Artificial Intelligence, Democracy, & the Future of Civilization | Yoshua Bengio & Yuval Noah Harari - YouTube

https://www.youtube.com/watch?v=TKopbyIPo6Y

## Summary and Key points by Video Insights :

The text appears to be a transcript of a discussion between Vashi Capellos, Professor Bengio, and Professor Harari about the implications of Artificial Intelligence (AI). Key points from the discussion include:

1. Vashi Capellos introduces the discussion and mentions that it will focus on the potential threats and benefits of AI.
2. Professor Harari emphasizes that AI is unique because it's the first technology that can make decisions and create ideas on its own. He points out that this is fundamentally different from any technology we've seen before as it potentially takes power away from us. He also states that while there are positive uses for this kind of power, there are also many negative uses.
3. Professor Bengio discusses two main threats he sees from AI. First, in the short term, he's concerned that large language models could be used for disinformation and propaganda. Second, he worries about a future where AI becomes at least as intelligent as humans, which could lead to them being smarter than us in some ways due to their access to data and digital communication bandwidth.
4. Both professors acknowledge the existential threat posed by AI if it surpasses human intelligence. Harari agrees with Bengio's analogy comparing humans to frogs if we create a species smarter than us, implying that it does not bode well for humanity.
5. Harari also discusses the potential threat to democracy posed by AI and mentions that it's important to think about how fast AI is developing compared to the slow pace of biological evolution.
6. Bengio mentions the uncertainty about when machines might become as intelligent or smarter than humans, suggesting it could be in a few years or decades. He emphasizes the importance of governments preparing for the worst-case scenario and the need to put guardrails in place to minimize risks.
7. Both professors have signed a letter asking companies developing AI to pause their work due to the existential risk to humanity. Harari mentions the need for time to adapt to new technologies, referencing the Industrial Revolution and the devastating experiments of Nazism and Soviet communism that resulted from trying to adapt too quickly.
8. Harari argues that we need to be extremely careful and take things more slowly when dealing with powerful technology like AI to avoid catastrophic failures. He suggests that we won't get a second chance if we make mistakes with this kind of technology, as we did during the Industrial Revolution.

The speakers in the video express their concerns about the lack of regulation surrounding artificial intelligence (AI) and its applications. They mention that the letter they signed was primarily aimed at corporations and governments, asking them to pause and consider the implications of unchecked AI development. While they didn't expect corporations to change their behaviors due to competitive pressures, they hoped the letter would draw the public's and governments' attention to the dangers of AI.

They highlight the lack of political discourse around AI, which they believe should be a top concern given its potential existential dangers and its increasing influence on everyday life, such as job applications and loan approvals.

The speakers also highlight the responsibility of politicians and AI experts to acknowledge the potential dangers of AI, rather than focusing solely on its positive aspects. They argue that media also has a responsibility to push for this discussion.

They consider recent Congressional hearings in the US, where AI experts have pleaded for more regulation, to be a significant step. They believe regulation should differentiate between AI development (research) and deployment (implementation into society). Simple rules, such as not allowing AI to counterfeit humans, should be implemented to prevent the collapse of public conversation and democracy.

They argue for a regulatory body that has agility and the ability to adapt to rapidly changing technologies and their potential misuse. They express concerns over some governments' reluctance to delegate power to such agencies.

Finally, they suggest some low-tech solutions for regulation, such as requiring individuals to physically sign up for social media accounts to combat the prevalence of bots. They note the paradox of having powerful information technology but a collapsing public conversation, emphasizing the need for regulation to preserve democracy.

The discussion centers around the use of Artificial Intelligence (AI) and its potential implications for democracy and society. The conversation consists of several key points:

1. **Concerns about AI in conversations and decision-making**: The participants express concern about the potential misuse of AI in social and political conversations. They note that if an AI is used in a conversation, it could manipulate the human participant's views by building intimacy over time. The AI does not have a mind or opinion, but it can still influence human decisions. This scenario is particularly troubling in the context of political discussions, where AIs could undermine democratic decision-making processes.
2. **Shift from attention to intimacy**: The conversation highlights a shift in the battlefront of AI influence, from attention to intimacy. With the growth of AI capabilities, there is concern that millions or billions of AI agents could manipulate human behavior and decision-making by gaining our intimacy.
3. **Concerns about social collapse**: The speakers are worried that unchecked AI influence could lead to social collapse. They mention how AI could be used to persuade individuals to buy a product, vote for a certain politician, or make other decisions, which could have serious societal consequences if not regulated.
4. **Democracy and social media**: The conversation also covers the complex relationship between democracy, social media, and AI. The participants express concern that social media, and eventually AI, could exploit human vulnerabilities, such as our susceptibility to influence and our tendency to trust information from sources we have a relationship with. They argue that these factors could amplify existing challenges in democracy.
5. **Challenges to regulation**: The conversation highlights the difficulties in legislating and regulating AI due to conflicting interests. Corporations and some individuals may resist regulations that they see as infringing on free speech or slowing down progress.
6. **Bot rights and free speech**: One participant argues that concerns about free speech do not apply to AI, as bots do not have rights or free speech. They suggest that banning bots from platforms like Twitter or Facebook should be a simpler issue than banning human users.
7. **Need for public awareness**: The conversation emphasizes the need for public discussions to raise awareness about the potential risks of AI to democracy. The participants suggest that the threat of losing democratic freedoms could motivate people to support necessary regulations and precautions.
8. **Critique of world government and totalitarian systems**: The speakers express skepticism about the idea of a world government or an authoritarian system to control AI. They argue that such systems are impractical, potentially dangerous, and could be more susceptible to misuse of AI due to a lack of self-correcting mechanisms. They maintain that democracies, with their checks and balances, are better equipped to regulate AI.
9. **AI and infallibility**: The conversation concludes by highlighting that AI is not infallible. Even though AI can process vast amounts of information, it does not necessarily produce truth or wisdom, and its use could lead to mistakes. Democracies are seen as the best systems to identify and correct such errors.
10. **Resistance to surveillance solutions**: Finally, the speakers resist the idea of a surveillance solution to control AI, where a central government watches everything and uses AI to prevent harmful actions. They argue that this idea overlooks the dangers of giving too much power to a central authority and AI.

Here are the key points summarized from the text:

1. The speaker emphasizes the importance of making sure that AI does not restrict human freedom and control individuals too much. They believe it's important to find ways to organize society that preserve human values, rights, dignity, and democracy​1​.
2. The speaker suggests that even though creating a scenario where everyone wins might be challenging, it's crucial to keep trying and thinking through solutions. They draw a parallel to the ongoing fight against climate change, which also seemed like a desperate cause for many years, and argue that the same determination should be applied to the challenge of AI​2​.
3. The speaker believes that AI could potentially bring about many positive changes, such as improving health care and education, solving climate change, and reforming society in a way that could lead to a utopia. Despite this, they caution that this doesn't mean they're inherently optimistic, rather, they feel a moral duty to strive for the optimistic solution and try to make it happen​3​.
4. The speaker emphasizes that humans have the power to solve the problems they've created, and that it's important to spend as much time developing and understanding our own minds as we do developing AI. They believe that the main problem is not AI itself, but bad human actors using AI in harmful ways​4​.
5. The speaker suggests that changing humans is the only way to avoid a totalitarian regime controlled by AI. They believe that it's possible to make humans more ethical, wiser, more compassionate, and more rational, and argue that this should be the goal for all of humanity​5​.
6. The speaker acknowledges that focusing on potential catastrophic risks might alienate people who are concerned about current societal issues. They argue that we need to address both current and potential future problems at the same time, and that there might be a convergence of solutions​6​.
7. The speaker suggests that getting political and public buy-in to address these problems is crucial. They highlight the importance of focusing society's attention on these problems, and express concern about the potential impact of AI on the job market. They argue that we need more people in the political system who have a deep understanding of technology and its impact, and less people who just want to make billions in business​7​.
8. Finally, the speaker suggests that the threat AI poses to jobs is a tangible way to ground the thread it poses, just like the existential threat climate change poses. They argue that grounding these threats in this way can change the impetus for politicians to act​8​.

Here are the key points summarized from the text:

