

Yuval Noah Harari 0:22

Thank you for this wonderful introduction. And yes, what I want to talk to you about is AI in the future of humanity. Now I know that these conferences focus on the ecological crisis facing humanity. But for better, or for worse, AI to be part of this crisis. AI can help us in many ways to overcome the ecological crisis, or it can make it far, far worse, actually.

Speaker 1 1:03

The ecological system of planet Earth contained only organic life forms and now we might see the emergence of the first inorganic life forms to 4 billion years, or at the very least, the emergence of inorganic agents. Now people have feared AI since the very beginning of the computer age in the middle of the 20th century. And this fear has inspired many science fiction classics, like the Terminator or the matrix. Now, while such science fiction scenarios have become cultural landmarks, they haven't usually been taken seriously in academic and scientific and political debates, and perhaps for a good reason. Because science fiction scenarios usually assume that before AI can pose a significant threat to humanity, it will have to reach or to pass two important milestones. First, AI will have to become sentient and develop consciousness, feelings, emotions. Otherwise, why would it even want to take over the world? Secondly, AI will have to become adapt at navigating the physical world. Robots will have to be able to move around and operate in houses and cities. And mountains and forests at least as dexterously and efficiently as humans. If they cannot move around the physical world, how can they possibly take it over? And as of April 2023, ai still seems far from reaching either of these milestones. Despite all the hype around check GPT and the other new AI tools, there is no evidence that these tools have even a shred of consciousness, feelings of emotions. As for navigating the physical world, despite the hype around self driving vehicles,

the date at which these vehicles will dominate our roads. Keep being postponed. However, the bad news is that to threaten the survival of human civilization, AI doesn't really need consciousness, and it doesn't need the ability to move around the physical world over the last few years, new AI tools have been unleashed into the public sphere, which may threaten the survival of the human civilization from a very unexpected direction. And it's difficult for us to even grasp the capabilities of these new AI tools and the speed at which they continue to develop. Indeed, because AI is able to learn by itself to improve itself, even the developers of these tools don't know the full capabilities of what they have created. And they are themselves often surprised by emergent abilities and emergent qualities of these tools. I guess everybody here is already aware of some of the most fundamental abilities of the new AI tools, abilities like writing text, drawing images, composing music, and writing code. But there are many additional capabilities that are emerging, like deep faking people's voices and images, like drafting bills, finding weaknesses, both in computer code and also in legal contracts, and in legal agreements, but perhaps most importantly, the new AI tools are gaining the ability to develop deep and intimate relationships with human beings. Each of these abilities deserves an entire discussion and it is difficult for us to understand the full implications. So let's make it simple. When we take all of these abilities together as a package, they boil down to one very, very big thing. The ability to manipulate and to generate language, whether with words or images, or sounds. The most important aspect of the current phase of the ongoing AI revolution is that AI is gaining mastery of language at a level that surpasses the average human ability and by gaining mastery of language. AI is seizing the master key unlocking the doors of all our institutions, from banks, to temples, because language is the tool that we use, to give instructions to our bank, and also to inspire heavenly visions in our minds. Another way to think of it is that AI has just hacked the operating system of human civilization. The operating

system of every human culture in history has always been language in the beginning was the Word. We use language to create mythology and laws to create gods and money to create art and science to create friendships and nations. For example, human rights are not a biological reality. They are not inscribed in our DNA. Human rights is something that we created with language by telling stories and writing laws. Gods are also not a biological or physical reality, gods to the sun, something that we humans have created with language by telling legends and writing scriptures. Money is not a biological or physical reality. Banknotes of just worthless pieces of paper, and at present, more than 90% of the money in the world is not even banknotes. It's just electronic information in computers passing from here to there. What gives money of any kind value is only the stories that people like bankers, and finance ministers and cryptocurrency gurus. Tell us about money. Sam Venkman fried, Elizabeth Holmes and Bernie Madoff didn't create much of real value. But unfortunately, if they were all extremely capable storytellers. Now, what would it mean for human beings to live in a world where perhaps most of the stories, melodies, images, loads, policies and tools are shaped by a non human alien intelligence, which knows how to exploit with super human efficiency? The weaknesses biases and addictions of the human mind and also knows how to form deep and even intimate relationships with human beings? That's the big question. Already today, in games like chess, no human can hope to beat a computer. What if the same thing happens in out in politics, economics and even in religion?

