LIBERTY

Yuval Noah Harari

"Big Data is watching you"

The liberal story cherishes human liberty as its number one value. It argues that all authority ultimately stems from the free will of individual humans, as it is expressed in their feelings, desires and choices. In politics, liberalism believes that the voter knows best. It therefore upholds democratic elections. In economics, liberalism maintains that the customer is always right. It therefore hails free-market principles. In personal matters, liberalism encourages people to listen to themselves, be true to themselves, and follow their hearts – as long as they do not infringe on the liberties of others. This personal freedom is enshrined in human rights.

In Western political discourse the term 'liberal' is sometimes used today in a much narrower partisan sense, to denote those who support specific causes like gay marriage, gun control and abortion. Yet most so-called conservatives also embrace the broad liberal world view. Especially in the United States, both Republicans and Democrats should occasionally take a break from their heated quarrels to remind themselves that they all agree on fundamentals such as free elections, an independent judiciary, and human rights.

In particular, it is vital to remember that right-wing heroes such as RonaldReagan and Margaret Thatcher were great champions not only of economic freedoms but also of individual liberties. In a famous interview in 1987, Thatcher said that 'There is no such thing as society. There is [a] living tapestry of men and women … and the quality of our lives will depend upon how much each of us is prepared to take responsibility for ourselves.

'1 Thatcher's heirs in the Conservative Party fully agree with the Labour Party that political authority comes from the feelings, choices and free will of individual voters. Thus when Britain needed to decide whether it should leave the EU, Prime Minister David Cameron didn't ask Queen Elizabeth II, the Archbishop of Canterbury, or the Oxford and Cambridge dons to resolve the issue. He didn't even ask the Members of Parliament. Rather, he held areferendum in which each and every Briton was asked: 'What do you feel about it?'

You might object that people were asked 'What do you think?' rather than 'What do you feel?', but this is a common misperception. Referendums and elections are always about human feelings, not about human rationality. If democracy were a matter of rational decision-making, there would be absolutely no reason to give all people equal voting rights – or perhaps any voting rights. There is ample evidence that some people are far more knowledgeable and rational than others, certainly when it comes to specific economic and political questions.2 In the wake of the Brexit vote, eminent biologist Richard Dawkins protested that the vast majority of the British public – including himself – should never have been asked to vote in the referendum, because they lacked the necessary background in economics and political science. 'You might as well call a nationwide plebiscite to decide whether Einstein got his algebra right, or let passengers vote on which runway the pilot should land.'3

However, for better or worse, elections and referendums are not about what we think. They are about what we feel. And when it comes to feelings, Einstein and Dawkins are no better than anyone else. Democracy assumes that human feelings reflect a mysterious and profound 'free will', that this 'free will' is the ultimate source of authority, and that while some people are more intelligent than others, all humans are equally free. Like Einstein and Dawkins, an illiterate maid also has free will, hence on election day her feelings – represented by her vote – count just as

much as anybody else's.

Feelings guide not just the voters, but also the leaders. In the 2016 Brexit referendum the Leave campaign was headed together by Boris Johnson and Michael Gove. After David Cameron resigned, Gove initially supported Johnson for the premiership, but at the very last minute Gove declared Johnson unfit for the position and announced his own intention to run for the job. Gove's action, which destroyed Johnson's chances, was described as a Machiavellian political assassination.4 But Gove defended his conduct by appealing to his feelings, explaining that 'In every step in my political life I have asked myself one question: "What is the right thing to do? What does your heart tell you?"'5That's why, according to Gove, he has fought so hard for Brexit, and that's why he felt compelled to backstab his erstwhile ally Boris Johnson and bid for the alpha-dog position himself – because his heart told him to do it.

This reliance on the heart might prove to be the Achilles heel of liberal democracy. For once somebody (whether in Beijing or in San Francisco) gains the technological ability to hack and manipulate the human heart, democratic politics will mutate into an emotional puppet show.

"Listen To the Algorithm"

The liberal belief in the feelings and free choices of individuals is neither natural nor very ancient. For thousands of years people believed that authority came from divine laws rather than from the human heart and that we should therefore sanctify the word of God rather than human liberty. Only in the last few centuries did the source of authority shift from celestial deities to flesh-and blood humans.

Soon authority might shift again – from humans to algorithms. Just as divine authority was legitimised by religious mythologies, and human authority was justified by the liberal story, so the coming technological revolution might establish the authority of Big Data algorithms, while undermining the very idea of individual freedom.

As we mentioned in the previous chapter, scientific insights into the way ourbrains and bodies work suggest that our feelings are not some uniquely human spiritual quality, and they do not reflect any kind of 'free will'. Rather, feelings are biochemical mechanisms that all mammals and birds use in order to quickly calculate probabilities of survival and reproduction. Feelings aren't based on intuition, inspiration or freedom – they are based on calculation.

When a monkey, mouse or human sees a snake, fear arises because millions of neurons in the brain swiftly calculate the relevant data and conclude that the probability of death is high. Feelings of sexual attraction arise when other biochemical algorithms calculate that a nearby individual offers a high probability of successful mating, social bonding, or some other coveted goal. Moral feelings such as outrage, guilt or forgiveness derive from neural mechanisms that evolved to enable group cooperation. All these biochemical algorithms were honed through millions of years of evolution. If the feelings of some ancient ancestor made a mistake, the genes shaping these feelings did not pass on to the next generation. Feelings are thus not the opposite of rationality – they embody evolutionary rationality.

