


Building the future in the 21st century: In conversation with Yuval Noah Harari

**Author:**Anton A. van Niekerk^{1,2} **Affiliation:**¹Centre for Applied Ethics, Stellenbosch University, Stellenbosch, South Africa²Department of Philosophy, Faculty of Arts and Social Sciences, Stellenbosch University, Stellenbosch, South Africa**Corresponding author:**

Anton A. van Niekerk, aavn@sun.ac.za

Dates:

Received: 24 Apr. 2020

Accepted: 18 June 2020

Published: 20 Nov. 2020

How to cite this article:Van Niekerk, A.A., 2020, 'Building the future in the 21st century: In conversation with Yuval Noah Harari', *HTS Theologiese Studies/Theological Studies* 76(1), a6058. <https://doi.org/10.4102/hts.v76i1.6058>**Copyright:**

© 2020. The Authors. Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Yuval Noah Harari has proved himself to be one of the most prominent and accessible historians of the 21st century. He has not only popularised a so-called dialectic with the past but also encouraged speculation about the history of the future. This article critically engages with Harari's revolutionary projections in an attempt to evaluate the lessons and concerns that one ought to take away from his work. More specifically, the ever-increasing achievements in the world of science and technology need to be balanced by humility. *Homo sapiens* are in the unprecedented position to shape their own evolution – thanks to what is often termed 'the Fourth Industrial Revolution' – and speedily move towards the emergence of a new kind of being. The very nature of being in such an unparalleled situation is that there now exists an unprecedented discontinuity between the past and present. With this in mind, is it even possible to make valid predictions about what the future holds in the manner that Harari does? Is it our responsibility to take on this task at all? These are some of the questions that this article grapples with as an impetus for suggesting provisional guidelines for humanity to follow when we inevitably take the future into our own hands.

Contribution: This article forms part of a collection that not only reflects on the origin and development of creation, but also on what the possible future might look like. It is based on historical thought and source interpretation that fits well with the intersectional and multi-disciplinary approach of this journal and collection.

Keywords: Algorithms; Harari; Historicism; History of the future; *Homo sapiens*; Longevity; Morality; Regulation; Responsibility; Story-telling.

A history of the future?

It is quite striking that historians – with the Israeli thinker Yuval Noah Harari as the most prominent example – are increasingly concerned with and writing about the 'future'. Historians since the days of the ancient Greeks like Thucydides and Herodotus have always been assumed to be writing about the past. The future is something that, almost by definition, cannot be known as it has not (yet) been experienced. This also held for whatever answer was offered in response to the question: What is history? Is history the 'events' that took place in the past and which, for that very reason, are in principle no longer fully accessible to us living in the present? Or is history the story that we tell about the past? Is history therefore essentially a creation of historians, on the assumption that they take considerable trouble to verify the stories that they tell as accurately as possible, drawing on whatever sources that present themselves as methodologically reliable? The latter, to my mind, is the more plausible view. History cannot in any way be exempted from the necessity of 'interpretations' by historians (Gadamer 1975:153–332). The First World War, for example, is not a brute (historical) fact waiting to be registered in the annals of historians. Today, the 'First World War' (as a 'historical fact') is assumed to have lasted from 1914 to 1918. Yet before the Second World War came about (from 1939 to 1945; probably more accurately from December 1941 to 1945), there was no 'First World War' – only 'the Great War', (the 'war to end all wars') – as is attested to in numerous monuments commemorating that pivotal event in European history between 1918 and 1939.

Despite this fact, Harari unashamedly calls his second book *Homo Deus*¹ (i.e. man has become God) 'a brief history of tomorrow', broadly expanding the vision that he set out in the last part of his initial best seller, *Sapiens* (Harari 2011). *Sapiens* is an internationally widely acclaimed book. It tells, in a riveting story of about 500 pages, how we as a species came to be what we currently

.....
1. *Homo Deus* will henceforth be abbreviated to HD (Harari 2017).

Note: Special Collection entitled Challenging Building Blocks: The coherence of cosmological evolution, life, consciousness, morality and building the 21st century, sub-edited by Chris Jones (SUN) and Juri van den Heever (SUN).

Read online:

Scan this QR code with your smart phone or mobile device to read online.

are. It is the story of how a seemingly insignificant group of ape-like creatures that were hunted down by 'natural predators' in a way similar to what currently occurs in nature to other species redeemed themselves in order to become masters of the current-day world. This was particularly made possible by the inadvertent development of language – language that enabled these creatures to not only communicate with one another in order to facilitate more security but also, above all, act as the key instrument of social organisation. This social organisation was the basis of the historical progress in the evolution of humans that we have seen throughout the history.

Sapiens' secret: Language, myth and storytelling

Before pursuing Harari's riveting tale further, we need to know something more about the kinds of language we see in the world in order to better understand the significance of what has just been claimed about language and its role in social organisation. Karl Popper identified four functions of language (see Popper 1972:235–238). Firstly, there was the 'significatory' function, in terms of which a sound or sign functions to identify something else, typically the kind of language use in terms of which a certain bird makes only one kind of noise when it sees a snake. Secondly, there was the 'expressive' function, in terms of which (a) sound(s) is/are made to express feelings or emotions; for example, the barks and cries of dogs. To this very day, *Homo sapiens* share these two language functions with other species in the animal kingdom. What elevated Sapiens above the rest, however, were two additional language functions. The first of these was the 'descriptive' function – the ability to format, by means of sounds (and later written signs), a description of something that is not at the moment present for either speaker or hearer. No other species has, to our best knowledge, ever succeeded in mastering this skill. The final function which was key to elevate Sapiens into capacities that were instrumental in the development of their mastery of the world was what Popper called the 'critical' (or argumentative) function of the language. This refers to our ability to not only, by means of sounds and signs, provide a description of something not present but also respond critically to that description and thus develop science. 'Science', as the Dutch philosopher RC Kwant argued, is the 'critical purification of whatever we claim to be the case'², that is, the effort to say, as accurately as possible, whatever we claim to be the case in the world, respectively, to make assertions about what is the case in as 'uncontroversial' a manner as possible.

