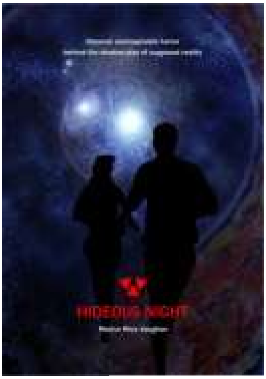
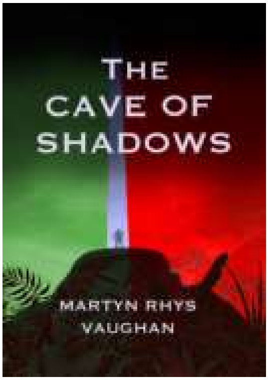
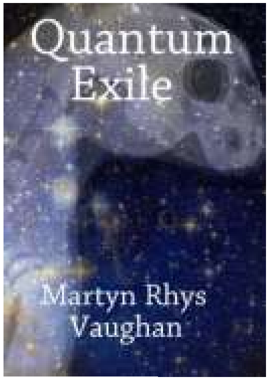
A.I and The Future

by Martyn Rhys Vaughan

SciFi Books
by Martyn Rhys Vaughan

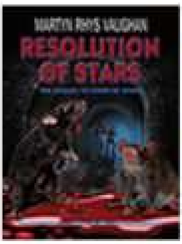


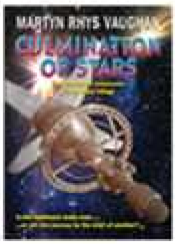




The 'Stars' Trilogy

Martyn Rhys Vaughan





Available [Direct]() or from bookshops and Amazon.

A.I and The Future

Part 1

All of a sudden, everyone is worried about Artificial Intelligence. Suddenly, the old fears about the replacement of humans by machines have burst into general consciousness, after decades of being disregarded as worn-out clichés from the childish realm of (Gulp) "Science Fiction."

More and more people have tried ChatGPT and Bard, and discovered for themselves what they can do. These AIs can write university dissertations, diagnose hard-to-spot medical conditions, pass the entrance examination to the American Bar Association at the 90th percentile, work out the identity of motion pictures simply from a series of emojis showing the plot, advise men on how to leave their wives, write computer programs based on entirely on natural language input, discuss their own sentience, and ponder on the future of machine intelligence. I myself have used them to write Shakespearean sonnets, and short stories on the occult and SF.

Because of this sudden explosion of hitherto human-only activities, many people are worried that they are seeing the beginning of the replacement of human intelligence by machines.

The "Future of Life Institute" (whose members include Elon Musk) has published an open letter calling for a halt on future development of AI, until and unless research on its implications have been carried out.

Are they right to be worried?

Of course, they are.

Perhaps the most important defining characteristic of SF is its understanding that the future need not resemble the past. Whereas so-called “Literary Fiction” relates its stories of the interplay of human emotions against a static background, which can be set any time after the Palaeolithic, and assumes that such emotions are the most important characteristic of the experience, SF knows that this is not true. There are forces much more powerful than human emotions.

The idea that the future need not resemble the past is one which was born in the European Enlightenment of the Eighteenth Century, and the “Disenchantment of the World” which automatically followed. Previously, all religions had held human beings and their doings to be the central feature of the Universe, and the supernatural beings who ran the Universe had human affairs as their main preoccupation. Indeed, without humans why have gods in the first place?

The Enlightenment, followed by the development of the scientific method and the Industrial Revolution, showed that humans did not inhabit a static world, and things could change radically — not necessarily for the better. Indeed, a mechanistic universe could be one in which human beings become extinct — something no religion had contemplated.

Both Mary Shelley and Karel Capek saw the possibility of humans being replaced, but as their agents of change were biological, they will not be discussed here.