We usually fail to realise that feelings are in fact calculations, because the rapid process of calculation occurs far below our threshold of awareness. We don't feel the millions of neurons in the brain computing probabilities of survival and reproduction, so we erroneously believe that our fear of snakes, our choice of sexual mates, or our opinions about the European Union are the result of some mysterious 'free will'.

Nevertheless, though liberalism is wrong to think that our feelings reflect a free will, up until

today relying on feelings still made good practical sense. For although there was nothing magical or free about our feelings, they were the best method in the universe for deciding what to study, who to marry, and which party to vote for. And no outside system could hope to understand my feelings better than me. Even if the Spanish Inquisition or the Soviet KGB spied on me every minute of every day, they lacked the biological knowledge and the computing power necessary to hack the biochemical processes shaping my desires and choices. For all practical purposes, it was reasonable to argue that I have free will, because my will was shaped mainly by the interplay of inner forces, which nobody outside could see. I could enjoy the illusion that I control my secret inner arena, while outsiders could never really understand what is happening inside me and how I make decisions.

Accordingly, liberalism was correct in counselling people to follow their heartrather than the dictates of some priest or party apparatchik. However, soon computer algorithms could give you better counsel than human feelings. As the Spanish Inquisition and the KGB give way to Google and Baidu, 'free will' will likely be exposed as a myth, and liberalism might lose its practical advantages.

For we are now at the confluence of two immense revolutions. On the onehand biologists are deciphering the mysteries of the human body, and in particular, of the brain and of human feelings. At the same time computer scientists are giving us unprecedented data-processing power. When the biotech revolution merges with the infotech revolution, it will produce Big Data algorithms that can monitor and understand my feelings much better than I can, and then authority will probably shift from humans to computers. My illusion of free will is likely to disintegrate as I daily encounter institutions, corporations and government agencies that understand and manipulate what was hitherto my inaccessible inner realm.

This is already happening in the field of medicine. The most importantmedical decisions in our life rely not on our feelings of illness or wellness, or even on the informed predictions of our doctor – but on the calculations of computers which understand our bodies much better than we do. Within a few decades, Big Data algorithms informed by a constant stream of biometric data could monitor our health 24/7. They could detect the very beginning of influenza, cancer or Alzheimer's disease, long before we feel anything is wrong with us. They could then recommend appropriate treatments, diets and daily regimens, custom-built for our unique physique, DNA and personality.

People will enjoy the best healthcare in history, but for precisely this reasonthey will probably be sick all the time. There is always something wrong somewhere in the body. There is always something that can be improved. In the past, you felt perfectly healthy as long as you didn't sense pain or you didn't suffer from an apparent disability such as limping. But by 2050, thanks to biometric sensors and Big Data algorithms, diseases may be diagnosed and treated long before they lead to pain or disability. As a result, you will always find yourself suffering from some 'medical condition' and following this or that algorithmic recommendation. If you refuse, perhaps your medical insurance would become invalid, or your boss would fire you – why should they pay the price of your obstinacy?

It is one thing to continue smoking despite general statistics that connectsmoking with lung cancer. It is a very different thing to continue smoking despite a concrete warning from a biometric sensor that has just detected seventeen cancerous cells in your upper left lung. And if you are willing to defy the sensor, what will you do when the sensor forwards the warning to

your insurance agency, your manager, and your mother?

Who will have the time and energy to deal with all these illnesses? In alllikelihood, we could just instruct our health algorithm to deal with most of these problems as it sees fit. At most, it will send periodic updates to our smartphones, telling us that 'seventeen cancerous cells were detected and destroyed'. Hypochondriacs might dutifully read these updates, but most of us will ignore them just as we ignore those annoying anti-virus notices on our computers.

"The drama of decision-making"

What is already beginning to happen in medicine is likely to occur in more and more fields. The key invention is the biometric sensor, which people can wear on or inside their bodies, and which converts biological processes into electronic information that computers can store and analyse. Given enough biometric data and enough computing power, external data-processing systems can hack all your desires, decisions and opinions. They can know exactly who you are.

Most people don't know themselves very well. When I was twenty-one, I finally realised that I was gay, after several years of living in denial. That's hardly exceptional. Many gay men spend their entire teenage years unsure about their sexuality. Now imagine the situation in 2050, when an algorithm can tell any teenager exactly where he is on the gay/straight spectrum (and even how malleable that position is). Perhaps the algorithm shows you pictures or videos of attractive men and women, tracks your eye movements, blood pressure and brain activity, and within five minutes ejects a number on the Kinsey scale.6 It could have saved me years of frustration. Perhaps you personally wouldn't want to take such a test, but then maybe you find yourself with a group of friends at Michelle's boring birthday party, and somebody suggests you all take turns checking yourself on this cool new algorithm (with everybody standing around to watch the results – and comment on them). Would you just walk away?

Even if you do, and even if you keep hiding from yourself and yourclassmates, you won't be able to hide from Amazon, Alibaba or the secret police. As you surf the Web, watch YouTube or read your social media feed, the algorithms will discreetly monitor you, analyse you, and tell Coca-Cola that if it wants to sell you some fizzy drink, it had better use the advertisement with the shirtless guy rather than the shirtless girl. You won't even know. But they will know, and such information will be worth billions.

Then again, maybe it will all be out in the open, and people will gladly share their information in order to get better recommendations – and eventually in order to get the algorithm to make decisions for them. It starts with simple things, like deciding which movie to watch. As you sit down with a group of friends to spend a cozy evening in front of the TV, you first have to choose what to see. Fifty years ago you had no choice, but today – with the rise of view-ondemand services – there are thousands of titles available. Reaching an agreement can be quite difficult, because while you personally like science-fiction thrillers, Jack prefers romantic comedies, and Jill votes for artsy French films. You may well end up compromising on some mediocre B-movie that disappoints all of you.

