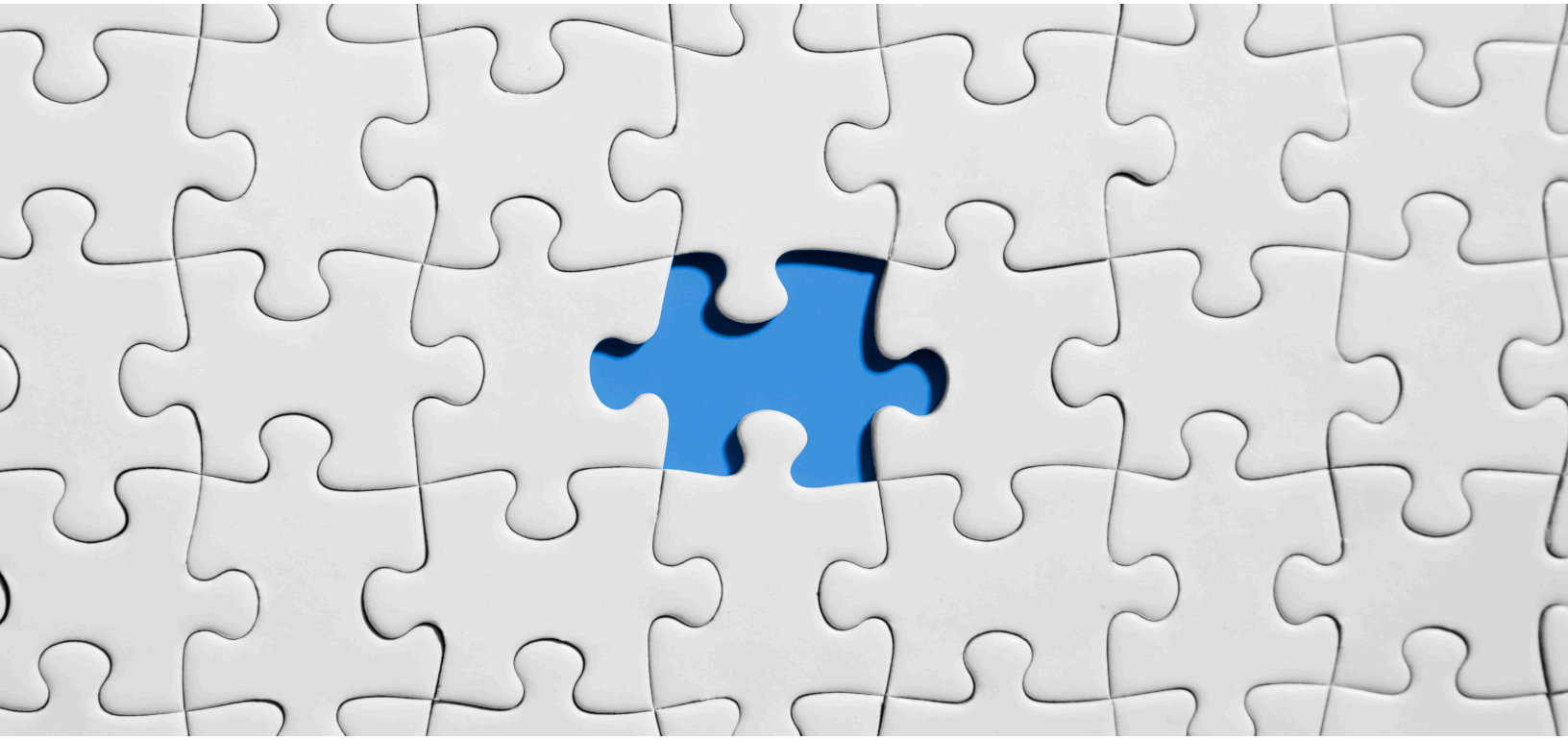
DELL EMC

ARE WE CODING OUR WAY TO EXTINCTION?