According to Harari, the most powerful capacity with which language endowed our species is the possibility of 'social organisation'. He (Harari 2017) explains as follows:

Over those 20000 years [i.e. the period stretching from our background as hunter-gatherers, through agricultural settlements, to the modern world] humankind moved from hunting mammoth with stone-tipped spears to exploring the solar system with

²'De wetenschappelijke kritiek is immers op de eerste plaats zelfkritiek van het spreken' (Kwant 1964:54).

spaceships, not thanks to the evolution of more dexterous hands or bigger brains (our brains today seem actually to be smaller). Instead the crucial factor in our conquest of the world was *our ability to connect many humans to one another*. Humans nowadays completely dominate the planet not because the individual human is smarter or more nimble-fingered than the individual chimp or wolf, but because *Homo sapiens* is the only species on earth capable of *cooperating flexibly in large numbers*... If cooperation is the key, how come that ants and bees did not beat us to the nuclear bomb even though they learned to cooperate *en masse* millions of years before us? Because their cooperation lacks flexibility³...To the best of our knowledge, only Sapiens can cooperate in very flexible ways with countless numbers of strangers. This concrete capability – rather than an external soul or some unique kind of consciousness – explains our mastery of planet Earth. (pp. 153–154; [author's added emphasis])

One might think that the arrival of language as key to social organisation was also primarily key to promoting the biological security and survival of the tribe of Sapiens wandering as hunter-gatherers in the hostile environment in which our ancestors found themselves roundabout 20 000 years ago. As soon as settlements for the sake of agricultural activities developed, followed by the development of writing and money by the Sumerians 5000 years ago, the potential of descriptive and argumentative language showed itself in what Harari reckons to be the key phenomenon in the evolution of Sapiens. This phenomenon was the discovery of the force and impact of 'storytelling' in order to mobilise and organise large groups of humans. The stories were devoid of all truth in the modern, scientific sense of the word, but we must appreciate that this occurred in a world where no sense of the value or desirability of scientifically argued or rationally demonstrable 'truth' existed. The stories were the origin and drivers of religion, politics and economic organisation. The great discovery was that people could be persuaded to believe master narratives about their origins, identities and, particularly, final destinations; and they were willing to be mobilised into action (war, sacrifice, worship, servitude, etc.) on the basis of these stories.

The force of myth in constituting human identities cannot be overestimated. Myth had three characteristics. Firstly, mythical stories were believed without experiencing any compulsion that the claims of the myth ought to be the outcome of independent investigation and discovery. Secondly, people believed myths as key instruments to influence and to protect themselves against their hostile environment. The 'knowledge' facilitated by myths was never attained purely for the sake of a disinterested understanding of 'how things actually work'. It was always the knowledge acquired for the sake of influencing (and hopefully controlling) the hostile, power-infused natural environment. Finally, no critical analysis of, or conversation with, the myth was possible. Critical questions

³This correlates with Popper's appreciation of both the descriptive and critical functions of language. It is, in particular, the critical function that facilitates the flexibility that Harari so highly appreciates. In social organisation – the key to human mastery – there is not only one way of doing things; we learn from mistakes and continue to do it better or more effectively.

about myths did not occur and would not have been understood. For ancient humans, the discovery was much more important in the sense that myths or master narratives could be utilised to frighten people about all kinds of made-up eventualities and thus enable the power in society to manipulate and control people on a large scale.

Harari (2017) writes in this regard:

Sapiens...live in triple-layered reality. In addition to trees, rivers, fears and desires, the Sapiens world also contains stories about money, gods, nations and corporations...Humans think they make history, but history actually revolves around this web of stories... It all started about 70 000 years ago when the Cognitive Revolution [facilitated by language as explained earlier] enabled Sapiens to start talking about things that existed only in their imagination [like gods, nations, devils, etc.]. (pp. 181–182)

People were mobilised by these stories to live, work and interact in ways prescribed by these stories. 'The Agricultural Revolution, which began about 12 000 years ago, provided the necessary material base for enlarging and strengthening the inter-subjective networks' (HD:182). Harari explains further and also applies this insight to modern-day people: 'Fictions enable us to co-operate better. The price we pay is that these same fictions determine the goals of our cooperation which are harnessed to serve fictional aims and desires' (HD:203).

Much has changed since the mythical times. With the advent of modernity, people started to insist that knowledge claims ought to be the outcome of independent reflection and investigation. Knowledge came to be ascertained from the perspective of a theoretical approach in which insights are sought for their own sake, and not as the key to manipulating the world in my and/or our favour. Finally, the critical and/or argumentative function of language particularly kicked in as far as the scientific investigation of the world developed.

What did not change, however, was reliance upon the force of stories or master narratives to organise and control people. Harari (2017) writes in this regard:

In the twenty-first century, we will create more powerful fictions and more totalitarian religions than in any previous era. With the help of biotechnology and computer algorithms, these religions will not only control our minute-by-minute existence, but will be able to shape our bodies, brains and minds, and to create entire virtual worlds complete with hells and heavens. Being able to distinguish fiction from reality and religion from science will therefore become more difficult but more vital than ever before. (p. 207)

Harari's *Homo Deus* is then his story of the future that awaits us in the 21st century. It is a dazzling array of stories about the heights and depths that the rapid growth of new technologies – some of which are present already, and some of which are only imagined and not necessarily realisable – will take us to. It is far too complex a story to cover adequately in the space available to me in this article. What I will do in the rest of the article is to, with very broad brush strokes, share some of the revolutionary possibilities that Harari

relates. This will be followed by a section in which a few critical questions and remarks about the project of his book will be pursued. The chapter will be concluded by an independent evaluation of what we ought to learn from Harari's work and what the concerns are that we ought to take into account in order to prevent developments that would be detrimental to the outcome of the futuristic history that Harari wishes us to foresee and experience.