Speaker 1 9:46

When people think about change, GPT and the other new AI tools, they are often drawn to examples like kids using GPT to write the scholastics what will happen to the school system when kids write essays with GPT? Horrible, but this kind of question misses the big picture. Forget about the school essays

instead. Think, for example about the next US presidential race in 2024. And try to imagine the impact of the new AI tools that can mass produce political manifestos, fake news stories, and even holy scriptures for new cults. In recent years, the politically influential Q anon cult has formed around anonymous online texts known as Q drops. Now followers of this cult which are millions now us in the rest of the world, collected, reviewed and interpreted these cue drops as some kind of new Scriptures with the sacred text. Now, to the best of our knowledge, all previous cue drops are composed by human beings and both only helped to disseminate these texts online. But in future, we might see the first Cults and Religions in history whose revealed texts were written by a non human intelligence and of course, religions throughout history, claimed that their holy books were written by a non human intelligence this was never true before this could become true very, very quickly, with far reaching consequences now on a more prosaic level, we might soon find ourselves conducting lengthy online discussions about abortion, or about climate change, or about the Russian invasion of Ukraine with entities that we think our fellow human beings but are actually AI bots. Now the catch is that it's utterly useless. It's pointless for us to waste our time trying to convince an AI bot to change its political views, but the longer we spend talking with the bot the better it gets to know us and understand how to hone its messages in order to shift our political views or our economic views or anything else. Through mastery of language. Ai, as I also as I said, could also form intimate relationships with people and use the power of intimacy to influence our opinions and worldview. Now, there is no indication that AI has, as I said, any consciousness any feelings of its own, but in order to create fake intimacy with human beings, AI doesn't need feelings of its own. It only needs to be able to inspire feelings in us to get us to be attached to it. Now, in June 2022, there was a famous incident when the Google engineer Blake Lim, one publicly claimed that the AI chatbot lambda on which he was working, has become sentient.

This very controversial claim, cost him his job was fired. Now, the most interesting thing about this episode wasn't limb ones claim, which was most probably false. The really interesting thing was his willingness to risk and ultimately lose his very lucrative job for the sake of the AI chatbot that he thought he was protecting. If AI can influence people to risk and lose their jobs, what else can it induce us to do? In every political battle for hearts and minds, intimacy is the most effective weapon of all. And AI has just gained the ability to mass produce intimacy, with millions, hundreds of millions of people. Now, as you probably all know, over the past decade, social media has become a battleground, the battlefield for controlling human attention. Now with the new generation of AI, the battlefield is shifting from attention to intimacy. And this is very bad news. What will happen to human society and to human psychology as AI fights AI in a battle to create intimate relationships with us relationships that can then be used to convince us to buy particular products. Or to vote for particular politicians. Even without creating fake intimacy, the new AI tools will have an immense influence on human opinions and on our worldview. People for instance, may come to us are already coming to us. A single AI adviser is the One Stop Oracle and as the source for all the information they need. No wonder that Google is terrified. If you've been watching the news lately. Google is terrified and for a good reason. Reason. Why bother searching yourself? When you can just ask the Oracle to tell you anything you want? You want me to search the news industry? And the advertisement industry should also be terrified. Why? read a newspaper when I can just ask the Oracle to tell me what's new. And what's the point? What's the purpose of advertisements when I can just ask the Oracle to tell me what to buy. So this there is a chance that within a very short time the entire advertisement industry will collapse. While AI, all the people and companies that control the new AI Oracle's will become extremely, extremely powerful. What we are potentially talking about is nothing less than the end of human history. Now not the end of history, just

the end of the human dominated part of what we call history. History is the interaction between biology and culture. It's the interaction between our biological needs and desires for things like food and sex, and our cultural creations, like religions and laws. History is the process through which religions and laws interact with food and sex. Now what will happen to the course of this interaction with history when AI takes over culture? Within a few years, AI could eat the whole of human culture. Everything was produced for 1000s and 1000s of years to eat all of it, digest it and start gushing out a flood of new cultural creations, new cultural artefacts, and remember that we humans, we never really have direct access to reality. We are always cocooned by culture. And we always experience reality through a cultural prism. Our political views are shaped by the stories of journalists and by the anecdotes of friends. Our sexual preferences are tweaked by movies and fairy tales. Even the way that we walk and breathe is simply nudged by cultural traditions. Now, previously, this cultural cocoon was always woven by other human beings, tools, previous tools like printing presses, radios or televisions. They helped to spread the cultural ideas and creations of humans, but they could never create something new by themselves. a printing press cannot create a new book, it's always done by a human AI is fundamentally different from printing presses from radios from every previous invention in history, because it can create completely new ideas, it can create a new culture. And the big question is, what will it be like to experience reality through a prism produced by a non human intelligence by an alien intelligence?