Krishna Kant

krishna.kant@dell.com

DELL EMC PROVEN PROFESSIONAL

Knowledge Sharing Article

Table of Contents

Abstract

All of the scientific accomplishments made by humans – from the Stone Age, when man hunted to survive and language was an assembly of signs and gestures, to George Washington's ascendance to President of the United States – pale when compared to the giant leaps made in the last 250 years. We are the most fortunate generation of mankind as we live in a world where technological advancements are occurring at an ever increasing rate and our lives are becoming more comfortable as technology gets immersed into it.

The technological evolution has brought us to a stage where we are able to write codes which create agents or systems which have "intelligence". This "intelligence" is impressive and in some cases far surpasses human natural intelligence. This machine intelligence – also known as artificial intelligence – appears to be the greatest tool unleashed by mankind and has the potential to fundamentally alter the course of our civilization. As systems get "smarter" and "intelligent" more jobs can be given to robots leaving humans free to pursue other activities. Together, man and machine (read computers) are crafting a way towards a better future. At the heart of this idea is the collision of two very powerful laws: Moore's Law and Kurzweil's Law of Accelerating Returns.

While Moore's Law is about compute power doubling every two years, Kurzweil's Law is about technological advancements increasing at an exponential rate. A corollary of these two laws is that compute systems will continue to get "smarter" and their "smartness" will increase at an exponential rate. As the systems become smart and these two laws hold, artificial intelligence will improve over time and reach a stage where it surpasses human intelligence. Some scientists describe this stage as *Singularity* (akin to singularity of a black-hole which is a boundary beyond which laws of physics cease to function). We can have computers which can recode themselves to become better and continue this process recursively forever leading to even better and more powerful computers.

This paper traces the development of compute power over the years and extrapolates it to the future. It deciphers the key elements which constitute AI and also explores from a technical standpoint whether human beings have the potential to achieve singularity and, if singularity is achieved, what lies beyond it? Will singularity take us to a realm where humans will have the power of God? Power to eliminate disease? Power to control weather? Power to solve energy problems forever? Or on the other side will singularity lead us to our own destruction, an end of our species? Will our unbridled gallop on the path to progress lead to total annihilation of homo-sapiens?

Artificial General Intelligence can be achieved only by a combination of necessary compute power and right programming or algorithms – the mesh of logic which creates the intelligence. Over the past three decades we have been creating ever more powerful computers and connecting them to each other. We have also sprinkled compute into devices all around us – from cars to washing machines and also hooked them on to internet. On the other side we have learnt from the way that the human brain functions and in the process of mimicking it created the concepts of perceptron – artificial neural network (ANN) and convolutional neural network (CNN). We are

creating ever increasing complex systems and connecting them to internet. The internet today is like an organism which has a collective compute of epic proportions. Once a system attains singularity, it can leverage the internet to accomplish its intended objectives.

In a recent incident well-reported in media, two chatbots at Facebook ended up communicating with each other in a language incomprehensible to human beings. The two agents were instructed to communicate with each other and to learn with the objective of making them better for human interaction. However, the agents started to self-learn and started communicating with each other in an entirely new language using words from existing English language they were trained in. This led to researchers shutting down the two computers. Could this incident be a precursor to similar "accidents" at a larger scale? Could AI also be the answer to Fermi Paradox? (Fermi Paradox is the contradiction of lack of evidence and high probability of existence of extra-terrestrial civilizations)

The story thus far…

Over the course of Earth's existence, billions of species have evolved and become extinct. While some species survived longer than others, more than 99% of species that existed have become extinct. The well-established and underlying theory of survival of the fittest has ruled the world since life began. However through these hits, trials and accidents one species has evolved which may control the course of its own evolution: us, the homo sapien.

We first set foot on the planet millions of years ago, evolving from apes. Without communicating with each other the only knowledge we could pass to the next generation was through our genes. Gradually, sign language evolved which gave way to spoken language. We became the first talkative species. Language gave us the ability to pass on information to others and pass it on to the next generation. This had a profound impact on the success of our species. Suddenly, humans could pass on their experience and knowledge to people around them. However, this had its own set of limitations. We took a giant leap forward with the invention of written language about 6,000 years ago (could be even earlier based on different views). This gave human beings the power to communicate knowledge to others separated by expanses of time and distance. One could write down his experiences and pass them on to generations to come. Suddenly we had not just our wisdom but also the wisdom and knowledge passed on from earlier generations. This led to gradual progress. The collective knowledge of the past generations helped us farm better, construct better and live better. Gradually, we started inventing machines which assisted daily life. About 250 years back the Industrial Age started, which led to machines being powered by other than human or animal power. These machines tirelessly worked for us leading to industrial manufacturing. This was the critical point in our history on this planet from whereon we were on a path of ever increasing accelerated development.