Increased longevity

The gist of Harari's tale is that our species has amassed great achievements, but are set to see a much more dramatic set of developments in the years to come. There are many things that are possible and that represent excellent news to look forward to. Harari is outspoken about the central technological project that will be driven in the 21st century, namely 'significantly increased longevity'. Life expectancy will at the very least double in the coming century. It increased from 40 to 70 in the course of the 20th century. In the 21st century, it will again double to roundabout 150 (Harari 2017):

Though falling far short of immortality, this would still revolutionise human society. For starters, family structure, marriages and child-parent relationships would be transformed. Today, people still expect to be married 'until death us do part', and much of life revolves around having and raising children. Now try to imagine a person with a lifespan of 150 years. Getting married at 40, she still has 110 years to go. Will it be realistic to expect her marriage to last 110 years? (pp. 29–30)

Harari is still quite modest in his projections of expected longevity. Aubrey de Grey has recently drawn much attention to the considerably bolder claims he makes in this regard. He identifies himself as a 'biogerontologist' and is currently chief science officer of the Strategies for Engineered Negligible Senescence Foundation in Cambridge. He claims that ageing is no more than a widespread, extremely destructive, yet avoidable, cluster of diseases that could all be reversed or cured in view of extending our lifespan almost indefinitely. 'We can grow old without becoming aged'. To quote De Grey (2008) himself:

Recent biotechnological progress indicates that many aspects of aging may indeed be effectively treatable by regenerative medicine in the foreseeable future. We cannot yet know whether all aspects will be curable, but extensive scrutiny has failed to identify any definite exceptions. Therefore, at this point there is a significant chance that such therapies would postpone age-related decline by several years, if not more, which constitutes a clear case for allocating significant resources to the attempt to develop those therapies. (pp. 67–68)

Just about every disease or degeneration of the human body associated with the process of ageing can, according to De Grey, be 'reverse-engineered'. The big challenge is to arrest the process of ageing, not at the equivalent of the bodily prowess of a 90-year old, but rather at that of a 40-year old. De Grey is confident that the first person to reach the age of 1000 has already been born, and he has high hopes that it will indeed be himself.

Adversity as a 'mere managerial project'

This technologically mediated Nirvana of longevity is by far not the only revolution that Harari foresees for the coming century. Steven Shapin (2017) provides a useful summary of some of the other transformations envisioned by Harari:

Once upon a time...we reckoned there was not much we could do to prevent or counter epidemic disease; we looked on death and famine as bad hands dealt by faith or divine judgement; we considered war to be in the nature of things; and we believed that personal happiness was a matter of fortune. Now, Harari says, these problems all have been reconfigured as managerial projects, subject to political will, but not limited by the insufficiencies of our knowledge or technique. We have [nowadays] become the masters of our own fate – and 'fate' itself should be reconceived as an agenda for further research and intervention. That is what it means to refer to the world era in which we live as the Anthropocene: one biological species, *Homo sapiens*, has become a major agent in shaping the natural circumstances of its own existence. [John Harris as well as the author of this article and others have, with reference to the genetic revolution and the accompanying possibilities of radical biomedical enhancements of humans, noted that few things illustrate our uniqueness as a species as much as our currently demonstrated capability of taking our own evolution into our own hands. Harris 1998; Van Niekerk 2014]. The gods once made sport of us; the future will upgrade humans into gods, and turn *Homo sapiens* into *Homo deus*. (n.p.)

With reference to the phenomenon of historical change, Harari stringently argues against the modernist conception that there are distinct 'laws of change'. He is also strongly polemicising against the 19th century idea that history shows inevitable progress, as, for example, Herbert Butterfield's book *The Whig interpretation of history* (Butterfield 1932) argues. The trajectories of the past that Harari discusses in *HD* are not driven by the regularities of alleged historical laws. In as far as these can be identified as recent historical phenomena, the script for that history is written by the 'innovations emerging from the minds of a visionary technical elite' (Shapin 2017). 'History is often shaped by small groups of forward-looking innovators rather than by backward-looking masses' (Harari as quoted by Shapin 2017).

Harari is also full of confidence that our control over infectious disease will increase dramatically in the 21st century – an encouraging word for a world that at the time of writing (March 2020) seems to be spinning out of control because of the threatening ravages of the Corona or SARS-COVID-19 respiratory virus. Where problems remain, it will not be because the future threats will be insurmountable or the science is inadequate. It will be because of a lack of political will or the proper mobilisation of resources.

Less good news

When Harari predicts the future, he is, however, far from reality and always buoyant with optimism. What is surprising is that, at the same time, he is not so much pessimistic about

themes that we have become accustomed to be pessimistic about. He does offer words of caution about topics like the environmental crisis and the threat of nuclear war. However, Harari is much rather concerned, not with potential developments that pose the threat of annihilating our species, but rather with the technologically driven developments that could transform our species into something that challenges the nature of our humanity. In other words, his foremost concern is with what we and our world might become on the basis of the spectacular potential of species-transformation via science and technology.