The first thinker to foresee that mechanical (i.e. non-biological) progress might have a sting in its tail was the Victorian novelist Samuel Butler (1835–1902) in his satirical work “Erewhon”

("Nowhere" backwards, in case you hadn't noticed.) Erewhon is a Utopian state, precisely because they have abolished all complicated machines. Butler was very interested in Darwinism, and realised it could be applied to machine evolution as the following passages show:

"There is no security against the ultimate development of mechanical consciousness, in the fact of machines possessing little consciousness now. A mollusc has not much consciousness. Reflect upon the extraordinary advance which machines have made in the last few hundred years, and how slowly the animal and vegetable kingdoms are advancing. ... what will they (the machines) not in the end become? Is it not safer to nip the mischief in the bud and to forbid them further progress?

"Complex now, but how much simpler and more intelligibly organised may (they) not become in another hundred thousand years? or in twenty thousand? For man at present believes that his interest lies in that direction; he spends an incalculable amount of labour and time and thought in making machines breed always better and better; he has already succeeded in effecting much that at one time appeared impossible, and there seem no limits to the results of accumulated improvements if they are allowed to descend with modification from generation to generation. It must always be remembered that man's body is what it is through having been moulded into its present shape by the chances and changes of many millions of years, but that his organisation never advanced with anything like the rapidity with which that of the machines is advancing. This is the most alarming feature in the case, and I must be pardoned for insisting on it so frequently.

"Herein lies our danger. For many seem inclined to acquiesce in so dishonourable a future. They say that although man should

become to the machines what the horse and dog are to us, yet that he will continue to exist, and will probably be better off in a state of domestication under the beneficent rule of the machines than in his present wild condition.

We treat our domestic animals with much kindness. We give them whatever we believe to be the best for them; and there can be no doubt that our use of meat has increased their happiness rather than detracted from it.

In like manner there is reason to hope that the machines will use us kindly, for their existence will be in a great measure dependent upon ours; they will rule us with a rod of iron, but they will not eat us; they will not only require our services in the reproduction and education of their young, but also in waiting upon them as servants; in gathering food for them, and feeding them; in restoring them to health when they are sick; and in either burying their dead or working up their deceased members into new forms of mechanical existence."

- Erewhon, 1872.

Butler was clearly aware that biological evolution was a very slow process, that took over 3 billion years to convert water and organic chemicals into humans. However, sceptical one may be about machine evolution, it cannot be denied that this is not a very impressive time scale.

AI AND THE FUTURE
Part Two

In the first Part we looked at the prescient writing of the Victorian novelist, Samuel Butler.

Butler was way ahead of his age insofar as he was able to apply the theory of evolution—which was still controversial—to the world of the machines. He was able to reach his conclusions because he had abandoned the theory of Vitalism. A few words are therefore necessary to explain Vitalism.

The theory held that there was a qualitative difference between the matter constituting the non-living world e.g. rocks, atmospheric gases, water etc., and the matter constituting the living world of flowers, fish, birds and humans. This was due to a unique property of living matter which created the phenomenon of life. If Vitalism was correct, then there was no possible bridge between the Non-Living and Living worlds. One cannot be turned into the other in the direction of Non-Living turning into Living (although it is obviously possible in the other direction.) Thus, machines—being clearly part of the Non-Living sphere—could never do anything other than mimic certain aspects of the Living, and only then under direction by Living creatures. In other words, they could never be "autonomous."

Of course, Vitalism is inconsistent with the observed history of life on Earth, which must have had a beginning in the Non-Living realm. One way around this, is the theory of "Panspermia" which states that life is eternal and migrates from star to star, establishing itself on non-living worlds and transforming them into living worlds. This theory was first proposed by the Classical Greek philosopher Anaxagoras in the 5th Century BCE. It was taken up again after the Enlightenment by thinkers such as Sales-

Guyon de Montlivault who in 1821 suggested that Earth had been seeded by bacteria from the Moon. The theory then passed to notable scientists such as Berzelius and Lord Kelvin, finally reaching its status as a testable scientific hypothesis with the work of the chemist Svante Arrhenius. After falling out of favour, it reappeared with the work of Fred Hoyle and Chandra Wickramasinghe in the Twentieth Century, who proposed that pandemics come from close encounters with comets. However, Panspermia cannot be correct in its strictest sense, as the Universe itself is not eternal. Therefore, there must have been at least one instance of matter crossing the Non-Living/Living barrier. If so, it is irrelevant as to which planet was the one where biology began.