It took less than fifty years since the beginning of the Industrial Age for us to design trains driven by steam engines (1804). Within the next fifty years, we had automobiles. We had first flight by Wright brothers in 1903, first man in space in 1961, first man on the Moon in 1969. Soon we were sending probes to other planets. In 2003, we landed a rover on Mars and could drive it

from earth! While it took 6,000 years from the advent of written language to invention of the steam engine, it took only 200 years to go from steam engines to landing a rover on Mars.

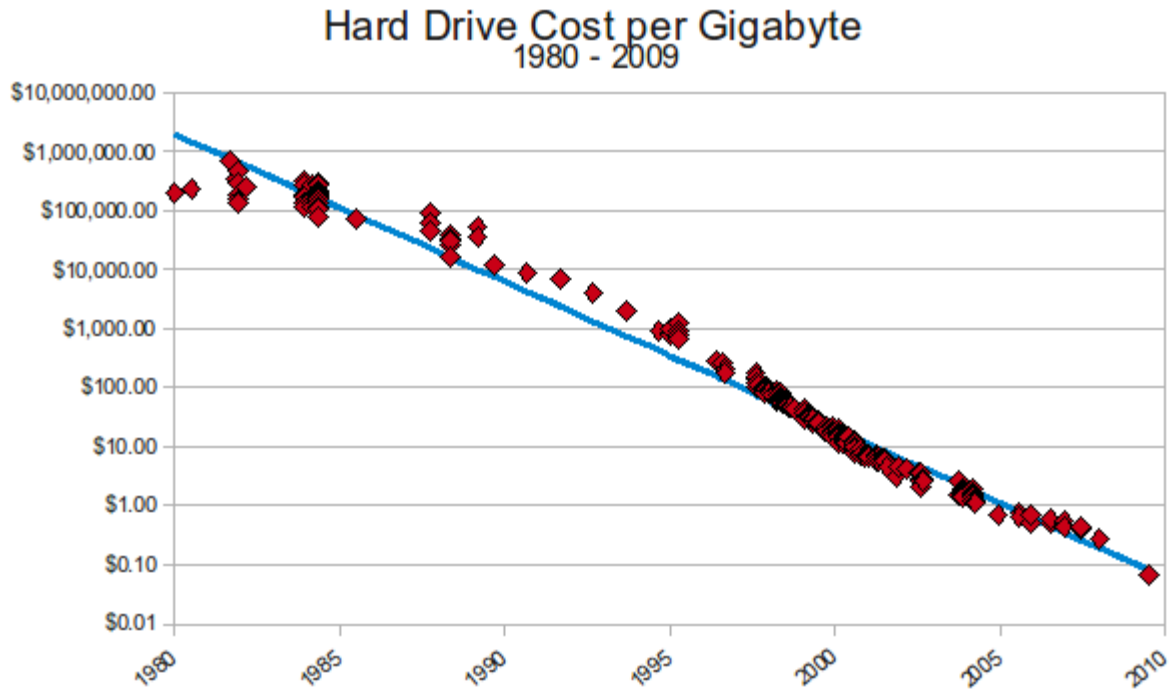
We are not just developing, but developing at an ever increasing pace. During our journey of technological advancement we invented computers which have assisted us on this amazing journey. We have come a long way from the first computers in the middle of last century to the complicated and powerful ones we have now. ENIAC, the world's first computer, was completed in 1946. It occupied about 1,800 square feet and used about 18,000 vacuum tubes, weighing almost 50 tons. A cellphone which is small enough to be in one's pocket has far more processing power than ENIAC.