The core of these (actual and potential) developments is again aptly summarised by Shapin (2017):

Homo sapiens may cease to be, not because earth will become uninhabitable or because Donald Trump or Kim Jong-un will push the button, but because we will become new kinds of beings: our bodies, minds and relationships with the environment and with mechanical devices will be altered in fundamental ways. Our capacities to know and to do things have always depended on our tools, but now these tools are acquiring the potential to transform what it means to be human. New tools will become parts of our bodies: we will have bionic hands, feet and eyes; nanorobots will cruise the bloodstream keeping an eye out for disease and repairing the damages of age and injury; wearable and implanted devices will expand our sensory repertoires and alter our moods; biological tools will infiltrate our cells, redesign our genes and give us new and improved flesh, blood and neurons. So the biggest threat to *Homo sapiens* is that the technological 'upgrade' that makes us as gods will, at the same time, redefine human capacities. (n.p.)

Harari's 'building blocks of the future' are therefore no longer historical regularities or the idealistic or material dialectics that thinkers like Hegel and Marx promoted so powerfully in the 19th century. The blocks much rather are the remarkable achievements of science and technology that are busy becoming the formatters of history in our time. The realisation of the reality of these developments is indeed a humbling experience. Harari, as we have seen, is quite critical of religion, in spite of its earlier capacity to, on the basis of key stories or master narratives, mobilise people to the extent that it did. Nowadays we are increasingly realising that humans hold no necessarily central place in the realm of the animals. What elevated us to our current position is the contingency, yet unforeseeable force of language and its power to facilitate social organisation. However, it also means that (Shapin 2017):

[T]here is no essential human 'self'; our thoughts and emotions are the product of the electrochemical impulses which can, in principle, be modelled by the formal problem-solving rules we call algorithms; our bodily frames and mental capacities have evolved over time and there is nothing fixed in our 'nature'. [Harari argues] that the only thing that can be predicted with certainty about human nature is that it will change...we will become more god-like as we become more machine-like and as machines' capacities become more god-like. Humanity's future is in the hands of technical experts – in biotechnology, artificial intelligence, cognitive and computer science. (n.p.)

Harari does not foresee a much longer future for *Homo sapiens*. We are destined to see the emergence of a new kind of being. The latter will have physical characteristics vastly superior to ours. They will be our superiors in both cognitive capacities and emotional dispositions. Although he does not explicitly discuss it, a vast literature is today also emerging on the likelihood of these new beings as not only 'cognitively', but also 'morally' enhanced. One of the main arguments for moral enhancement is the simple claim that we as a species in our current state of moral dispositions simply are not (morally) good enough to make decisions that, for example, have to be made in order to counter the ecological crisis precipitated by climate change (cf. Persson & Savulescu 2011).

In the dystopia that Harari sketches, war will be waged by drones and work will be done by robots. It is not clear what the place and function of unenhanced human beings might still be in such a world. What we are already seeing, and will increasingly see in future, is that tasks and everyday decision-making processes will increasingly be performed, no longer by the spontaneous exercise of free will, but by the operations of 'algorithms'. This requires some explanation.

Algorithms

What is an 'algorithm'? For Harari, this term refers to a 'recipe for decision-making'. It can occur not only in organisms but also in machines. Hence, Harari (2017) quoted:

An algorithm is a methodical set of steps that can be used to make calculations, resolve problems and reach decisions. An algorithm is not a particular calculation but the method followed when making the calculation. (p. 97)

An example is the 'recipe' for calculating the average of two numbers or for baking a cake.

We (humans) have, every time, to think and/or read in order to bake the cake or to construct vending machines. The recipe can, however, be 'inserted' into a machine which will do it perfectly every time it is so instructed. According to Harari, we, as humans, are also algorithms, or do many things according to algorithms operating in our bodies and brains.

Harari (2017) quoted:

Humans are algorithms that produce, not cups of tea [or vending machines] but copies of themselves...The algorithms controlling vending machines work through mechanical gears and electric circuits. The algorithms controlling humans work through sensations, emotions, thought. (pp. 98–99)

Human and/or animal bodies calculate (like Deep Blue playing chess) their best chances of survival, for example, a chameleon changing colour in order to remain invisible to predators. 'These [human/animal] algorithms undergo constant quality control by natural selection. Only animals that calculate probabilities correctly leave offspring behind' (HD:99–100).

What is now bound to happen in future is that 'algorithms embedded in silicon and metal will replace algorithms embedded in flesh which, Harari reminds us, is what biology and computer science tell us is all we really are anyway' (as has just been shown above; Shapin 2017). To this will be added the fact that new life forms will be created – life forms that will transform the assumption that life itself is an exclusive function of organic compounds. This will all be the outcome of what Harari calls the 'decoupling' of intelligence and consciousness, brought about by the advances of AI (HD:361). Henceforth, intelligence as produced and increasingly upgraded by artificial intelligence (AI) machines will no longer require any organic basis. Consciousness, by which he means sensations such as grief, sympathy and the like, still requires an organic basis, and we still do not know whether or how it might one day be transferred to robots and other AI machines.

Critical concerns

The 'history of the future' conjured up in Harari's *Homo Deus* is, indeed, a world of dazzling, almost inconceivable and often highly disconcerting possibilities – possibilities that, in many instances, have progressed to reality. Harari is a gifted and highly entertaining author who has, without any doubt, opened up the public discussion of the advent of the so-called Fourth Industrial Revolution with unprecedented prominence.

Of course, that does not mean that the vision he develops is a *fait accompli* that we can, with certainty, expect to be realised in the rest of the coming century. Whereas much of Harari's work is surprising, gripping and disturbing, a range of critical questions and concerns do offer themselves and require to be formulated and addressed.

