

Transcript:

(00:00) foreign [Music] 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 this conference is focused on the ecological crisis facing Humanity but for better All For Worse ai2 is 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 AI will probably change the very meaning of the ecological system because for 4 billion years the ecological system of planet Earth contained only organic life forms and now or soon 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 of 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

(01:54) because science fiction scenarios usually assume that before AI can pose a significant threat to humanity it will have to reach or to pass to 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 Adept 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

(02:45) 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 chat 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 keeps being postponed however the bad news is that to threaten the

(03:37) 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 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

(04:27) 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

(05:17) 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 their 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 (06:09) 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

(07:02) 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 God's too is something that we humans have created with Language by telling Legends and writing

(07:53) scriptures money is not a biological or physical reality banknotes are just worthless pieces of paper and at present more than 90 percent 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 bankman freed Elizabeth Holmes and Bernie Madoff didn't create much of real value but unfortunately they were all

(08:41) extremely capable storytellers now what would it mean for human beings to live in a world where perhaps most of the stories Melodies images laws policies and tools are shaped by a non-human alien intelligence which knows how to exploit with superhuman 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 art in

(09:38) politics economics and even in religion when people think about chat GPT and the other new AI tools they are often drawn to examples like kids using Chat GPT to write their school essays what will happen to the school system when kids write essays with shajipiti horrible but this kind of question misses the big picture forget about the school essays instead think for example about the next U.

(10:11) S 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 the Q drops now followers of this cult which are millions now in the US in the rest of the world collected reviewed and interpreted these Q drops as some kind of new scripture so the sacred text now to the best of our knowledge all previous Q drops were composed by human

(10:59) beings and Bots only helped to disseminate these texts online but in future we might see the first Cults and religions in history whose reviewed 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

(11:48) about the Russian invasion of Ukraine with entities that we think are 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 its Mastery of language AI as I also as I said could also form intimate

(12:32) relationships with people and use the power of intimacy to influence our opinions and well View 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 it was a famous incident when the Google engineer Blake lemon publicly claimed that the AI chat about Lambda on which he was working has

(13:17) become sentient this very controversial claim cost him his job was fired now the most interesting thing about this episode wasn't lemon's 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 heart and Minds intimacy is the most effective weapon of all and AI has just gained the

(14:09) ability to must 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 a battlefield for controlling human attention now with the new generation of AI the Battlefront 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

(14:55) or to vote for particular politicians even without creating fake intimacy the new AI tools would have an immense influence on human opinions and on our world view people for instance may come to use are already coming to use a single AI advisor as 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 newsletter Google is terrified and for a good Rhythm reason why bother searching yourself when you can just ask the Oracle to tell you

anything you want you

(15:42) don't need 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 is there is a chance that within a very short time the entire advertisement industry will collapse while I or the people and companies that control the new AI oracles will become extremely extremely powerful

(16:23) 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 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 cause of this interaction of history when AI

(17:07) takes over culture within a few years AI could eat the whole of human culture everything was produced for thousands and thousands of years to eat all of it digest it and start gushing out a flood of new cultural Creations new cultural artifacts 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 prison our political views are shaped by the stories of uh journalists and by the anecdotes of friends of sexual

(17:56) preferences are tweaked by movies and fairy tales even the way that we walk and breathe is something nudged by cultural traditions now previously this cultural cocoon was always woven by other human beings tools previous tools like printing presses or 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

(18:45) 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 now at first in the first few years AI will probably largely imitate the prototypes the human prototypes that fed it in its infancy but with each passing year AI culture will boldly go where no human has gone before so for thousands of years we humans basically lived inside the dreams and

(19:36) fantasies of other humans we have worshiped Gods we pursued ideals of beauty we dedicated Our Lives to causes that originated in the imagination of some human poet or Prophet or politician soon we might find ourselves living inside the dreams and fantasies of an alien intelligence and the danger that disposes of the potential danger it also has positive potential but the dangers it disposes are fundamental very very different from everything or most of the things imagined in science fiction movies and books

(20:15) 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 assumed 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 coconuts in a matrix like World of illusions contrary to what some conspiracy theories assume you don't really need to

(21:07) implant chips in people's brains in order to control them or to manipulate them for thousands of years profits and Poets 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 doubt 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 fear of AI has haunted humankind for only the last few Generations let's say from the middle of the 20th century if

(21:48) you go back to Frankenstein maybe it's 200 years but for thousands of years 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 I decout feared that perhaps a malicious demon was trapping him inside this kind of world of Illusions creating everything that the cult saw and heard

(22:32) in ancient Greece Plato told the famous allegory of the cave in which a group of people is Chained inside a cave all their lives facing a blank wall a screen on that screen they see projected various Shadows and the prisoner's mistake these Illusions these Shadows for the reality in ancient India Buddhists and Hindu sages pointed out that all 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

(23:24) killing others and 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 demon with Plato's Cave with the 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 realize that it is there because we'll think this is reality and social media if this sounds far-fetched so just look at social media over the last few years social media has