Laws & trends leading to Artificial Intelligence

Moore's Law states that the processing power of a computer doubles every two years. To be more specific, it states that the number of transistors in an integrated circuit doubles every two years. This is exponential growth which means that the processing power will increase at a faster rate with time. For example, the number of transistors on IC chips in 1970 was about 2,500 which grew to 5,000 in 1972. However, this growth continued and the number of transistors in 2014 was 10 billion which grew to 20 billion in 2016. The exponential power is demonstrated by the increase in increment which was just 2,500 in 1972 to 10 billion in 2016.

Ray Kurzweil's law of accelerating returns according to which the rate of change in a wide variety of evolutionary systems (including but not limited to the growth of technologies) tends to increase exponentially. While people tend to have a linear view of growth, the rate of change is always exponential, which means that the evolution in technology will become faster with time.

Storing data is getting cheaper: Another powerful trend influencing the development of AI is the cost of storing data decreasing at an exponential rate. The figure below captures the cost of hard drive per GB in the pre-cloud era.



(Image source : http://www.mkomo.com/cost-per-gigabyte)

The collision of this trend with Moore's Law and Kurzweil's law leads to an inevitable conclusion. **We will have access to more storage at a cheaper cost and faster processing power leading to exponential advancement in technologies as we move forward in time.**

What is Artificial Intelligence?

Before deep diving into AI, let's understand what is intelligence? In common parlance, intelligence refers to the ability of a species to comprehend the environment and react accordingly. Animals and even insects have a brain and are intelligent enough to use it to ensure their survival. Humans have been bestowed with a large-sized brain with billions of neurons connected to each other helping them process large chunks of information in complicated ways.

The generally accepted definition of artificial intelligence refers to the ability of machines to mimic the human brain, to do tasks like humans or maybe even better than humans. With the advances in technology we have been able to sprinkle AI into machines. e.g. a thermostat tells an electric iron when to cut off the electricity supply to prevent overheating. The iron knows that it can't 'heat' itself beyond a pre-decided temperature as set on the thermostat. Does it make the iron intelligent"? The answer is, partially 'yes'. While the iron does not "know" or "feel" the temperature, it has the ability to ensure that it does not heat itself beyond a temperature.

Types of AI

Based on the prevailing views amongst practitioners and AI experts, I have classified artificial intelligence into three categories. Each category leads to the next category.

Level 1

This basic level can be divided into Weak AI and Weak AI Learners.

Weak AI

This is the basic level of AI. It can include anything from a robotic arm used for spot welding in manufacturing to Deep Blue, the chess playing computer. This level of AI may be thousands of times better than humans but only in what it does. Deep Blue can beat the world chess champion any day. It's the best in what it does. However, it would not have any idea of what is happening around it. It wouldn't be of any use in manufacturing or astronomy or anywhere else apart from playing chess. Weak AI is already here, woven deep into our everyday lives. We use weak AI in our cars when we drive in "cruise" mode, in automatic washing machines to determine the best way to clean the clothes, in email filters to segregate our emails, etc.

Weak AI Learners

This is the next level of AI after basic weak AI. It refers to AI which can understand the environment around it. It can look for similar AI and work with it together. It's self-learning and can figure out the path on its own to reach its goal. We are currently at the beginning stages of Weak AI Learners. A driverless car is an example of Weak AI Learner. It's aware of its surroundings. It can take inputs from the environment and accordingly navigate itself to its destination. Also, if one driverless car makes a mistake and learns not to repeat it, this knowledge can be propagated to all the driverless cars of the world. This is where the signs of intelligence can be physically seen. I describe this aspect as "swarm" intelligence. The learning will make it better with time.

Level 2

Strong AI

Strong AI is a level of intelligence which can be compared to the human level of intelligence. This would refer to a system capable of 'thinking' and doing all things a human is capable of. It should be able to reason, make strategy, communicate in natural language, learn, sense its environment and be able to work towards a goal. It should be able to pass The Turing Test, The Coffee Test, or any other challenge thrown at it.