The first and most self-evident question that comes to the fore is that about the ambition proposed by his overall project, which is to 'predict the future' – particularly the envisaged story (or 'history'!) of the 21st century.

Now, if we are to learn anything from history, it is indeed the need for humility in our endeavours to predict future events accurately. One is often dumbstruck by the reality of unfulfilled historical predictions. The final triumph over cancer has often been predicted, particularly as a result of the remarkable progress made in this regard in the second half of the 20th century; yet, while significant strides have been made in managing this disease, we are far from a situation where the triumph over cancer is ready to be announced.

Predictions were similarly made in terms of radical increases in the speed of civil air travel that have up till now come to naught. With the advent of personal computers since the 1980s, there were high hopes that the 'paperless office' would soon be realised. In reality, computers have produced more 'paper factories' than ever before. The history of aviation, looked at from the perspective of the

mere 66 years it took to progress from Kitty Hawk to the Moon landing of 1969, is arguably one of the greatest technological feats that humankind has achieved. It in turn created high hopes of significantly increased and prolonged space travel – even the ‘conquering of space’; yet, more than half a century later, we are not seeing any developments that nearly approximate those expectations; even a human journey to the next colonisable planet (Mars) is probably decades away, if it will be realised at all. As late as 1932, Albert Einstein could not see that nuclear energy would ever become a reality. The Decca record company announced when first listening to the Beatles in 1962: ‘We don’t like their sound, and guitar music is on the way out’. Dr Dionysius Lardner predicted in 1830 that rail travel (still a novelty in his time) at high speed is not possible because passengers, ‘unable to breathe’, would ‘die of asphyxia’. Shapin (2017) rightly claims: ‘Predictions worth paying attention to should have an expiration date. Harari’s [and De Grey’s!] extrapolations from historical trends of a vast increase in human longevity may prove similarly unsafe’.

When it comes to predicting the future, one is at the same time reminded of events that were not (and probably could not) be predicted. Cell phones could indeed be predicted as a ‘wireless’ modification of the standard telephone – but not as hand-sized computers as they turned out to become; but, for the latter, there was initially little enthusiasm. A Western Union internal memo stated in 1876 that ‘this ‘telephone’ has too many shortcomings to be seriously considered as a means of communication. The device is inherently of no value to us’. The best known examples of predictions that seemingly just could not be made were the personal computer and the World Wide Web. The ‘equally unanticipated bitcoin will [probably] end up radically transforming the nature of commercial transactions’ (Shapin 2017). The same author concludes: ‘What can’t be imagined can’t be predicted’.

Harari’s book is indeed one that commands interest and attention. However, it is a book that, for its eventual credibility to be established beyond reasonable doubt, depends on the judgment of history. It commands, in a way very similar to the ‘singularity’ predictions of Kurzweil (2006), an attitude of ‘we shall see’.

One of the most controversial claims of Harari’s historical predictions is the implied view that the continuity between past, present and future will be destroyed or seriously compromised. Harari is indeed claiming that we are nowadays, and increasingly in what lies ahead, confronted with a new kind of historical actor, namely the formatters of technological innovation which will, increasingly, manifest the achievements of AI on a scale that is entirely discontinuous with the past. In addition, the ‘black box’ phenomenon is indicative of the fact that computers and other AI machines are able to learn spontaneously and on their own, on the condition that they are ‘fed’ enough with relevant data. A good example of this is the so-called ‘melanoma machine’ which has generated its own algorithm for the detection – generally more accurately than by dermatologists – of the life threatening melanoma

tumour as its output. The intriguing thing about this black box phenomenon is that computer scientists and doctors are not yet able to reconstruct exactly how that learning process occurred; the machine ‘learned of its own accord’. The learning occurred spontaneously as the outcome of so-called ‘deep learning’ based on data inputs. This probably constitutes one of the greatest revolutions of our time.

Harari’s predictions do often create the impression that he does regard history to be predictable, although not in terms of the way in which it was done in the past. The latter refers to the approach of historical determinism – a position that has widely been discredited. By ‘historical determinism’ is meant the claim that history is the playing out of one or a series of unchangeable patterns or directive models. The best known caution against misconceptions in this regard was expressed by Sir Karl Popper (1957) in his classic book *The Poverty of Historicism*.

Historicism is a position that Popper spiritedly opposes in this book, which, according to him, is one of the main causes of totalitarian thinking. He (Popper 1957) defines historicism as:

[A]n approach to the social sciences which assumes that historical prediction is their principal aim, and which assumes that this aim is attainable by discovering the ‘rhythms’ or the ‘patterns’, the ‘laws’ or the ‘trends’ that underlie the evolution of history. (p. 3)

Once such ‘rhythms’ or ‘trends’ are identified in history; they are reified, that is, they are seen as fixed, inevitable and unstoppable. They become ‘history’, that is, a metaphysical force that operates persistently and inevitably. Historicism entails the position that insists on the extrapolation and projection of these trends to the future. In that sense, it insists that we ‘learn from history’. However, ‘learning from history’ in this case does not mean the creative discernment of that which is of greater or lesser value in the past in order to improve our situation. Much rather it means to subject oneself and one’s society to alleged ‘regularities’ that will impose themselves and will establish their effects with ‘historical necessity’.

Examples of such ‘trends’ include Hegel’s idea of the ‘cunning of reason’, that is, the inevitability with which the ‘Objective Spirit’ of dialectical reason holds the development of world history under its sway (Hegel 1991), as well as Marx’s idea of historical materialism, that is, the idea that history is both generated and operationalised by class struggle, that the exploitation of the working class is a necessary condition for the survival of capitalism and that the unavoidable contradictions of capitalist society point towards the inevitable demise of capitalism in the final apotheosis of the communist revolution and the classless society (see Marx & Engels 1975:477–519).