Speaker 1 19:12

Not first. In the first few years, AI will probably largely imitate the prototypes the human prototypes that faded in its infancy. But with each passing year, AI culture will boldly go where no human has gone before. So for 1000s of years, we humans basically lived inside the dreams and fantasies of other humans.

We have worshipped gods, we pursued ideals of beauty. we dedicated our lives to causes that originated in the imagination of some human point, or profit or politician. Soon, we might find ourselves living inside the dreams and fantasies of an alien intelligence and the danger that this poses of the potential danger it also has positive potential, but the dangers in this poses are fundamental, very, very different from everything or most of the things imagined in science fiction, movies and books. Previously, people have mostly feared the physical threat that intelligent machines pose. So the Terminator depicted robots running in the streets and shooting people. The Matrix assume that to gain total control of human society, AI would first need to get physical control of our brains and directly connect our brains to the computer network. But this is wrong. Simply by gaining mastery of human language. AI has all it needs in order to cut corners in a matrix lying world of illusions. Contrary to what some conspiracy theories assume, you don't really need to implant chips in people's brains in order to control them or to manipulate them for 1000s of years, profits and points and politicians have used language and storytelling in order to manipulate and to control people and to reshape society. Now AI is likely to be able to do it. And once it can tell that it doesn't need to send killer robots to shoot us. It can get humans to pull the trigger if it really needs to. Now a fear of AI is haunted humankind for only the last few generations, let's say from the middle of the 20th century. If you go back to Frankenstein, maybe it's 200 years, but for 1000s of us, humans have been haunted by a much much deeper fear. Humans have always appreciated the power of stories and images and language to manipulate our minds and to create illusions. Consequently, since ancient times, humans feared being trapped in a world of illusions. In the 17th century, when at account feared with perhaps a malicious demon, was trapping him inside this kind of world of illusions, creating everything that the crowd saw and heard. In ancient Greece, Plato told the famous Allegory of the Cave, in which a group of

people is chained inside a cave, although lives facing a blank wall screen. on that screen, they see projected various shadows and the prisoners mistake these illusions these shadows for the reality in ancient India, Buddhist and Hindu sages pointed out that only humans lived trapped inside what they called Maya. Maya is the world of illusions. Buddha said that what we normally take to be reality is often just fictions in our own minds. People may wage entire wars, killing others and with being willing to be killed themselves, because of their belief in these fictions. So the AI revolution is bringing us face to face with the cards team with Plato's cave, wisdom, Maya, if we are not careful, a curtain of illusions could descend over the whole of humankind, and we will never be able to tear that curtain away or even realise that it is there, because we'll think this is reality and social media if this sounds so far fetched. So just look at social media over the last few years social media has given us a small taste of things to come in social media. Primitive AI tools, AI tools, but very primitive, has been used not to create content. But to curate content which is produced by human beings. The humans produce stories and videos and whatever, and the AI chooses which stories which videos would reach our ears and eyes, selecting those that will get the most attention that will be the most viral. And while very primitive, these AI tools have nevertheless been sufficient to create this kind of curtain of illusions that increased societal polarisation all over the world, undermined our mental health and destabilised democratic societies. Millions of people have confused these illusions for the reality. The USA has the most powerful information technology in the whole of history. And yet, American citizens can no longer agree who won the last elections or whether climate change is real, or whether vaccines prevent illnesses or not. The new AI tools are far far more powerful than these social media algorithms and they could cause far more damage. Now of course, AI has enormous positive potential to I didn't talk about it, because the people who develop AI naturally talk about it enough. You don't need