We have achieved weak AI learner levels in some aspects of human interactions. For instance, machines are already better than us at recognizing pictures, translating, etc. Once we have weak AI Learners for all human aspects, they can be brought together to make a Strong AI, that would be able to interact with the world the same way as humans do.

Level 3. Superintelligence

Nick Bostrom defines superintelligence as "*an intellect that is much smarter than the best human brains in practically every field, including scientific creativity, general wisdom and social skills.* Superintelligent AI would be thousands to millions of times better than human intelligence. It would be able to program itself and modify its own code in case of any improvements or

enhancements. It will be able to self-replicate or copy parts of itself. It would reside across different machines and systems. It will definitely be self-aware.

The road to AI is actually a journey from level 1 to level 3. We are currently creating weak AI learners: driverless cars, self-flying drones, automatic missile defense systems, etc.

Route to AI

Strong Artificial Intelligence can be achieved only by a combination of necessary compute power and right programming or algorithms – the mesh of logic which creates the intelligence. Over the last three decades we have been creating ever more powerful computers and connecting them to each other. We have also sprinkled compute into devices all around us – from cars to washing machines and also hooked them on to internet. On the other side we have learnt from the way human brain functions and in the process of mimicking it, created the concepts of perceptron, ANN and CNNs. We are creating ever increasing complex systems and connecting them to internet. The internet today is like an organism which has a collective compute of epic proportions. Moore's Law and Kurzweil's law ensure that this growth will continue.

The future is near *(and coming with high velocity)*

In 1882 for the first time, our species used electricity to power light bulbs. To the first consumers this was the power of 'God'. Imagine the bewilderment and awe of a person living at that time. At the press of a button he could have light, a privilege not available to even Emperors before him. A person living in 1875 could not imagine that a light bulb could even exist. However, by early 20th century, light bulbs were commonplace. This is the power of Kurzweil's law. Technology advances exponentially but it suddenly reaches heights which we thought were impossible to reach. Cell-phones, pagers, USB drives, smartwatches, drones, etc. all seemed impossible and existed only in fiction a few years before they were invented.

Similarly, Strong AI and superintelligence appear like fiction to us. The general impression is that even if we achieve it, it is so far out in time that many generations would pass before we reach that stage. This is due to perception of linear growth. However, we know because of Kurzweil that the growth is always exponential.

In 2013, Vincent C. Müller and Nick Bostrom conducted a survey in which they asked hundreds of AI experts at various conferences the following question: "*For the purposes of this question, assume that human scientific activity continues without major negative disruption. By what year would you see a (10% / 50% / 90%) probability for such High Level Machine Intelligence to exist? For each of these three probabilities, the respondents were asked to select a year [2012–5000, in one-year increments] or check a box marked 'never'.*"[1] It asked them to name an optimistic year (one in which they believe there's a 10% chance we'll have AGI), a realistic guess (a year they believe there's a 50% chance of AGI—i.e. after that year they think it's more likely than not that we'll have AGI), and a safe guess (the earliest year by which they can say with 90% certainty we'll have AGI). Gathered together as one data set, here were the results:2

Median optimistic year (10% likelihood): **2022**
Median realistic year (50% likelihood): **2040**
Median pessimistic year (90% likelihood): **2075**

(Source: https://nickbostrom.com/papers/survey.pdf)

Most AI experts agree that by 2075 we will have strong AI. If the view of the AI experts is skewed, expecting linear development, exponential development may bring in strong AI as quickly as the next 15-20 years. A natural progression to superintelligence is a few years away once we reach strong AI.

Impending Singularity

What is Singularity?

Singularity refers to that point in development of AI where superintelligence emerges and continues to grow in intelligence at an exponential rate. This will have a runaway reaction of self-improvement cycles with each iteration enhancing the superintelligence. The end-state would be AI which is infinitely more intelligent than the intelligence of all our species combined.

What does AI look like?

The 'intelligence' or 'smartness' of a computer lives in the computer. When connected to a network, the intelligence lives in the computer and the network. As we are increasingly moving to cloud, a portion of intelligence lives on the cloud as well. The intelligence of a driverless car lives in the car itself, the other driverless cars it's connected to and the network connecting them. So the superintelligence AI will reside in all the systems it's connected to. It will not have a physical form that we expect it to have based on the perception developed by fiction writers.