This kind of extrapolation of the future from the past is not only empirically and scientifically dubious. It is politically

and ideologically dangerous. The ‘trends’ or ‘patterns’ that historicism identifies are never indubitable historical realities, but invariably ideological constructs that fit the political and social agenda of certain people in society. A future based on such ideas is seldom the emancipation of which Hegel and Marx dreamed, and more often, the criminal enslavement of ordinary people who are the victims of the unattainable utopias conjured up in the minds of the derailed tyrants of totalitarian states such as Hitler and Stalin – people who were prepared to trample on anything to attain their misguided utopias, even if it meant stepping on and over dead bodies.

It would be unfair to accuse Harari of full-fledged historicism. However, the main thrust of his *HD* is a clear attempt to predict the future and to tie that prediction to one model of historical change, namely technological innovation to the point that we will probably see the human species in its present guise disappear completely. This, to say the least, represents, if not an unduly arrogant, then at least immodest agenda. Harari’s text is one that could easily incline readers to flirt with historicist ideas, which would undermine the credibility of his text considerably.

Closely related to this is Shapin’s point about the recursive nature of some of Harari’s predictions ‘where the outcome is not in fact independent of the prediction’ (Shapin 2017). Shapin (2017) writes:

Some prophecies are self-fulfilling: if people are agreed that certain things can happen, or will happen, they may devote resources and co-ordinate their energies to make them happen. In the case of Moore’s law, you could say the increase in computing power was the result of technological expertise and self-fulfilling prophecy combined: the ‘law-like’ increase in semi-conductor density held up in part because technologists’ confidence that such things were possible encouraged the mobilisation of resources. Then there are self-negating prophecies. If you believe, for instance, that automated computerised systems are safe, you may neglect due care and vigilance, so making them dangerous; but if you believe that buildings may be vulnerable to earthquakes, or that nuclear power stations are inherently dangerous, you may take steps to ensure their integrity, so making them safe. (n.p.)

These are possibilities that any attempt to seriously ‘predict the future’ ought to reckon with. The special kind of ‘building blocks of the future’ that Harari provides us with do not function *ex opere operato*⁴; the very way in which the future is predicted can significantly contribute to foreseen and unforeseen outcomes.

Decision

In spite of the array of critical concerns raised against Harari, the value of his work can hardly be overrated. The totally unforeseen and often disconcerting revolutions that

await us in the rest of the present century (let alone in what comes after that) have been elevated by his work to a centre piece of public discussion in a way that almost has no precedent. Harari has a remarkable ability to guide a (mostly) lay audience through a labyrinth of highly complex insights and developments, and to succeed in fusing that complex set of insights with ordinary public discourse. In that way, he has succeeded where scientific and technological experts have often failed. Harari has, at the same time, also transformed the genre of historiography itself; the writing of history will, because of his work, never be the same, if for no other reason than that he has elevated the alleged impenetrability of the future to the stuff of everyday informed intellectual conversation.

The world of *Homo sapiens* has always been subject to change. Right up to the advent of modernity (i.e. since the 17th century) the world changed, but people were hardly aware of it. The chances of a man earning his livelihood in exactly the same way that his father and grandfather(s) did was the order of the day in (current-day terms) ‘developed’ and ‘developing’ societies. In ‘developed’⁵ societies, this only started to change significantly with the advent of the industrial revolution in the 18th and 19th centuries. The big change came in the 20th century when, what Alvin Toffler (1970) rightly called ‘future shock’ set in, that is, the experience, on an annual basis, of dramatic changes that people experienced (in transport, communication, health care and information). Now, with full force, people became aware that the world changes almost on a daily basis and that, although we have to deal with future shock, we are adaptable and learn remarkably soon to deal with a constantly changing environment.

Harari brings to our attention that the changes we have to deal with do not only pertain to our physical and intellectual environment, but also to the very structures by means of which we live our lives, namely our bodies. We are on the verge of seeing dramatic machine technologies incorporated into our bodies, even up to the possibilities of cyborg engineering. Harari (2017) explains that:

Bioengineering is not going to wait patiently for natural selection to work its magic. Instead, bioengineers will take the old Sapiens body, and intentionally rewrite its genetic code, rewire its brain circuits, alter its biochemical balance, and even grow entirely new limbs. They will thereby create new godlings [i.e. ‘cyborgs’] who might be as different from us Sapiens as we are different from *Homo erectus*. (p. 50)

Harari (2017) continues:

This may sound like science fiction, but it’s already a reality. Monkeys have recently learned to control bionic hands and feet disconnected from their bodies, through electrodes implanted in their brains. Paralysed patients are able to move their bionic

4. That is, in a self-operative, spontaneous way. This phrase was often used in disputes between Catholics and Reformers when disagreeing about the operations of the sacraments. Reformers used to accuse Catholics that they regard the sacraments to function *ex opere operato*, that is, by and of itself, and do not require faith (as the Reformers insist) to bring the working and effect of the sacraments to full fruition.

5. This term is used in inverted commas since the distinction, in current-day terms, is more accurately indicated as that between rich and poor countries. The fact that people in rich countries are characterised by their general possession of or access to symbols of material wealth such as a house, a car, a TV asset and a computer in itself is hardly necessarily indicative of ‘development’, unless the latter term is used exclusively in purely materialist-economic terms.

limbs or operate computers by the power of thought alone. If you wish, you can already remote control electric devices in your house using an electric 'mind-reading' helmet. (p. 51)

All of this may seem very strange and fiction-like; but in every one of these (seemingly) 'extreme' examples that he discusses, the undoubted benefits are quite apparent. A defining argument in this respect is: Where has it ever been declaimed or written in judicial or moral law that our evolution as humans has to settle for the format of the bodies that we currently have? Our bodies as we know them are the outcome of evolution. Now we have reached the stage, as is patently illustrated by the cyborg examples, where we have started to take our future evolution into our own hands.

