David Chalmers and the Singularity

Accompanying notes

‘The ever-accelerating process of technology gives the appearance of approaching some essential singularity in the history of the race beyond which human affairs, as we know them, could not continue’

John Von Neumann 1958

Technology

In the space of around one hundred years, successive waves took humanity from an era of candles and horse cards to one of power stations and space stations. Something similar is going to occur in the next 30 years, a new wave of technology will force us to confront the most foundational questions our species has ever faced.

Technology proliferates in waves of unpredicted causality. Most technologies span out evolve and dominate through the ages. Scientists and inventors cannot always predict the way that their technology is used or ways in which it could cause potential damage. In addition, they cannot predict how successful or how quickly their technology will spread. There have been countless people and groups who have tried to stop technologies from proliferating, most failed. Science and technology are based on ideas, and ideas do not die. Technology comes in successive waves, starting from the early technologies like the wheel and fire to the most recent waves of quantum computing, genetic engineering and now AI.

Ray Kurzweil, a well-known futurist, and author of The Singularity is Near (2005) has popularised the concept of the technological singularity. He believes that artificial intelligence will eventually surpass human intelligence, leading to a point where technological growth becomes uncontrollable and irreversible. He predicts that this event, known as the AI singularity, could bring about profound changes to society in the human condition. He predicts 20,000 years of technological advancements into a single century. His prediction, although much debated, is that the singularity will occur around the year 2045.

Research on the AI singularity explores the hypothetical future scenario where artificial intelligence surpasses human intelligence leading to rapid technological growth and potentially profound societal changes. Recent studies focus on the implications of AI advancements including concerns about AI safety and ethics. Governance researchers are investigating ways to ensure AI systems align with human values, mitigate risks of unintended consequences, and promote beneficial AI development.

Additionally, discussions revolve around the potential impact of super intelligent AI on various sectors such as healthcare, finance, and transportation. Ongoing research aims to address the challenges and opportunities associated with the AI singularity.

Mustafa Suleyman (the founder of Deep Mind and author of The Coming Wave) says he never forgets the moment AI became a reality for him. With his team he created an algorithm called DQN (Deep Q-Network) that was trained to learn I how to play games by itself. *The self-learning was the key*

distinction of the system. DQN taught itself unexpected and brilliant strategies to winning games and was the precursor for AlphaGo. AlphaGo initially learned by watching 150,000 games played by human experts playing the East Asian game of Go. Then, it played against itself over and over. This algorithm was able to simulate millions of new games, trying out combinations of moves that had never been played before. In March 2016 AlphaGo won against Lee Sedol, a virtuoso world champion. This brough AI to the forefront of the world's attention. In a few months, the team had trained it with algorithms to discover new knowledge and find new, seemingly superhuman insights, to win the game. Later models like Alpha Zero dispensed with any prior human knowledge and simply played itself millions of times, learning from scratch to reach a level of performance that trounced the original AlphaGo without any of the received wisdom or input of human players.

Mustafa Suleyman goes on to discuss the rapid emergence of powerful Large Language Models (LLMs) like ChatGPT. He also explores the exponential progress in synthetic biology, catalysed by innovations such as CRISPR gene editing, leading to groundbreaking applications in medicine, agriculture, and manufacturing. Suleyman explains how the interconnectedness of AI and synthetic biology will bring about a future where these technologies converge to drive unprecedented advancements. Beyond AI and biology, he explores the broader spectrum of the impending wave, encompassing robotics, quantum computing, and clean energy, each poised to reshape society in profound ways. Suleyman underscores the accelerating pace of technological convergence and the potential for unforeseen breakthroughs as these disparate fields intersect, heralding a new era of innovation and transformation.

AI for you

A clever way to think of AI is to imagine Einstein is in your basement and you have instant access to all human knowledge. However, he can make mistakes, jump to conclusions, and misunderstand you. Einstein's biggest limitation is you. In order to use AI effectively there are two main questions to ask yourself 'What can I do? How do I do it?' These two questions are the foundation of prompt engineering skills.