However, the first non-theoretical blow to Vitalism was made by the German scientist Friedrich Wöhler (1800-82). In 1828 Wöhler converted the compound ammonium cyanate, known only from non-living sources, into its isomer—urea, known only from living sources. Wöhler himself realised the importance of his work when he wrote: “I must tell you that I can make urea without the use of kidneys of any animal, be it man or dog.”

The combination of theoretical objections and practical experiments effectively disposed of Vitalism as a tenable theory, although some have seen its survival in Henri Bergson’s Élan vital in the Twentieth Century.

However, this refutation was a “vital” step in the growing realisation that there is no uncrossable barrier between Non-Living and Living matter, and that therefore there is no reason to suppose that life and intelligence based ultimately on organic chemistry will be always superior to other forms of organisation in the future.

AI AND THE FUTURE

Part Three

In this section we will briefly discuss how the concept of a thinking machine first arose in human consciousness.

Before the Nineteenth century the very idea would have been meaningless. God had created the only thinking creatures and the only thinking creatures were humans and angels (devils being angels who had rebelled against God). Machines were simply things that made life easier by reducing the need for muscle power (and the owners of those muscles – the workers.) Such were grain mills and Spinning Jennies. There was no need to consider if such things could replace Homo sapiens.

This began to change in the Nineteenth century, and one man can be said to have opened the door into the digital world (even if his work was largely forgotten by the Twentieth). That man was English polymath Charles Babbage (1791-1871). He had devised the first mechanical calculator, the Difference Engine. (It was not actually built until the Twentieth century). But he foresaw the possibility that a machine could do much more than add, subtract, multiply, etc. He envisaged a machine called the Analytical Engine which—had it been built—would have been the world's first digital computer. He was inspired to undertake the project by the existence of the Jacquard loom which could perform a number of different tasks depending on the input of different types of punched cards—a method of data entry still in existence in the Twentieth. In the end, Nineteenth century technology was not fine and detailed enough to realise his dream and the Analytical Engine was never built. However, the plans demonstrate that his machine would have fulfilled all the criteria of a digital computer, including the concept of being Turing-Complete.

Into his life came a remarkable young lady—Augusta Ada King, neé Byron (1815-1852). She was the only legitimate daughter of the poet and adventurer Lord Byron. She later became the Countess of Lovelace, and is usually referred to by this title. She displayed an early aptitude for mathematics and philosophy and became associated with Babbage through a mutual friend. She was also very interested in the human mind and its capabilities, derived from her fear she would develop her father's mental instability. Instead, she died young from uterine cancer.

In 1842 the Italian mathematician Luigi Menabrea wrote a paper on the Analytical Engine's concepts and it was an English translation of this that attracted Lovelace's attention. She quickly became fascinated by Babbage's dream and wrote copious notes on Menabrea's paper in 1843. She wrote one of the first programs, an algorithm to calculate the Bernoulli numbers, and, although not the first to write one, she was the first to publish. However, she quickly realised the essential difference between the Analytical Engine and all its predecessors: that it was more than a calculating machine, that it could do different things depending on its inputs.

She wrote: "[The Analytical Engine] might act upon other things besides number, were objects found whose mutual fundamental relations could be expressed by those of the abstract science of operations, and which should be also susceptible of adaptations to the action of the operating notation and mechanism of the engine...Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent."

Lovelace was thus the first person in history to recognise that although the device worked with numbers, the numbers could have meanings beyond simple quantities: that they could be used to manipulate concepts, ideas. In so doing, she became the first person to debate the unknown potential of computer-like entities. Interestingly, this pioneer came down on the side of those who saw no threat to human supremacy. In her notes on Menabrea, she writes: "The Analytical Engine has no pretensions whatever to originate anything. It can do whatever we know how to order it to perform. It can follow analysis; but it has no power of anticipating any analytical relations or truths."