Of course, the latter is a high-risk operation that might yield formations that are unnecessary and sometimes dangerous. We must, however, avoid the serious mistake of thinking that because something can be brought about, it should also be brought about. When Mallory was asked why he wishes to reach Everest's summit, his famous answer was: 'because it's there'. There is no definitive reason why that should be our attitude when assessing the desirability of the possible ensuing revolutions in info-tech and biotech.

Conclusion: Three concepts

This section introduces the final arguments of this article. We have dwelt extensively on the actual and potential possibilities that the technological revolutions of the next century may bring. It is clear that the building blocks of this new era are going to differ dramatically from the way our world has been constituted up to the present time. However, in the argument up till this point, three key concepts have not yet been adequately mentioned – concepts, the understanding of which will be key to the way in which we eventually embark on the building of the future. These concepts are 'morality', 'responsibility' and 'regulation'. They are the real trumps that we hold in our efforts to, analogously to the card game from which the notion of 'trump' emerged (nothing to do with the world's current most powerful man!), try to build a bridge – not to an optimal contract with some individual partner, but to a future that will be inhabitable and shared by the rest of humanity.

Morality, in simple words, is a remarkable creation and inheritance of our species as is science, technology, medicine and literature. Harari is correct in his analysis that morality has for way too long been driven by religion. In the latter sense, religion was instrumental in conjuring up the mythical stories or grand narratives by means of which people were originally mobilised into action. However, Harari is also right in his assessment that those days are irrevocably over.

What is morality? It is the universal acknowledgement that all our actions can and ought to be submitted to the set of expectations known as the demand of obligation⁶. A current culture that is oblivious to the demand of obligation on their

6. Afrikaans has a better word for this, namely 'behorenses'.

behaviour is something that does not exist. We cannot simply do what we want. This applies as much to our behaviour as fathers, mothers, children, students, lawyers and teachers as it applies to the technologies that we create as scientists, engineers and technologists. The law of obligation that morality confers on us all is the recognition that, in the words of the Afrikaans poet DJ Opperman, 'an eternity borders on our deeds'⁷, that is, whatever we do have possible universal consequences and will necessarily be evaluated by others.

If morality is then not derivable from religions, how did it come about and from what does it derive its authority? The most persuasive answer to this question is to argue that society's morality is essentially the outcome of a historically achieved, broad consensus about values and norms to which right and good actions must conform. The fact that this value constellation derives from an achieved consensus is not to argue that this consensus is a timeless and unchangeable set of laws. It is also persuasive to think of society's set of values as a series of progressive settlements that occurs in the course of history. *Magna Carta* reflects something of it. So does the French Declaration of the Rights of Man. A dramatic and highly efficacious example is the American Declaration of Independence and Bill of Rights in the late 18th century. In the 20th century, we have seen both the United Nations' Declaration of Human Rights and the South African Bill of Rights and Constitution. Many could be added. Two insights are pertinent in this regard. The first is to note that morality as concept refers to both the documents that inform moral behaviour, and the behavioural practice inspired by these consensus documents. The second is to realise that these documents are not infinitely valid. They express an achieved consensus that came out of tumultuous historical events, and they are, every time, an expression to the measure of moral growth that their times testify to.

Ethics ought to be distinguished from morality. Ethics is the regular reflection about the nature of morality. Ethics, in order words, is a more intellectual enterprise by means of which we contemplate or reflect about the validity of a value constellation or moral consensus. Central to the argument in this article is the claim that morality and ethics need to remain central 'building blocks of the future', particularly in view of the potential and actual developments discussed in this article. Under no circumstances may the notion be tolerated that the new technologies have or are making morality and ethics obsolete. Much of the concerns provoked by the discussion in this article can and will be adequately addressed if a moral stance about the developments is maintained at all costs. This moral orientation must from the outset accompany all processes of scientific research and technological developments. A moral assessment ought to remain an integral part of the research process, and under no circumstances a mere 'add-on' that is (sometimes inadvertently) raised when the scientific work and technological applications are done and dusted.

7. The Afrikaans phrase is: 'Aan jou dade grens 'n ewigheid', from the poem 'Negester en Stedelig' in the poetry volume *Negester oor Niveé*.

The question may well be (and is!) asked as to whether morality is reserved for the sphere of conventional human affairs, and whether a 'morality algorithm' or some semblance of a function of moral choice cannot also be infused in AI. The (now) classic example (a variation of philosophy's classic 'trolley problem') is the self-driving car and the question as to how it should be programmed if confronted with one of two options: (1) swinging on to the curb and killing three pedestrians in order to save the one passenger in the car's life, or (2) not swinging away but keeping to the road and thereby precipitate a head-on collision that will kill the passenger in the car. The argument in this respect is that an algorithm can and must be programmed into this vehicle which will result in a moral decision. The implication is that morality will in future not be the exclusive domain of organic human life. (Legal issues are also relevant here; who is, after all, legally liable in case an accident of this nature occurs? Is it the manufacturer of the car, the creator of the algorithm, or even the AI itself? If the latter, how in practice can it be held liable?)