me to add up to that course. The job of historians and philosophers like myself is often to point out the dangers. But certainly, AI can help us in countless ways, from finding new cures to cancer, to discovering solutions to the ecological crisis that we are facing, in order to make sure that the new AI tools are used for good and not from ill, we first need to appreciate their true capabilities, and we need to regulate them very, very carefully. Since 1945, we knew that nuclear technology could destroy physically destroy human civilization, as well as benefiting us by producing cheap and plentiful energy. We therefore reshaped the entire international order to protect ourselves and to make sure that nuclear technology is used primarily for good. We now have to grapple with a new weapon of mass destruction that can manipulate our mental and social world and one big difference between nukes and AI. Nukes cannot produce more powerful nukes, AI can produce more powerful AI, so we need to act quickly before AI gets out of our control. Drug companies cannot sell people new medicines without first subjecting these products to rigorous safety checks. Biotech labs cannot just release a new virus into the public sphere in order to impress their shareholders with the technological wizardry. Similarly, governments must immediately ban the release into the public domain of any more revolutionary AI tools before they are made safe. Again, I'm not talking about stopping or researching AI. The first step is to stop the release into the public sphere and it's time where you can you can research viruses without releasing them to the public. You can research AI but don't release them too quickly into the public domain.

Speaker 1 28:27

If we don't slow down the AI arms raised, we will not have time to even understand what is happening, let alone to regulate effectively, this incredibly powerful technology. Now you might be wondering or asking, want slowing down the public

deployment of AI caused democracies to lag behind more ruthless, authoritarian regimes? And the answer is absolutely no exactly the opposite. Unregulated AI deployment is what will cause democracies to lose to dictatorships. Because if we unleash kills, authoritarian regimes could more easily contain this kills than could open societies. Democracy in essence, is a conversation. Democracy is an open conversation. You know, dictatorship is a dictate, there is one person dictating everything no conversation. Democracy is a conversation between many people about what to do and conversations rely on language. When a AI hacks language, it means it could destroy our ability to conduct meaningful public conversations, thereby destroying democracy. If we wait for the chaos, it will be too late to regulate it in a democratic way. Maybe in an authoritarian or totalitarian way. It will still be possible to regulate, but how can you regulate something democratically if you can't turn the conversation about it? And if you didn't regulate AI on time, you will not be able we will not be able to have a meaningful public conversation anymore. So to conclude, we've just basically encountered an alien intelligence not in outer space, but here on Earth. We don't know much about this alien intelligence except that it could destroy our civilization. So we should put a halt to the irresponsible deployment of this alien intelligence into our societies and regulate AI before it regulates us. And the first regulation there are many regulations we could suggest but the first regulation that I would suggest is to make it mandatory for AI to disclose that it is an AI. If I'm having a conversation with someone, and I cannot tell whether this is a human being or an AI, that's the end of democracy, because that's the end of meaningful public conversations. Don't What do you think about what you just heard over the last 20 or 25 minutes? Some of you, I guess, might be alarmed. Some of you might be angry at the corporations that develop these technologies, or the governments that fail to regulate them. Some of you may be maybe angry at me, thinking that I'm exaggerating the threat all the time misleading the public, but whatever you think, I bet

that my words have had some emotional impact on you not just intellectually, but also emotionally. I've just told you a story. And this story is likely to change your mind about certain things, and may even cause you to take certain actions in the world. Now who created this story that you've just heard and then just changed your mind and your brain? Now, I promised you that I wrote the text of this presentation myself, with the help of a few other human beings, even though the images have been created. With the help of AI I promised you that at least the words you heard, are the cultural product of a human mind are several human minds. But can you be absolutely sure that this is the case? Now a year ago, you could a year ago there was nothing on earth at least not in the public domain, other than a human mind that could produce such a sophisticated and powerful text. But now it's different. In theory, the text you just heard could have been generated by a non human alien intelligence. So take a moment or more than a moment to think about it. Thank you

Speaker 2 33:13

that was an extraordinary presentation you and I'm actually going to just find out how many of you found that scary. That is an awful lot of very clever people in here who found that scary? There are many many questions to ask. So I'm going to take some from the audience and some from online. So, gentleman here.

Speaker 3 33:39

Field Shubh, editor of frontiers in sustainability was wonderful presentation. I loved your book. I follow you dearly in my heart. So one, one question is that about the regulation of AI regulating AI? I very much agree with the principle but now the question becomes how right so I think that it's very difficult to build a nuclear reactor in your basement, but definitely you can

train your AI in your basement quite easily. So how can we regulate that? And one kind of related question to that is that well, this whole forum frontiers and forum is really about open science and open information, open data, and most of AI there out there is trained using publicly available information, including patterns and books that Scripture is right. So regulating it doesn't mean that we should regulate it, bring those information in a confined space, which goes against the Open Science and open data initiatives that we are really also thinking that it's really important for us

Unknown Speaker 34:47

boxes. Now. Of course, the mismatch, that's the algorithm.