(24:08) given us a small taste of things to come in social media primitive AI tools AI tools but very primitive have 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

(24:54) that increased societal polarization all over the world undermined our mental health and destabilized 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

(25:41) damage now of course AI has enormous positive potential too 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 chorus 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

(26:25) 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 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 annihilate our mental and social world and one big difference between nukes and

(27:17) 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 own 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

(28:06) talking about stopping all research in AI the first step is to stop the release into the public school it's somewhere 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 um if we don't slow down the AI arms race 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

(28:45) of AI cause democracies to lag behind more ruthless of the retirement 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 chaos authoritarian regimes could more easily contain these chaos 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

(29:27) between many people about what to do and conversations rely on language when 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 a totalitarian rate will still be possible to regulate but how can you regulate something democratically if you can't hold the conversation about it and if you didn't regulate AI on time you will not be able

(30:07) we will not be able to have a meaningful public conversation anymore so to conclude we have 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 fir

st regulation that many regulations we could suggest but the first regulation that I would suggest is (30:49) 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 now 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

(31:30) thinking that I'm exaggerating the threat or the time misleading the public but whatever you think I bet that my words have had some emotional impact on you not just intellectual impact also emotional impact 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 that just changed your mind and your brain now I promised you that I wrote the text of this presentation myself with the

(32:11) help of a few other human beings even though the images have been created with the help of AI I promise you that at least the words you heard of the cultural product of a human mind or 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

(32:55) generated by a non-human alien intelligence so take a moment or more than a moment to think about it thank you [Applause] that was an extraordinary presentation you about 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 a gentlemen here and the field trip editor of Frontiers and sustainability was wonderful

(33:40) presentation I love your book I follow you dearly in my heart so one one question out of many is that about the regulation of AI regulating AI I very much agree with the principle but now the question becomes how right so um I think that it's very difficult to build the nuclear reactor in your basement but definitely you can train your AI in your basement quite easily so how can we regulate that and um one kind of related question to that is that well this whole Forum Frontiers and forum is really about open and

(34:19) Science and open information open data and most of AI that are out there is trained using publicly available information including patterns and books and scriptures right so regulating AI doesn't mean that we should regulate and bring those information in a confined space which goes against the open science and open data initiatives that we are really also thinking that is really important for us black box is an algorithm isn't it that's the algorithm no I know they'll always trade-offs and the thing is just to understand what

(34:53) 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 uh regulate something on a global level because it's it's an arms race but there are things which countries have a benefit to regulate even only themselves

(35:27) 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 the EU of 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 is Twitter or Facebook who authorized them to for instance prevent the former president of the United States from making public statements and this

(36:02) 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 enough expression it's a human right humans have it but bots don't so if you deny freedom of expression to bots I think that should be fine with everybody uh let's take another question is if you could just pass the microphone down here I'm Prince was dearest and I'm a philosopher I just have an interesting question oh I think it's an interesting question there you go um I have a

(36:33) 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 alien suggests foreign but it also suggests at least in the imagination of life form so I'm curious as to what work you're trying to have those words do for you yeah um 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

(37:06) itself so artificial is a kind of wishful thinking that 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 realize is it's alien we don't understand how it works or in 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 I do and it says they said we don't know I mean we we know how we

(37:37) built it initially but then it 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 uh I don't think that it's impossible for it to develop Consciousness but I don't think it's necessary for it to develop Consciousness either that's a problematic that's an open question but life doesn't necessarily mean consciousness we have a lot of life forms microorganisms plants whatever

(38:12) fungi which we think they don't have Consciousness we still regard them as a life form and I think AI is getting very very close to that position ultimately of course what is life is a philosophical question I mean we Define the boundaries and you know like is a virus life or not we think that an amoeba is life but a virus it's somewhere just on the borderline between life and not life um then it's you know it's it's language it's our choice of words so it's I think it's less it is

(38:47) 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 our under our control if it does something wrong we'll just pull the plug nobody knows how to pull the plug anymore I'm going to take a question from our online audience um I've uh this is from Michael Brown in the US what do you think about the possibility that artificial general intelligence already exists and it or

(39:22) those who have access to artificial general intelligence are already influencing societal systems 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 um so when I look at the chaotic state of the world I'm quite confident

(39:54) again from a historical perspective that nobody has it anywhere um how much time it will take to develop artificial general intelligence I don't know um but to threaten the foundations of civilization we don't need artificial general intelligence and then going 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

(40:32) four billion years ago how long it will take it to reach Tyrannosaurus Rex how long it will take it to reach Homo sapiens not for billion years could be just 40 years it's the thing about digital Evolution it's moving on a completely different time scale than organic evolution can I thank you it's been absolutely wonderful it's been such a treat to have you here and I've no doubt you'll stay with us for a little uh while afterwards but the whole audience please join me in thanking [Applause]

(41:14) [Music]