Chat GPT stands for Chat Generative Pretrained Transformer. It is a large language model (LLM)that is built from an artificial neural network where a bunch of numbers and parameters are connected to each other. A large language model will have millions or trillions of parameters and these parameters are set through training. The large neural network is given lots and lots of text and corrected through back propagation. The second stage of training is reinforcement learning with human feedback where humans will spend thousands of hours testing and evaluating the output from the model and giving it feedback. For example, if you asked Chat GPT 'How do I rob a bank?' it would answer 'I can't assist with that.' This is why large language models are tested and evaluated, and then frozen (pretrained.)

There are now multimodal models that are so good they can role play, write poetry, code, strategize, coach, teach and provide legal and medical advice. These models can be absolutely brilliant and sometimes terribly stupid. Human judgement is still needed to compensate for AI weaknesses to double check its work for legal compliance, for data security, bias etc. Prompt engineering and design is at the heart of using AI. Better prompt engineering skills equals better results.

At this stage of AI developments, we still exert control of the domain, the parameters, the legal, the cultural the security of the models and the AI we programme.

Stages of AI

According to future business tech there are ten stages of AI, four have been reached:

Stage 1 Rule based AI e.g. thermostats, microwaves

Stage 2 Context based I. e.g. Siri, Google assistant, Alexa, they analyse vast amounts of data and can recall your browsing history they are made to cure a cure rate and experience for the individual new line

Stage 3 Narrow domain AI e.g. IBM's Watson, AlphaGo

Stage 4 Reasoning AI e.g. ChatGPT, autonomous vehicles

Stage 5 Artificial general intelligence AI can perform any software task that a human being can.

stage 6 Super intelligent AI Superintelligent AIs could possess intelligence that eclipses the combined cognitive abilities of every human that has ever existed.

stage 7 Self-aware AI

stage 8 Transcendent AI

stage 9 Cosmic AI

stage 10 God like AI.

Speed superintelligence

"Speed superintelligence" describes an AI that can function like a human mind, only much faster. For example, with a million-fold increase in the speed of information processing relative to that of humans, a subjective year would pass in thirty physical seconds. Such a difference in information processing speed could drive the singularity.

Technology forecasters and researchers disagree regarding when, or whether, human intelligence will be surpassed. Some argue that advances in AI will result in general reasoning systems that bypass human cognitive limitations. Others believe that humans will evolve or directly modify their biology so as to achieve radically greater intelligence. A number of future studies focus on scenarios that combine these possibilities, suggesting that humans are likely to interface with computers or upload their minds to computers that enables substantial intelligence amplification.

The book The Age of Em by Robin Hanson describes a hypothetical future scenario in which human brains are scanned and digitized, creating "uploads" or digital versions of human consciousness. In this future, the development of these uploads may precede or coincide with the emergence of superintelligent artificial intelligence.

Who cares about the Singularity?

Suleyman proposes that debates about the Singularity is a colossal red herring. He says that obsessing over this concept of super intelligence leads to overlooking the growing frequency of milestones occurring now. He believed that raising questions about synthetic media and misinformation, privacy, lethal autonomous weapons should be at the forefront of discussions. His belief is that getting sidetracked into a debate about whether consciousness requires some indefinable spark forever lacking in machines, or whether it will just emerge from neural networks is currently immaterial. He goes on to say that it does not matter whether the system is self-aware or has humanlike intelligence, but what really matters is what the system can do. The real challenge is that these systems can do more, much more, with every passing day.

Our next move

For most of history the challenge of technology lay in creating and unleashing its power. That has now flipped: the challenge of technology is today about containing its unleashed power, ensuring it continues to serve us and our planet. As David Chalmers say the age of AI could lead to an age of outcomes from the very worst to the very best, a utopia or complete annihilation.

Are we committed to holding on to our place at the top of the evolutionary pyramid?

Can we control the emergence of AI systems that are smarter and more capable than we can ever be?

Will we allow the emergence of AI systems that are smarter and more capable humans?

How will our morality be challenged?

‘We always overestimate the change that will occur in the next two years and underestimate the change that will occur in the next 10 years

Bill Gates

LB.