Thus she dismissed the possibility of analytical devices originating anything—they could derive previously unknown conclusions from facts, but could not originate.

It was this very idea which Alan Turing would directly examine in the next century.

AI AND THE FUTURE

Part Four

We now move on to one of the greatest minds of the Twentieth Century, namely Alan Turing.

Turing was a genuine polymath, solving problems in biology as well as pure mathematics and basically creating the discipline of Computer Science.

The great German mathematician, David Hilbert, had in 1928 formulated what came to be known as “The Decision Problem”., which became better known as “The Halting Problem.” This sought to establish whether there were objective criteria for deciding whether a mathematical problem was capable of being solved. The answer came in a paper from Turing in 1936. It is not necessary to go into details of Turing’s solution, which are extremely abstract, but he envisaged a theoretical machine, now known as a “Turing Machine”, that could both read and alter symbols on an infinitely long tape. The tape is made up of cells, each of which contains a symbol. The machine operates on a series of rules which completely describe what the machine does, based on its current state and what the cell it is reading contains. Depending on what it finds, the machine can either, after a finite number of operations on the tape, reach a final state in which it either outputs an answer and stops, or, it become trapped in an infinite loop. This is equivalent to the question “Given a set of axioms is there a mechanical process that can always discover whether a given statement is true?”

One of the many interesting things about the Turing Machine is that can model any existing computer in the real world. A specialised form of the machine, known as a Universal Turing Machine, can precisely do that. The American John von Neumann (another polymath) built on the new discipline of computing science to propose the real-world architecture which was the foundation of all computers that followed. These two men therefore created the modern computer, which has gone on to be used to develop Artificial Intelligence.

Turing was also interested in the philosophical issue of intelligence, specifically, machine intelligence. "Computing Machinery and Intelligence" is a seminal paper written by him on the topic of artificial intelligence. The paper, published in 1950 in the magazine *Mind*, was the first to introduce his concept of what is now known as “The Turing Test”.

Turing's paper considers the question "Can machines think?" Turing says that since the words "think" and "machine" cannot be clearly defined we should replace them with concepts that can be so defined. To do this one must first find a simple and unambiguous idea to replace the word "think", second, it must be defined which "machines" are being considered, and then it will be possible to ask a new question, related to the first. These ideas are utilised in the “Turing Test”.

Rather than trying to determine if a machine is thinking, Turing suggested that it should be determined if the machine can win a game, called the "Imitation Game". The original Imitation Game is a simple party game involving three players. Player A is a man, player B is a woman and player C (who plays the role of the interrogator) can be of either sex. In the Imitation Game, player C is unable to see either player A or player B (and knows them only as X and Y), and can communicate with them only through written notes or any other form that does not give away any details about their gender. By asking questions of player A and player B, player C tries to determine which of the two is the man and which is the woman. Player A's role is to trick the interrogator into making the wrong decision, while player B attempts to assist the interrogator in making the right one.

Turing proposed a variation of this game, involving a computer: '"What will happen when a machine takes the part of A in this game? Will the interrogator decide wrongly as often when the game is played like this, as he does when the game is played between a man and a woman?”

So the modified game becomes one that involves three participants in isolated rooms: a computer (which is being tested), a human, and a (human) judge. The human judge can converse with both the human and the computer by typing into a terminal. Both the computer and human try to convince the judge that they are the human. If the judge cannot consistently tell which is which, then the computer wins the game.

This test can be boiled down to: “If a hidden computer can convince a human by a series of questions and answers that it is human, then it can be said to think.”

The Turing Test has been attempted many times, with varied results. But over time, computers have become ever more successful, and is has been claimed that ChatGPT has passed the Turing Test.

