us to get there faster. We cannot expect technologies to solve our problems, the most problems we have right now are within the society, within some of us or all of us. AI can't help us to save us from ourselves, neither could we really blame AI for not solving all of our problems. Such reality illustrates the importance of ethics.

Refernces

Barrat, James. *Our Final Invention: Artificial Intelligence and the end of human era*. Thomoas Dunes Books, New York,USA, 2013. Print.

Winston,H. Patrick. Artificial Intelligence, third edition. Addison-Wesley, USA, 1993. Print.

Dye, R.Thomas. *Power and sosciety: An introduction to the social science, second edition*. Florida State University, Massachusetts,USA, 1979. Print.