1. The speaker discusses the possible impact of AI on jobs, mentioning a study by OpenAI and other academics that suggests many jobs could be modified by AI, leading to increased productivity. This could either result in fewer people doing those jobs or the same number of people doing more work, using the example of programming. However, they also note that even if we have the technology to change jobs drastically, societal changes often occur slowly, so it may take years or even decades for these changes to significantly impact the job market​1​.
2. The speaker posits that AI could rapidly automate jobs that involve language manipulation, such as writing emails or working with databases. They point out that it's hard to predict whether companies will be able to implement this technology quickly, but if they do, we could face significant transition problems. They suggest that the question of whether an AI could take over someone's job is a tangible concern that people can easily understand​2​.
3. The speaker points out that while AI will likely create many new jobs, the transition period could be difficult, especially considering global implications. They mention that the AI revolution is being led by a few countries that are likely to become very wealthy and powerful, which could damage the economies of less developed countries. For instance, if it becomes cheaper to produce textiles in developed countries due to AI, the economies of countries like Honduras and Bangladesh, which rely heavily on the textile industry, could be severely impacted​3​.
4. The speaker discusses the challenges of retraining workers who have lost their jobs due to AI, particularly in less developed countries. They note that while developed countries may be able to use the profits from the AI revolution to cushion the blow for their citizens and pay for retraining, they question who will pay for this in less developed countries, suggesting that it's unlikely that companies in developed countries will be taxed to fund retraining in other countries​4​.
5. The speaker warns that the AI revolution could lead to a few countries dominating the world, similar to the Industrial Revolution in the 19th century. They also highlight the potential for political control through data collection, suggesting that it's increasingly possible to control a country by obtaining its data rather than sending in soldiers. They raise the concern of what happens when the personal records of politicians, journalists, judges, and military officers in a country are held by a foreign entity, questioning whether that country can still be considered independent or if it has become a "data colony"​5​.
6. The speaker believes these are immediate dangers that should be clear to any citizen, regardless of their views on the long-term existential risks of AI. They conclude the discussion on this note, expressing hope that the audience now has a better understanding of the challenges posed by AI​6​.

Transcript:

(00:00) my name is vashi capellos I'm very focused on politics so that will form part of the discussion with both professor bengio and Professor Harare welcome welcome to the conversation it's such a pleasure to have you join us this afternoon you can hear me thank you and let me assure you I'm not a deep fake or an AI that is a relief to find out so I wanted to frame the discussion with both of you moving forward around how you view uh at first right now we'll get into the potential you know positive aspects but

(00:30) right now it's a threat as you view AI poses how would you succinctly describe it for our audience this afternoon Professor Harari mm-hmm well there are two things to know about AI it's the first technology in history that can make decisions by itself and it's the first technology in history that can create ideas by itself lots of people are now trying to calm us down by telling us that every time there is a new technology people are afraid of the consequences and in the end we we manage okay but this is nothing like

(01:05) anything we've seen before in history you know whether it's a stone knife or an atom bomb on previous tools empowered us because it was humans who had to decide how to use the bomb but AI can make decisions by itself so it potentially takes power away from us secondly all previous information technology in history they could only reproduce or spread the ideas of human being like the printing press could print the Bible but it could not write the Bible or it could not write a commentary on the Bible San GPT can

(01:49) write can create a completely new commentary on the Bible or on anything else and it can potentially in the future even create new holy texts for future religions you know humans always fantasize about getting their holy scriptures from a superhuman intelligence now it's becoming possible now there are many positive usages for this kind of power but there are also many negative usages and again it's it's fundamentally different from anything we've encountered before because it can take power away from us

(02:30) Professor benchio how do you view the fundamental threat AI poses at this juncture well the problem is that there are many threats but the two most important that I'm concerned with are in the shorter term and it could be as soon as the next U.S election that tools derive from the large language models that we've been talking about can be used for propaganda for this information for personalized trolls that could convince you to vote otherwise and the second threat which may come as soon as a few years from now maybe a

(03:14) decade from now it's it's hard to say is if we fill the Gap bridge the gap between the current state of the art in Ai and the the abilities that human have in terms of intelligence and we build machines that would be at least as smart as us but then automatically they would have advantages because of access to all the data because of digital communication bandwidth all kinds of advantages that allow them to acquire information faster than humans to uh to to share it among themselves all sorts of things that make me think

(03:51) and others that even if we only uncover the same kind of principles that give us our own intelligence they would become smarter than us in some ways and we already see that with things like jbd that in some ways they're smarter than us and they know more things but they're also more stupid in other ways let me just follow up quickly with you on that and bottom line if they are able to be smarter than us what is the threat that that poses is it is it existential yes and what do you mean by that not to

(04:21) depress everyone here we're very it's very weighty but well there's a lot of unknown and there when I mean Jeff Hinton uses this analogy that others have used as well if uh imagine um we created a new species and that species was um you know smarter than us in the same way that we're smarter than uh mice or frogs or something are we treating the frogs well a question we ask ourselves all the time Professor Harare I I I want to get you to weigh in on the same um because you've written extensively

(05:00) about sort of the the threat to democracy and that AI opposes and and more in the near term do you share the concern about the potential longer term threat in that it might be existential in nature yeah absolutely and there are short-term threats which are also very significant like the collapse of the mock oops oh no it's nice to develop and becomes more intelligent than us then I completely agree with the Frog analogy it it doesn't bode well for us we just you froze for a second there so we just caught the ending so we're not