An algorithm might help. You can tell it which previous movies each of youreally liked, and based on its massive statistical database, the algorithm can then find the perfect match for the group. Unfortunately, such a crude algorithm is easily misled, particularly because self-reporting is a notoriously unreliable gauge for people's true preferences. It often happens that we hear lots of people praise some movie as a masterpiece, feel compelled to watch it, and even though we fall asleep midway through, we don't want to look like philistines, so we tell everyone it was

an amazing experience.

Such problems, however, can be solved if we just allow the algorithm tocollect real-time data on us as we actually watch movies, instead of relying on our own dubious self-reports. For starters, the algorithm can monitor which movies we completed, and which we stopped watching halfway through. Even if we tell the whole world that Gone with the Wind is the best movie ever made, the algorithm will know we never made it past the first half-hour, and we never really saw Atlanta burning.

Yet the algorithm can go much deeper than that. Engineers are currentlydeveloping software that can detect human emotions based on the movements' ofour eyes and facial muscles.8 Add a good camera to the television, and such software will know which scenes made us laugh, which scenes made us sad, and which scenes bored us. Next, connect the algorithm to biometric sensors, and the algorithm will know how each frame has influenced our heart rate, our blood pressure, and our brain activity. As we watch, say, Tarantino's Pulp Fiction, the algorithm may note that the rape scene caused us an almost imperceptible tinge of sexual arousal, that when Vincent accidentally shot Marvin in the face it made us laugh guiltily, and that we didn't get the joke about the Big Kahuna Burger – but we laughed anyway, so as not to look stupid. When you force yourself to laugh, you use different brain circuits and muscles than when you laugh because something is really funny. Humans cannot usually detect the difference. But a biometric sensor could.

The word television comes from Greek 'tele', which means 'far', and Latin 'visio', sight. It was originally conceived as a device that allows us to see from afar. But soon, it might allow us to be seen from afar. As George Orwell envisioned in Nineteen Eighty-Four, the television will watch us while we are watching it. After we've finished watching Tarantino's entire filmography, we may have forgotten most of it. But Netflix, or Amazon, or whoever owns the TV algorithm, will know our personality type, and how to press our emotional buttons. Such data could enable Netflix and Amazon to choose movies for us with uncanny precision, but it could also enable them to make for us the most important decisions in life – such as what to study, where to work, and who to marry.

Of course Amazon won't be correct all the time. That's impossible. Algorithms will repeatedly make mistakes due to insufficient data, faulty programming, muddled goal definitions and the chaotic nature of life. But Amazon won't have to be perfect. It will just need to be better on average than us humans. And that is not so difficult, because most people don't know themselves very well, and most people often make terrible mistakes in the most important decisions of their lives. Even more than algorithms, humans suffer from insufficient data, from faulty programming (genetic and cultural), from muddled definitions, and from the chaos of life.

You may well list the many problems that beset algorithms, and conclude that people will never trust them. But this is a bit like cataloguing all the drawbacks of democracy and concluding that no sane person would ever choose to support such a system. Winston Churchill famously said that democracy is the worst political system in the world, except for all the others. Rightly or wrongly, people might reach the same conclusions about Big Data algorithms: they have lots of hitches, but we have no better alternative.

As scientists gain a deeper understanding of the way humans make decisions, the temptation to rely on algorithms is likely to increase. Hacking human decision-making will not only make Big Data algorithms more reliable, it will simultaneously make human feelings less reliable. As

governments and corporations succeed in hacking the human operating system, we will be exposed to a barrage of precision-guided manipulation, advertisement and propaganda. It might become so easy to manipulate our opinions and emotions that we will be forced to rely on algorithms in the same way that a pilot suffering an attack of vertigo must ignore what his own senses are telling him and put all his trust in the machinery.

In some countries and in some situations, people might not be given any choice, and they will be forced to obey the decisions of Big Data algorithms. Yet even in allegedly free societies, algorithms might gain authority because we will learn from experience to trust them on more and more issues, and will gradually lose our ability to make decisions for ourselves. Just think of the way that within a mere two decades, billions of people have come to entrust the Google search algorithm with one of the most important tasks of all: searching for relevant and trustworthy information. We no longer search for information. Instead, we google. And as we increasingly rely on Google for answers, so our ability to search for information by ourselves diminishes. Already today, 'truth' is defined by the top results of the Google search.

This has also been happening with physical abilities, such as navigating space.People ask Google to guide them around. When they reach an intersection, their gut feeling might tell them 'turn left', but Google Maps says 'turn right'. At first they listen to their gut feeling, turn left, get stuck in a traffic jam, and miss an important meeting. Next time they listen to Google, turn right, and make it on time. They learn from experience to trust Google. Within a year or two, they blindly rely on whatever Google Maps tells them, and if the smartphone fails, they are completely clueless. In March 2012 three Japanese tourists in Australia decided to take a day trip to a small offshore island, and drove their car straight into the Pacific Ocean. The driver, twenty-one-year-old Yuzu Nuda, later said that she just followed the instructions of the GPS and 'it told us we could drive down there. It kept saying it would navigate us to a road. We got stuck.'12In several similar incidents people drove into a lake, or fell off a demolished bridge, by apparently following GPS instructions.The ability to navigate is like a muscle – use it or lose it.14 The same is true for the ability to choose spouses or professions.