This brings us to the second of the three concepts to be introduced in conclusion; that is 'regulation'. AI is changing the world at a rapid pace, and we do not nearly have a clear idea as to what possible moral and legal problems the application of AI is bound to confront us with. On the other hand, we now live in a world where moral and legal accountability is by and large (though not completely) clear. We therefore have to regulate the outcomes of these actual and potential technologies as far as possible. Biomedical and info-technological ethics are not absent; a large literature and a world-wide conversation on these issues are available to us. It is because of fairly universally acknowledged regulation that, for example, reproductive cloning has been outlawed in most countries, in spite of the technology being available for more than 20 years. The moral and legal implications and problems of AI need to be revealed as widely as possible, and an inter-disciplinary conversation ought to be brought about to propose relevant regulations. It can never be accepted that simply because something can be done, it ought to be done or it is in order if it is done. Our control of our lifeworld and environment may be threatened by aspects of applied AI. This, however, does not mean that we are helpless and unable to curb excesses that threaten our humanity. Whereas it is true that 'human nature' as we know it is not an absolutely essential or unchangeable desideratum, significant deviations from our broad consensus on human nature in the direction of beings that have the potential to challenge our way of life and our freedom as human beings could and should be regulated before things run out of hand. In conclusion, a short word about the final concept to be added to the argument in this article. This is the concept of 'responsibility'. Responsibility, it could be argued, is by far the most important category in the field of morality and ethics (see Van Niekerk & Nortjé 2013). What has been argued above about the centrality of morality, ethics and regulatory practices in the process of building the future in the 21st century, in the

end boils down to humanity's universal acceptance of responsibility for what is done with the new technologies.

Responsibility simply amounts to the ability and willingness to provide communicable and persuasive reasons for action taken. Responsible decisions cannot infallibly claim that they will always be correct; fallibilism is part and parcel of an ethics of responsibility. The latter, however, acknowledges that a point in moral disputes is often reached where the luxury of further observation and reflection is no longer available, and some decision has to be reached. The decision might turn out to be unsatisfactory. However, what cannot then be doubted is that the motivation of those decisions has been thought through and has been deemed possible in the light that was available to the moral actor(s) at the time.

An ethics of responsibility is, furthermore, not only concerned with the present, but also with the future. It considers the interests, not only of us living here and now, but also of those who will come and live after us. As was claimed in the beginning of this article, at no stage in our history as humans has the future become as important to us as at the moment. It therefore is fitting that the ethics that will drive our decision-making will accordingly be facing up to and will be dealing with the demands of future generations.

Finally, as was argued by the great Greek ethicist Aristotle who still exerts so much influence on moral reflection in our time, an ethics of responsibility is driven by practical wisdom – which Aristotle called 'phronesis'. It is an approach which acknowledges the practical nature of ethics, that is, the demand for decisions about practical issues that have to be made on the basis of applying valid moral precepts or rules. *Phronesis* is a to-and-fro movement (in philosophical parlance: a 'dialectic') between the acknowledgement and validity of moral action guides on the one hand, and the demands of a practice that can stand the test of a meaningful life, on the other. This is, in the end, the great challenge for our future in the 21st century, particularly against the backdrop of the new, revolutionary and for a long time unforeseen building blocks that we have to build such a future.

Acknowledgements

The author would like to thank his research assistant, Sarah Gardner, for much help in finding and summarising the resources, writing the first draft of the abstract and carefully proof-reading and editing the manuscript.

Competing interests

The author declares that he has no financial or personal relationships that may have inappropriately influenced him in writing this research article.

Author's contribution

A.A.v.N. is the sole author of this research article.

Ethical considerations

This article followed all ethical standards for research without direct contact with human or animal subjects.

Funding information

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Disclaimer

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any affiliated agency of the author.

References

Butterfield, H., 1932, *The Whig interpretation of history*, G. Bell, London.

De Grey, A., 2008, *Ending aging*, St Martin's Press, New York, NY.

Gadamer, H.-G., 1975, *Truth and method*, Sheed and Ward, London.

Harari, Y.N., 2011, *Sapiens: A brief history of humankind*, Vintage, London.

Harari, Y.N., 2017, *Homo Deus: A brief history of tomorrow*, Vintage, London.

Harris, J., 1998, *Clones, genes and immortality: Ethics and the genetic revolution*, Oxford University Press, Oxford.

Hegel, G.W.F., 1991, *The philosophy of history*, transl. J. Sibree, Prometheus Books, New York, NY.

Kurzweil, R., 2006, *The singularity is near: When human transcend biology*, Penguin, London.

Kwant, R.C., 1964, *Mens en kritiek*, Het Spectrum, Utrecht/Antwerp.

Marx, K. & Engels, F., 1975, *The communist manifesto*, Collected Works, Vol. 6, Lawrence and Wishart, London.

Persson, I. & Savulescu, J., 2011, 'Unfit for the future? Human nature, scientific progress and the need for moral enhancement', in J. Savulescu, R. Ter Meulen & G. Kahane (eds.), *Enhancing human capacities*, Wiley-Blackwell, Oxford.

Popper, K.R., 1957, *The poverty of historicism*, Routledge & Kegan Paul, London.

Popper, K.R., 1972, *Objective knowledge: An evolutionary approach*, Clarendon Press, Oxford.

Shapin, S., 2017, 'The superhuman upgrade', *London Review of Books*, viewed 05 March 2020, from <https://www.lrb.co.uk/the-paper/v39/n14/steven-shapin/the-superhuman-upgrade>.

Toffler, A., 1970, *Future shock*, Pan Books, London.

Van Niekerk, A.A., 2014, 'Biomedical enhancement and the pursuit of mastery and perfection: A critique of the views of Michael Sandel', *South African Journal of Philosophy* 33(2), 155–165. <https://doi.org/10.1080/02580136.2014.923697>

Van Niekerk, A.A. & Nortjé, N., 2013, 'Phronesis and an ethics of responsibility', *South African Journal of Bioethics and Law* 6(1), 28–31. <https://doi.org/10.7196/sajbl.262>