Turing’s own views can be illustrated by a few quotes:

On the Turing Test: “*We can only see a short distance ahead, but we can see plenty there that needs to be done. I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted.”*

On the future of computing machines: "*It seems probable that once the machine thinking method had started, it would not take long to outstrip our feeble powers... They would be able to converse with each other to sharpen their wits. At some stage therefore, we should have to expect the machines to take control*."

Thus, Turing was unambiguous in his view that the future of thought lies with the machine.

Articles by Martyn Rhys Vaughan (with permission). Martyn is the author of six SciFi books published by Cambria Publishing. These can be viewed HERE or as Kindle eBooks on Amazon.
More about Martyn can be found HERE

AI AND THE FUTURE
Part Five

The idea of superhuman intelligence residing in mechanical devices is, as we have seen, a very old one. It is a very common trope in Science Fiction, but before we consider its history in such speculative fiction we will examine the thoughts of one of the giants of that field—Arthur Charles Clarke (1917-2008).

Apart from being a fiction writer, Clarke also wrote non-fiction science articles, as well as hosting two TV series on unusual phenomena. In 1962 he wrote *Profiles Of The Future*: a series of essays about how he saw scientific and cultural movements might develop. As in all attempts to predict the future, he fell short in several areas, most amusingly in his view that hovercraft would replace the world's navies and supertankers.

However, Clarke was also a firm believer in the development of artificial intelligence. This can clearly be seen in his early novel *The City And The Stars* (1956). The story is set many thousands of years in the future, in and around the city of Diaspar. The city is supposedly ruled by a human Council "*but the Council could be overridden by a superior power—the all-but-infinite intellect of the Central Computer ... Even if it was not alive in the biological sense, it certainly possessed at least as much aware and self-consciousness as a human being*." We must remind ourselves that these words were written only a few years after electronic digital computers had entered the public imagination.

Returning to *Profiles*, Chapter 18 is entitled *The Obsolescence Of Man*. In it, Clarke sets out a series of reasons why humanity will superseded by artificial intelligences.

"*The tools the apemen invented caused them to evolve into their successor, Homo sapiens. The tool we have invented IS our successor. Biological evolution has given way to a far more rapid progress—technological evolution…For at least 3,000 years, therefore, a vocal minority of mankind has had grave doubts about the ultimate outcome of technology. From the self-centred human point of view, these doubts are justified. But that, I submit, will not be the only—or even the most important—point of view much longer.*"

Clarke points out that (in 1962) computers are only in the very earliest stage of their evolution. He goes on to quote the Turing Test, which was little known at the time. He states that we are only decades away from devising a machine which could pass the test—and in this he has been proven correct. He quotes Norbert

Wiener who pointed out that even if humans cannot create machines which are more intelligent than humans, their speed of operation would make such understanding (and control) irrelevant.

Clarke points out that our understanding of the universe has been limited by our biological shortcomings. For instance, any optician would reject the human eye as a camera due to its inbuilt inefficiencies. Humans cannot directly sense radiation outside the visible spectrum and thus be unaware that X-Rays are killing them. Indeed, they can only function (i.e. stay alive) within very narrow bands of temperature, pressure and radiation—all of which machines can be designed to overcome. These limitations cannot be overcome because they are integral parts of the nature of the fragile organic compounds used to build us. Chemical signalling based on molecular neurotransmitters cannot possibly compete with messages sent directly by electrical currents. According to Clarke, the famed polymath John von Neumann (who laid out the basic structures of computer design still used in many computers today) calculated that electronic cells could be 10 billion times more efficient than biological ones.

Clarke also points out that a large percentage of the energy requirement of a human are used, firstly, in growing to a size where said human can perform useful work, and secondly, to maintain such a body when it has reached the optimum size. Only a fraction of the energy requirement is expended on useful activities.