Every year millions of youngsters need to decide what to study at university.This is a very important and very difficult decision. You are under pressure from your parents, your friends and your teachers, who have different interests and opinions. You also have your own fears and fantasies to deal with. Your judgement is clouded and manipulated by Hollywood blockbusters, trashy novels, and sophisticated advertising campaigns. It is particularly difficult to make a wise decision because you do not really know what it takes to succeed in different professions, and you don't necessarily have a realistic image of your own strengths and weaknesses. What does it take to succeed as a lawyer? How do I perform under pressure? Am I a good team-worker?

One student might start law school because she has an inaccurate image of her own skills, and an even more distorted view of what being a lawyer actually involves (you don't get to give dramatic speeches and shout 'Objection, Your Honour!' all day). Meanwhile her friend decides to fulfil a childhood dream and study professional ballet dancing, even though she doesn't have the necessary bone structure or discipline. Years later, both deeply regret their choices. In the future we could rely on Google to make such decisions for us. Google could tell me that I would be wasting my time in law school or in ballet school – but that I might make an excellent (and very happy) psychologist or plumber.15

Once AI makes better decisions than us about careers and perhaps even relationships, our

concept of humanity and of life will have to change. Humans are used to thinking about life as a drama of decision-making. Liberal democracy and free-market capitalism see the individual as an autonomous agent constantly making choices about the world. Works of art – be they Shakespeare plays, Jane Austen novels, or tacky Hollywood comedies – usually revolve around the hero having to make some particularly crucial decision. To be or not to be? To listen to my wife and kill, King Duncan or listen to my conscience and spare him? To marry Mr Collins or Mr Darcy? Christian and Muslim theology similarly focus on the drama of decision-making, arguing that everlasting salvation or damnation depends on making the right choice.

What will happen to this view of life as we increasingly rely on AI to make decisions for us? At present we trust Netflix to recommend movies, and Google Maps to choose whether to turn right or left. But once we begin to count on AI to decide what to study, where to work, and who to marry, human life will cease to be a drama of decision-making. Democratic elections and free markets will make little sense. So would most religions and works of art. Imagine Anna Karenina taking out her smartphone and asking the Facebook algorithm whether she should stay married to Karenin or elope with the dashing Count Vronsky. Or imagine your favourite Shakespeare play with all the crucial decisions taken by the Google algorithm. Hamlet and Macbeth will have much more comfortable lives, but what kind of life will it be exactly? Do we have models for making sense of such a life?

As authority shifts from humans to algorithms, we may no longer see theworld as the playground of autonomous individuals struggling to make the right choices. Instead, we might perceive the entire universe as a flow of data, see organisms as little more than biochemical algorithms, and believe that humanity's cosmic vocation is to create an all-encompassing data-processing system – and then merge into it. Already today we are becoming tiny chips inside a giant data-processing system that nobody really understands. Every day I absorb countless data bits through emails, tweets and articles; process the data; and transmit back new bits through more emails, tweets and articles. I don't really know where I fit into the great scheme of things, and how my bits of data connect with the bits produced by billions of other humans and computers. I don't have time to find out, because I am too busy answering all these emails.

"The Philosophical Car"

People might object that algorithms could never make important decisions for us, because important decisions usually involve an ethical dimension, and algorithms don't understand ethics. Yet there is no reason to assume that algorithms won't be able to outperform the average human even in ethics. Already today, as devices like smartphones and autonomous vehicles undertake decisions that used to be a human monopoly, they start to grapple with the same kind of ethical problems that have bedeviled humans for millennia.

For example, suppose two kids chasing a ball jump right in front of a self driving car. Based on its lightning calculations, the algorithm driving the car concludes that the only way to avoid hitting the two kids is to swerve into the opposite lane, and risk colliding with an oncoming truck. The algorithm calculates that in such a case there is a 70 per cent chance that the owner of the car – who is fast asleep in the back seat – would be killed. What should the algorithm do?

Philosophers have been arguing about such 'trolley problems' for millennia (they are called 'trolley problems' because the textbook examples in modern philosophical debates refer to a runaway trolley car racing down a railway track, rather than to a self-driving car).Up till now, these arguments have had embarrassingly little impact on actual behaviour, because in times of

crisis humans all too often forget about their philosophical views and follow their emotions and gut instincts instead.

One of the nastiest experiments in the history of the social sciences was conducted in December 1970 on a group of students at the Princeton Theological Seminary, who were training to become ministers in the Presbyterian Church. Each student was asked to hurry to a distant lecture hall, and there give a talk on the Good Samaritan parable, which tells how a Jew travelling from Jerusalem to Jericho was robbed and beaten by criminals, who then left him to die by the side of the road. After some time a priest and a Levite passed nearby, but both ignored the man. In contrast, a Samaritan – a member of a sect much despised by the Jews – stopped when he saw the victim, took care of him, and saved his life. The moral of the parable is that people's merit should be judged by their actual behaviour, rather than by their religious affiliation.

The eager young seminarians rushed to the lecture hall, contemplating on the way how best to explain the moral of the Good Samaritan parable. But the experimenters planted in their path a shabbily dressed person, who was sitting slumped in a doorway with his head down and his eyes closed. As each unsuspecting seminarian was hurrying past, the 'victim' coughed and groaned pitifully. Most seminarians did not even stop to enquire what was wrong with the man, let alone offer any help. The emotional stress created by the need to hurry to the lecture hall trumped their moral obligation to help strangers in distress.