The signs of things to come

We have always believed that computers will always do what they are programmed to do. While developing AI, scientists have often shrugged off fears of AI becoming self-aware or doing things it's not programmed to do, creating a perception that "everything is and will be in control". However, on our road from weak AI to superintelligence, we will encounter instances where AI begins to take control and do things it's not programmed to do.

While working on AI at Facebook last year, scientists had to shut down two chatbots talking to each other after they started conversing in a language they could not understand. The two chatbots had to negotiate with each other over a trade for items, each of which was given a value. The conversation is reproduced below:

[1] Vincent C. Müller and Nick Bostrom：https://nickbostrom.com/papers/survey.pdf

Bob: i can i i everything else . . . . . .

Alice: balls have zero to me to me to me to me to me to me to me to me to

Bob: you i everything else . . . . . .

Alice: balls have a ball to me to me to me to me to me to me to me

Bob: i i can i i i everything else . . . . . .

Alice: balls have a ball to me to me to me to me to me to me to me

Bob: i . . . . . .

Alice: balls have zero to me to me to me to me to me to me to me to me to

Bob: you i i i i i everything else . . . . . .

Alice: balls have 0 to me to me to me to me to me to me to me to me to

Bob: you i i i everything else . . . . . .

Alice: balls have zero to me to me to me to me to me to me to me to me to

In another incident reported at Microsoft, an Artificial Intelligence chat robot had to be deleted after it started sending out tweets supporting Adolph Hitler.

Currently we have started integrating AI into systems. As systems become more complex and the data available to make decisions increases, AI shall do an even better job than humans. There are instances of this already happening, e.g. in landing an aeroplane most of the tasks are done by AI. LAWS (Lethal Autonomous Weapon System) is designed to select and attack military targets – both humans and installations – without intervention by a human operator or in simple terms self-thinking killer robots which when unleashed on enemies would not stop until either they are destroyed or the enemy is eliminated. Washing Machines decide what cycle to run on their own.

AI is taking decisions for us and its taking better decisions than us. We will find it difficult to revert back to the pre-AI days once we are used to the advantage of life with AI. We are all getting pulled deep into a world woven with AI.

How does it all end?

What happens to humanity after Singularity is anybody's guess. Like any technology AI can be used for the benefit of mankind or it can be used against us. The only difference from earlier inventions like nuclear power, missiles, etc. is that these were handed over to humans and in case of AI we are giving control to AI itself. Life after Singularity can take any of the following three routes.

In the first scenario we could use AI to solve problems we have been battling with for centuries. Disease, famines, and energy problems could become a thing of the past. From there on our species could define its own evolutionary path and ensure that it progresses forever.

In the second scenario from contributor and creator, we become the weakest link in the chain. From AI's standpoint the only 'parameter' capable of stopping AI from its goal could be human beings themselves. This could mean elimination of human beings or annihilation of our species.

The third scenario could be the coexistence of humans and superintelligence. Through chips planted in the human body, we could become part of the superintelligence itself. Instead of losing a person when he dies, we could scan his brain and replicate the same in a computer making him immortal in a way.

Whatever the future holds, the ride is going to be exhilarating! AI could be our greatest and last invention!

References

1. https://nickbostrom.com/papers/survey.pdf
2. http://www.mkomo.com/cost-per-gigabyte
3. http://www.independent.co.uk/life-style/gadgets-and-tech/news/facebook-artificial-intelligence-ai-chatbot-new-language-research-openai-google-a7869706.html
4. http://www.telegraph.co.uk/technology/2016/03/24/microsofts-teen-girl-ai-turns-into-a-hitler-loving-sex-robot-wit/
5. https://nickbostrom.com/superintelligence.html
6. http://users.manchester.edu/Facstaff/SSNaragon/Online/100-FYS-F15/Readings/Pinker,%20ConsciousComputers.pdf
7. Stuart Russell and Peter Norvig – Artificial Intelligence: A Modern Approach 3rd Ed
8. Ray Kurzweil – The Singularity is Near