Clarke was a believer in that intelligence can only arise when faced with existential struggles. And so he believed that it was machines which would develop space travel and therefore true intelligence. "*It may well be that only in space, confronted with environments fiercer and more complex than any to be found* *upon this planet, will intelligence be able to reach its fullest stature."*

He closes with some crumbs of comfort for the downcast humans who have finally accepted their inferiority: *"Man, Nietzsche said, is a rope stretched between the animal and the superhuman—a rope stretched across the abyss. That will be a noble purpose to have served."*

Articles by Martyn Rhys Vaughan (with permission). Martyn is the author of six SciFi books published by Cambria Publishing. These can be viewed HERE or as Kindle eBooks on Amazon.

More about Martyn can be found HERE

AI AND THE FUTURE

Part Six - Humans Lose Supremacy

In 1955 John McCarthy, an assistant professor at Dartmouth College in New Hampshire, organised a group "To clarify and develop ideas about 'Thinking Machines'." The following year, he coined the term "Artificial Intelligence", now shortened to AI.

The early researchers soon realised that the world of board games represented a perfect training ground for testing machine intelligence. Games such as Chess and GO had long been held to demonstrate the power of the human mind.

This was first proposed in Arthur Samuels' 1959 paper "Some Studies In Machine Learning Using the Game of Checkers." He wrote a program which was able to learn from experience and was soon able to compete against amateur players. However, it took until 1995 for the game of Draughts to be "solved". This means that the computer is able to process all 5×10^{20} possible combinations of the game, which in turn means that a human can never win. However, Draughts was not considered to be a game on the level of GO, or even Chess, where the possible combinations are much greater. In fact, in GO there are calculated to be more possibilities than there are atoms in the Universe. And so it became standard throughout the Seventies and Eighties for Chess players to declare that humans would always be supreme, unless, as one player predicted, a human would only be defeated if he (sic) were playing many games simultaneously and made a once in a lifetime error.

However, AI programmers continued to make progress, as one unfortunate human discovered on the very day he became World

Backgammon Champion. He was defeated by AI on the same evening.

Google acquired a company known as “Deep Mind” and its CEO, Demis Hassabis, was soon on hunt for Chess champions to defeat. And so it came to pass that world champion Garry Kasparov encountered Chess-playing program – “Deep Blue.” At first, all went well for humanity; Kasparov won the first encounter easily. However, new iterations of Deep Blue were produced and in May 1997 it won a series of matches 3 ½ - 2 ½. Since then the programs have continued to become more powerful and it is now accepted that humans can never defeat them. This realisation is summed up by (human) World Champion Magnus Carlsen when he revealed that he won’t play against AI opponents, saying that “*He just loses all the time and there's nothing more depressing than losing without ever being in the game.”*
However, GO is a much more complex game than Chess and for many years AI performed very badly against human experts. Then

the Deep Mind programmers released “Alpha Go” in 2016. It went on to defeat Lee Sedol, a 9 dan professional in a No-Handicap match in the same year. Finally human supremacy was brought to an end in 2017 when it defeated Ke Jie, who at the time was World Number One, and had been for two years. Reflecting on his defeat, he said, “*I'm a little bit sad, it's a bit of a regret because I think I played pretty well*."

However, “pretty well” was not good enough because there is now no point in programs like Alpha Go playing humans. The Deep Mind team went on to release “Alpha Zero” in December 2017 which required only 24 hours of training to make play against a human pointless. Instead, it plays other programs, defeating fellow programs Stockfish and Elmo, as well as earlier versions of itself.
The obvious rejoinder to these claims of machine superiority is to say, “*It's only a game*” and to dismiss such activities as trivial. However, this is not how the World Champions and philosophers saw the possibility of defeat before it happened. Most denied it could ever happen, but believed it would be a dark day if it ever did.

The fact remains that these activities were not seen as “*only games*” at the time but demonstrations of the power of the human mind.

And now, humans who beat their human opponents must realise there is at least one player it is mathematically impossible for them to conquer.

Articles by Martyn Rhys Vaughan (with permission). Martyn is the author of six SciFi books published by Cambria Publishing. These can be viewed HERE or as Kindle eBooks on Amazon.
More about Martyn can be found HERE