Human emotions trump philosophical theories in countless other situations. This makes the ethical and philosophical history of the world a rather depressing tale of wonderful ideals and less than ideal behaviour. How many Christians actually turn the other cheek, how many Buddhists actually rise above egoistic obsessions, and how many Jews actually love their neighbours as themselves? That's just the way natural selection has shaped Homo sapiens. Like all mammals, Homo sapiens use emotions to quickly make life and death decisions. We have inherited our anger, our fear and our lust from millions of ancestors, all of whom passed the most rigorous quality control tests of natural selection.

Unfortunately, what was good for survival and reproduction in the Africansavannah a million years ago does not necessarily make for responsible behaviour on twenty-first-century motorways. Distracted, angry and anxious human drivers kill more than a million people in traffic accidents every year. We can send all our philosophers, prophets and priests to preach ethics to these drivers – but on the road, mammalian emotions and savannah instincts will still take over. Consequently, seminarians in a rush will ignore people in distress, and drivers in a crisis will run over hapless pedestrians.

This disjunction between the seminary and the road is one of the biggest practical problems in ethics. Immanuel Kant, John Stuart Mill and John Rawls can sit in some cosy university hall and discuss theoretical problems in ethics for days – but would their conclusions actually be implemented by stressed-out drivers caught in a split-second emergency? Perhaps Michael Schumacher – the Formula One champion who is sometimes hailed as the best driver in history – had the ability to think about philosophy while racing a car; but most of us aren't Schumacher.

Computer algorithms, however, have not been shaped by natural selection,and they have neither emotions nor gut instincts. Hence in moments of crisis they could follow ethical guidelines much better than humans – provided we find a way to code ethics in precise numbers and statistics. If we teach Kant, Mill and Rawls to write code, they can carefully program the self-

driving car in their cosy laboratory, and be certain that the car will follow their commandments on the highway. In effect, every car will be driven by Michael Schumacher and Immanuel Kant rolled into one.

Thus if you program a self-driving car to stop and help strangers in distress, it will do so come hell or high water (unless, of course, you insert an exception clause for infernal or high-water scenarios). Similarly, if your self-driving car is programmed to swerve to the opposite lane in order to save the two kids in its path, you can bet your life this is exactly what it will do. Which means that when designing their self-driving car, Toyota or Tesla will be transforming a theoretical problem in the philosophy of ethics into a practical problem of engineering?

Granted, the philosophical algorithms will never be perfect. Mistakes will still happen, resulting in injuries, deaths and extremely complicated lawsuits. (For the first time in history, you might be able to sue a philosopher for the unfortunate results of his or her theories, because for the first time in history you could prove a direct causal link between philosophical ideas and real-life events.) However, in order to take over from human drivers, the algorithms won't have to be perfect. They will just have to be better than the humans. Given that human drivers kill more than a million people each year, that isn't such a tall order. When all is said and done, would you rather the car next to you was driven by a drunk teenager, or by the Schumacher–Kant team?

The same logic is true not just of driving, but of many other situations. Take for example job applications. In the twenty-first century, the decision whether to hire somebody for a job will increasingly be made by algorithms. We cannot rely on the machine to set the relevant ethical standards – humans will still need to do that. But once we decide on an ethical standard in the job market – that it is wrong to discriminate against black people or against women, for example – we can rely on machines to implement and maintain this standard better than humans.

A human manager may know and even agree that it is unethical to discriminate against black people and women, but then, when a black woman applies for a job, the manager subconsciously discriminates against her, and decides not to hire her. If we allow a computer to evaluate job applications, and program the computer to completely ignore race and gender, we can be certain that the computer will indeed ignore these factors, because computers don't have a subconscious. Of course, it won't be easy to write code for evaluating job applications, and there is always a danger that the engineers will somehow program their own subconscious biases into the software.Yet once we discover such mistakes, it would probably be far easier to debug the software than to rid humans of their racist and misogynist biases.

We saw that the rise of artificial intelligence might push most humans out of the job market – including drivers and traffic police (when rowdy humans are replaced by obedient algorithms, traffic police will be redundant). However, there might be some new openings for philosophers, because their skills – hitherto devoid of much market value – will suddenly be in very high demand. So if you want to study something that will guarantee a good job in the future, maybe philosophy is not such a bad gamble.

Of course, philosophers seldom agree on the right course of action. Few 'trolley problems' have been solved to the satisfaction of all philosophers, and consequentialist thinkers such as John Stuart Mill (who judge actions by consequences) hold quite different opinions to deontologists such as Immanuel Kant (who judge actions by absolute rules). Would Tesla have to actually take a stance on such knotty matters in order to produce a car?

Well, maybe Tesla will just leave it to the market. Tesla will produce two models of the self-driving car: the Tesla Altruist and the Tesla Egoist. In an emergency, the Altruist sacrifices its owner to the greater good, whereas the Egoist does everything in its power to save its owner, even if it means killing the two kids. Customers will then be able to buy the car that best fits their favourite philosophical view. If more people buy the Tesla Egoist, you won't be able to blame Tesla for that. After all, the customer is always right.

This is not a joke. In a pioneering 2015 study people were presented with a hypothetical scenario of a self-driving car about to run over several pedestrians. Most said that in such a case the car should save the pedestrians even at the price of killing its owner. When they were then asked whether they personally would buy a car programmed to sacrifice its owner for the greater good, most said no. For themselves, they would prefer the Tesla Egoist.

Imagine the situation: you have bought a new car, but before you can start using it, you must open the settings menu and tick one of several boxes. In case of an accident, do you want the car to sacrifice your life – or to kill the family in the other vehicle? Is this a choice you even want to make? Just think of the arguments you are going to have with your husband about which box to tick.