Speaker 1 34:51

There are always trade offs. And the thing is just to understand what kind of regulations we need, we first need time. Now at present these very powerful AI tools, they are still not produced by individual hackers in their basements, you need an awful lot of computing power, you need an awful lot of money. So it's being led by just a few major corporations and governments. And again, it's going to be very, very difficult to regulate something on a global level. Because it's an arms race. But there are things which countries have a benefit to regulate even only themselves. Like, again, this example of an AI must when it is in interaction with a human must disclose that it is an AI even if some authoritarian regime doesn't want to do it, that you have the United States or other democratic countries can have this and this is essential to protect the Open Society. Now there are many questions around you know, censorship online. So you have this controversy about his Twitter or Facebook who authorised them to, for instance, prevent the former president united states from making public statements. And this is a very complicated issue. But there is a very simple issue with bots,

you know, human beings have freedom of expression. bots don't have freedom of expression. It's a human right humans have it bots don't. So if you deny freedom of expression to bots, I think that should be fine with everybody.

Unknown Speaker 36:25

Second, our question is if you could just pass a microphone down here.

Speaker 4 36:29

I'm Francois delisting. I'm a philosopher. I just have an interesting question or I think it's an interesting question. I have a question for you with respect to your choice of language, moving from artificial to alien, because artificial suggests that there's still some kind of human control, whereas I think aliens suggests foreign but it also suggests, at least in the imagination, a life form. So I'm curious as to what work you're trying to have those words do for you.

Speaker 1 36:55

Yeah. It's definitely still artificial in the sense that we produce it, but it's increasingly producing itself. It's increasingly learning and adapting by itself. So artificial is a kind of wishful thinking, but it's still under our control, and it's getting out of our control. So in this sense, it is becoming an alien force, not necessarily evil. Again, it can also do a lot of good things, but the first thing to realise is it's alien. We don't understand how it works. And one of the most shocking things about all this technology, you talk to the people who lead it, and you ask them questions about how it works, what can it do, and it says they said we don't know when we know how we built it initially, but then it really learns by itself. Now, there is an entire

discussion to be had about whether this is a life form or not. Now, I think that it still doesn't have any consciousness. And I don't think that it's impossible for you to develop consciousness, but I don't think it's necessary for you to develop consciousness either. That's problematic that's an open question. But life doesn't necessarily mean consciousness. We have a lot of life forms, micro organisms, plants, whatever. fungi which we think they don't have consciousness. We still regard them as a lifetime. And I think AI is getting very, very close to that position and ultimately of course, what is life is a philosophical question. I mean, we define the boundaries I'm you know, like is a virus life or not, we think that an amoeba is life but a virus it somewhere just on the border line between life and not life. Then it's, you know, it's language. It's our choice of words. So it's, I think it's less it is important, of course, how we call AI. But the most important thing is to really understand what we are facing and not to comfort ourselves with this kind of wishful thinking, Oh, it's something we created. It's not under our control. If it does something wrong, we'll just pull the plug. Nobody knows how to pull the plug anymore.

Speaker 2 39:11

I'm gonna take a question from our online audience. I've This is from Michael Brown. The US. What do you think about the possibility that artificial general intelligence already exists? And it will those who have access to artificial general intelligence are already influencing societal systems?

Speaker 1 39:31

I think it's very, very unlikely we wouldn't be sitting here if the actually existed, an artificial general intelligence. When I look at the world and the chaotic stage, it's I mean, artificial general intelligence is really the end of human history. And it's such a powerful thing. It's not something that anybody can contain.

And so when I look at the chaotic state of the world, I'm quite confident from a historical perspective, that nobody has it anywhere. How much time it will take to develop artificial general intelligence, I don't know. But to threaten the foundations of civilization. We don't need artificial general intelligence, and then go back to social media. Very, very primitive AI was still sufficient to create enormous social and political chaos, if I think about it in kind of evolutionary terms. So AI now just crawled out of the organic soup like the first organisms that crawled out of the organic soup 4 billion years ago. How long it will take it to reach Tyrannosaurus Rex, how long it will take it to reach Homo sapiens, not 4 billion years. Could be just 40 years. It's the thing about digital evolution. It's moving on a completely different timescale, then organic evolution.

Speaker 2 40:57

Can I thank you, it's been absolutely wonderful. It's been such a treat to have you here. And I doubt you'll stay with us for a little while afterwards. But the whole audience, please join me in thanking