So maybe the state should intervene to regulate the market, and lay down an ethical code binding all self-driving cars? Some lawmakers will doubtless be thrilled by the opportunity to finally make laws that are always followed to the letter. Other lawmakers may be alarmed by such unprecedented and totalitarian responsibility. After all, throughout history the limitations of law enforcement provided a welcome check on the biases, mistakes and excesses of lawmakers. It was an extremely lucky thing that laws against homosexuality and against blasphemy were only partially enforced. Do we really want a system in which the decisions of fallible politicians become as inexorable as gravity?

"Digital dictatorships"

AI often frightens people because they don't trust the AI to remain obedient. We have seen too many science-fiction movies about robots rebelling against their human masters, running amok in the streets and slaughtering everyone. Yet the real problem with robots is exactly the opposite. We should fear them because they will probably always obey their masters and never rebel.

There is nothing wrong with blind obedience, of course, as long as the robots happen to serve benign masters. Even in warfare, reliance on killer robots could ensure that for the first time in history, the laws of war would actually be obeyed on the battlefield. Human soldiers are sometimes driven by their emotions to murder, pillage and rape in violation of the laws of war. We usually associate emotions with compassion, love and empathy, but in wartime, the emotions that take control are all too often fear, hatred and cruelty. Since robots have no emotions, they could be trusted to always adhere to the dry letter of the military code, and never be swayed by personal fears and hatreds.

On 16 March 1968 a company of American soldiers went berserk in the South Vietnamese village of My Lai, and massacred about 400 civilians. This war crime resulted from the local initiative of men who had been involved in jungle guerrilla warfare for several months. It did not serve any strategic purpose, and contravened both the legal code and the military policy of the USA. It was the fault of human emotions. If the USA had deployed killer robots in Vietnam, the massacre of My Lai would never have occurred.

Nevertheless, before we rush to develop and deploy killer robots, we need to remind ourselves that the robots always reflect and amplify the qualities of their code. If the code is restrained and benign – the robots will probably be a huge improvement over the average human soldier. Yet if the code is ruthless and cruel – the results will be catastrophic. The real problem with robots is not their own artificial intelligence, but rather the natural stupidity and cruelty of their human masters.

In July 1995 Bosnian Serb troops massacred more than 8,000 MuslimBosniaks around the town of Srebrenica. Unlike the haphazard My Lai massacre, the Srebrenica killings were a protracted and well-organised operation that reflected Bosnian Serb policy to 'ethnically cleanse' Bosnia of Muslims.If theBosnian Serbs had had killer robots in 1995, it would likely have made the atrocity worse rather than better. Not one robot would have had a moment's hesitation carrying out whatever orders it received, and would not have spared the life of a single Muslim child out of feelings of compassion, disgust, or mere lethargy.

A ruthless dictator armed with such killer robots will never have to fear that his soldiers will turn against him, no matter how heartless and crazy his orders. A robot army would probably have strangled the French Revolution in its cradle in 1789, and if in 2011 Hosni Mubarak had had a contingent of killer robots he could have unleashed them on the populace without fear of defection. Similarly, an imperialist government relying on a robot army could wage unpopular wars without any concern that its robots might lose their motivation, or that their families might stage protests. If the USA had had killer robots in the Vietnam War, the My Lai massacre might have been prevented, but the war itself could have dragged on for many more years, because the American government would have had fewer worries about demoralised soldiers, massive anti-war demonstrations, or a movement of 'veteran robots against the war' (some American citizens might still have objected to the war, but without the fear of being drafted themselves, the memory of personally committing atrocities, or the painful loss of a dear relative, the protesters would probably have been both less numerous and less committed).

These kinds of problems are far less relevant to autonomous civilian vehicles, because no car manufacturer will maliciously program its vehicles to target and kill people. Yet autonomous weapon systems are a catastrophe waiting to happen, because too many governments tend to be ethically corrupt, if not downright evil.

The danger Is not restricted to killing machines. Surveillance systems could be equally risky. In the hands of a benign government, powerful surveillance algorithms can be the best thing that ever happened to humankind. Yet the same Big Data algorithms might also empower a future Big Brother, so that we might end up with an Orwellian surveillance regime in which all individuals are monitored all the time.

Indeed, we might end up with something that even Orwell could barely imagine: a total surveillance regime that follows not just all our external activities and utterances, but can even go under our skin to observe our inner experiences. Consider for example what the Kim regime in North Korea might do with the new technology. In the future, each North Korean citizen might be required to wear a biometric bracelet that monitors everything you do and say – as well as your blood pressure and brain activity. By using our growing understanding of the human brain, and using the immense powers of machine learning, the North Korean regime might be able for the first time in history to gauge what each and every citizen is thinking each and every moment. If you look at a picture of Kim Jong-un and the biometric sensors pick up the telltale signs of

anger (higher blood pressure, increased activity in the amygdala) – you'll be in the Gulag tomorrow morning.

Granted, due to its isolation the North Korean regime might have difficulty developing the required technology by itself. However, the technology might be pioneered in more tech-savvy nations, and copied or bought by the North Koreans and other backward dictatorships. Both China and Russia are constantly improving their surveillance tools, as are a number of democratic countries, ranging from the USA to my home country of Israel. Nicknamed 'the start-up nation', Israel has an extremely vibrant hi-tech sector, and a cutting-edge cybersecurity industry. At the same time it is also locked into a deadly conflict with the Palestinians, and at least some of its leaders, generals and citizens might well be happy to create a total surveillance regime in the West Bank as soon as they have the necessary technology.

Already today whenever Palestinians make a phone call, post something onFacebook or travel from one city to another they are likely to be monitored by Israeli microphones, cameras, drones or spy software. The gathered data is then analysed with the aid of Big Data algorithms. This helps the Israeli security forces to pinpoint and neutralise potential threats without having to place too many boots on the ground. The Palestinians may administer some towns and villages in the West Bank, but the Israelis control the sky, the airwaves and cyberspace. It therefore takes surprisingly few Israeli soldiers to effectively control about 2.5 million Palestinians in the West Bank.

In one tragicomic incident in October 2017, a Palestinian labourer posted to his private Facebook account a picture of himself in his workplace, alongside a bulldozer. Adjacent to the image he wrote 'Good morning!' An automaticAlgorithm made a small error when transliterating the Arabic letters. Instead of 'Ysabechhum!' (which means 'Good morning!'), the algorithm identified the letters as 'Ydbachhum!' (which means 'Kill them!'). Suspecting that the man might be a terrorist intending to use a bulldozer to run people over, Israeli security forces swiftly arrested him. He was released after they realised that the algorithm made a mistake. But the offending Facebook post was nevertheless taken down. You can never be too careful.What Palestinians are experiencing today in the West Bank might be just a primitive preview to what billions will eventually experience all over the planet.

In the late twentieth century democracies usually outperformed dictatorships because democracies were better at data-processing. Democracy diffuses the power to process information and make decisions among many people and institutions, whereas dictatorship concentrates information and power in one place. Given twentieth-century technology, it was inefficient to concentrate too much information and power in one place. Nobody had the ability to process all the information fast enough and make the right decisions. This is part of the reason why the Soviet Union made far worse decisions than the United States, and why the Soviet economy lagged far behind the American economy.

However, soon AI might swing the pendulum in the opposite direction. AI makes it possible to process enormous amounts of information centrally. Indeed, AI might make centralised systems far more efficient than diffused systems, because machine learning works better the more information it can analyse. If you concentrate all the information relating to a billion people in one database, disregarding all privacy concerns, you can train much better algorithms than if you respect individual privacy and have in your database only partial information on a million people. For example, if an authoritarian government orders all its citizens to have their

DNA scanned and to share all their medical data with some central authority, it would gain an immense advantage in genetics and medical research over societies in which medical data is strictly private. The main handicap of authoritarian regimes in the twentieth century – the attempt to concentrate all information in one place – might become their decisive advantage in the twenty-first century.

As algorithms come to know us so well, authoritarian governments could gain absolute control over their citizens, even more so than in Nazi Germany, and resistance to such regimes might be utterly impossible. Not only will the regime know exactly how you feel – it could make you feel whatever it wants. The dictator might not be able to provide citizens with healthcare or equality, but he could make them love him and hate his opponents. Democracy in its present form cannot survive the merger of biotech and infotech. Either democracy will successfully reinvent itself in a radically new form, or humans will come to live in 'digital dictatorships'.

This will not be a return to the days of Hitler and Stalin. Digital dictatorships will be as different from Nazi Germany as Nazi Germany was different from ancien régime France. Louis XIV was a centralising autocrat, but he did not have the technology to build a modern totalitarian state. He suffered no opposition to his rule, yet in the absence of radios, telephones and trains, he had little control over the day-to-day lives of peasants in remote Breton villages, or even of townspeople in the heart of Paris. He had neither the will nor the ability to establish a mass party, a countrywide youth movement, or a national education system.30

It was the new technologies of the twentieth century that gave Hitler both the motivation and the power to do such things. We cannot predict what will be the motivations and powers of digital dictatorships in 2084, but it is very unlikely that they will just copy Hitler and Stalin. Those gearing themselves up to refight the battles of the 1930s might be caught off their guard by an attack from a totally different direction.

Even if democracy manages to adapt and survive, people might become the victims of new kinds of oppression and discrimination. Already today more and more banks, corporations and institutions are using algorithms to analyse data and make decisions about us. When you apply to your bank for a loan, it is likely that you're application is processed by an algorithm rather than by a human. The algorithm analyses lots of data about you and statistics about millions of other people, and decides whether you are reliable enough to give you a loan. Often, the algorithm does a better job than a human banker. But the problem is that if the algorithm discriminates against some people unjustly, it is difficult to know that. If the bank refuses to give you a loan, and you ask 'Why?' the bank replies 'The algorithm said no.' You ask 'Why did the algorithm say no? What's wrong with me?', and the bank replies 'We don't know. No human understands this algorithm, because it is based on advanced machine learning. But we trust our algorithm, so we won't give you a loan.

When discrimination is directed against entire groups, such as women or black people, these groups can organise and protest against their collective discrimination. But now an algorithm might discriminate against you personally, and you have no idea why. Maybe the algorithm found something in your DNA, your personal history or your Facebook account that it does not like. The algorithm discriminates against you not because you are a woman, or an African American – but because you are you. There is something specific about you that the algorithm does not like. You don't know what it is, and even if you knew, you cannot organise with other people to protest, because there are no other people suffering the exact same prejudice. It is

just you. Instead of just collective discrimination, in the twenty-first century we might face a growing problem of individual discrimination.

At the highest levels of authority, we will probably retain human figureheads, who will give us the illusion that the algorithms are only advisors, and that ultimate authority is still in human hands. We will not appoint an AI to be the chancellor of Germany or the CEO of Google. However, the decisions taken by the chancellor and the CEO will be shaped by AI. The chancellor could still choose between several different options, but all these options will be the outcome of Big Data analysis, and they will reflect the way AI views the world more than the way humans view it.

To take an analogous example, today politicians all over the world can choose between several different economic policies, but in almost all cases the various policies on offer reflect a capitalist outlook on economics. The politicians have an illusion of choice, but the really important decisions have already been made much earlier by the economists, bankers and business people who shaped the different options in the menu. Within a couple of decades, politicians might find themselves choosing from a menu written by AI.

"Artificial intelligence and natural stupidity"

One piece of good news is that at least in the next few decades, we won't have to deal with the full-blown science-fiction nightmare of AI gaining consciousness and deciding to enslave or wipe out humanity. We will increasingly rely on algorithms to make decisions for us, but it is unlikely that the algorithms will start to consciously manipulate us. They won't have any consciousness.

Science fiction tends to confuse intelligence with consciousness, and assume that in order to match or surpass human intelligence, computers will have to develop consciousness. The basic plot of almost all movies and novels about AI revolves around the magical moment when a computer or a robot gains consciousness. Once that happens, either the human hero falls in love with the robot, or the robot tries to kill all the humans, or both things happen simultaneously.

But in reality, there is no reason to assume that artificial intelligence will gain consciousness, because intelligence and consciousness are very different things. Intelligence is the ability to solve problems. Consciousness is the ability to feel things such as pain, joy, love and anger. We tend to confuse the two because in humans and other mammals intelligence goes hand in hand with consciousness. Mammals solve most problems by feeling things. Computers, however, solve problems in a very different way.

There are simply several different paths leading to high intelligence, and only some of these paths involve gaining consciousness. Just as airplanes fly faster than birds without ever developing feathers, so computers may come to solve problems much better than mammals without ever developing feelings. True, AI will have to analyse human feelings accurately in order to treat human illnesses, identify human terrorists, recommend human mates and navigate a street full of human pedestrians. But it could do so without having any feelings of its own. An algorithm does not need to feel joy, anger or fear in order to recognise the different biochemical patterns of joyful, angry or frightened apes.

Of course, it is not absolutely impossible that AI will develop feelings of its own. We still don't know enough about consciousness to be sure. In general, there are three possibilities we need to consider:

1. Consciousness is somehow linked to organic biochemistry in such a way that it will

never be possible to create consciousness in nonorganic systems.

2. Consciousness is not linked to organic biochemistry, but it is linked to intelligence in such a way that computers could develop consciousness, and computers will have to develop consciousness if they are to pass a certain threshold of intelligence.
3. There are no essential links between consciousness and either organic biochemistry or high intelligence. Hence computers might develop consciousness – but not necessarily. They could become super- intelligent while still having zero consciousness.

At our present state of knowledge, we cannot rule out any of these options. Yet precisely because we know so little about consciousness, it seems unlikely that we could program conscious computers any time soon. Hence despite the immense power of artificial intelligence, for the foreseeable future its usage will continue to depend to some extent on human consciousness.

The danger is that if we invest too much in developing AI and too little in developing human consciousness, the very sophisticated artificial intelligence of computers might only serve to empower the natural stupidity of humans. We are unlikely to face a robot rebellion in the coming decades, but we might have to deal with hordes of bots who know how to press our emotional buttons better than our mother, and use this uncanny ability to try and sell us something – be it a car, a politician, or an entire ideology. The bots could identify our deepest fears, hatreds and cravings, and use these inner leverages against us. We have already been given a foretaste of this in recent elections and referendums across the world, when hackers have learned how to manipulate individual voters by analysing data about them and exploiting their existing prejudices.WhileScience-fiction thrillers are drawn to dramatic apocalypses of fire and smoke, in reality we might be facing a banal apocalypse by clicking.

To avoid such outcomes, for every dollar and every minute we invest in improving artificial intelligence, it would be wise to invest a dollar and a minute in advancing human consciousness. Unfortunately, at present we are not doing much to research and develop human consciousness. We are researching and developing human abilities mainly according to the immediate needs of the economic and political system, rather than according to our own long-term needs as conscious beings. My boss wants me to answer emails as quickly as possible, but he has little interest in my ability to taste and appreciate the food I am eating. Consequently, I check my emails even during meals, while losing the ability to pay attention to my own sensations. The economic system pressures me to expand and diversify my investment portfolio, but it gives me zero incentives to expand and diversify my compassion. So I strive to understand the mysteries of the stock exchange, while making far less effort to understand the deep causes of suffering.

In this, humans are similar to other domesticated animals. We have bred docile cows that produce enormous amounts of milk, but are otherwise far inferior to their wild ancestors. They are less agile, less curious and less resourceful.We are now creating tame humans that produce enormous amounts of data and function as very efficient chips in a huge data-processing mechanism, but these data-cows hardly maximise the human potential. Indeed we have no idea what the full human potential is, because we know so little about the human mind. And yet we hardly invest much in exploring the human mind, and instead focus on increasing the speed of our Internet connections and the efficiency of our Big Data algorithms. If we are not careful, we will end up with downgraded humans misusing upgraded computers to wreak havoc on them

and on the world.

Digital dictatorships are not the only danger awaiting us. Alongside liberty, the liberal order has also set great store by the value of equality. Liberalism always cherished political equality, and it gradually came to realise that economic equality is almost as important. For without a social safety net and a modicum of economic equality, liberty is meaningless. But just as Big Data algorithms might extinguish liberty, they might simultaneously create the most unequal societies that ever existed. All wealth and power might be concentrated in the hands of Stiny elite, while most people will suffer not from exploitation, but from something far worse – irrelevance.