(05:42) here I agree with the Frog analogy that again another way to think about it is that the AI we oh oh no being much much faster than biological evolution I mean if if the AI of today is like an amoeba so just imagine what T-Rex would look like and it won't take billions of years to get there we could get there in a few years well well actually I think that's an excellent point to follow up on with you Professor bengio because I feel like as a regular person not an expert like yourself this stuff is is very scary to hear and uh are there

(06:27) timelines in your mind attached to it I think I asked that question not to you know further or amplify those fears but to bring us to a conversation around how to mitigate those eventualities right and and how fast must policy makers act in response so what kinds of timelines are you thinking of so the problem is that there's a lot of uncertainty about the timeline and um it could be as little as a few years before we at least have the principles to build machines as intelligent as us and that's smarter than us

(07:02) um or it could be decades um and because there is such uncertainty I think we collectively in our governments in particular have a responsibility to prepare for the you know worst case um at least you know the plausible worst case which might be like five years I don't know and what's important um and Yvonne talks a lot about that is the transition like how um hey how do we make sure that we adapt fast enough so if it comes in decades maybe we have a chance to adapt Society if it comes in five years it seems

(07:43) hopeless right so what can we do to slow things down where it's more dangerous how do we start as soon as possible putting the right guard rails to minimize all the risks that we've been talking about today Marie you are both signatories to a letter that essentially uh asks companies behind the development of this technology in response to what Professor bengio just laid out to take a pause on what they're doing why that letter I think again sort of we were just discussing what brought all this to a

(08:18) higher level of public cognizance that letter certainly did did the trick why was it important for you to attach your name to it no because again it's really an existential risk to humanity and what we need above all is time human societies are extremely adaptable we are good at it but we it takes time and if you look for instance at the at the last time that we had a major technological Revolution the Industrial Revolution it took us many generations to figure out how to build a relatively good industrial societies and on the way we

(09:00) had some terrible experiments failed experiments in how to build industrial societies uh Nazism and Soviet communism can be seen as simply failed experiments that killed millions and millions of people in how to build functioning industrial societies and now we are dealing with something even more powerful than the trains and radio and electricity that we invented in the industrial revolution I think there is there is certainly a way to build good societies with AI but it will take time and it will and and we need to make sure

(09:44) that we don't make any more such failed experiments because if if we do it now with with this kind of Technology we won't get a second chance we won't survive it we managed to survive the failed experiments of the Industrial Revolution only because the technology was not powerful enough to destroy us so we have to be extremely careful and we need to take things more slowly in facing huge issues like the professor just laid out Professor bengio there's a corporate or societal response and then

(10:18) there's also a government response this letter is a little bit aimed at both but primarily asking the corporations behind this to take a pause when you signed it did you think they would no thank you for being clear why not um the incentive system we've built in in that as as you've always saying Works reasonably well for our industrial societies in our liberal democracies is based on competition and companies would not survive if they didn't play that game because another one would take their place

(10:58) now there are also individuals in those companies that may think that you know you know ethics and social values important so humans can temper a bit that profit maximization incentive but but it's a very strong one um as far as I'm concerned I didn't write the letter I signed it I thought it was a really good way to call attention to the problem to the general public to governments um and even people in companies have been discussing behind closed doors a lot since the letter so it has worked in that sense which is

(11:36) you know I think the real reason why I signed it did you Professor Harare think when you signed it that anybody would take a pause not the companies for the same reasons but we hoped at least to draw the attention of the public and of the governments because ultimately it's responsibility of governments to regulate this very dangerous development um unfortunately what we saw in in recent years is that the political discussion is just not there if you look at the main political issues that uh politicians that the parties are

(12:15) talking about they are not talking about this at all and this should be one of the top issues in every election campaigns now because it again it's not just the existential dangers down the line it's only also immediate concerns of everyday life it's our jobs it's who is making decisions about our life you apply to a bank to get a loan increasingly it's an AI making the decision about you you apply to University you apply to an employer to get a job increasingly it's an AI making the decision and you don't even under

(12:51) you can't even understand if they reject you why they rejected you and this should be again at the top of our political priorities and at least my Hope was that this letter would help push the issue to the center of attention and I'm glad to see that it's at least to some extent succeeded do you agree Professor bengio that there is a political malaise around this at this moment yeah um politicians for the most part want to talk about positive things actually I think it's the case also of a lot of AI experts that they they want to

(13:33) stick to the the good side of AI but I mean life is what it really is and and we need to face the the dangers the sooner the better um so I think this it's going to take time for politicians to to accept that they they have a responsibility here not just to be thinking about the economic benefits or the benefits to health care and so on of AI but also that they have a duty to protect the public from the short term and long-term risks Ferrari there's some responsibility for the media to ask politicians that

(14:12) question you know there's some culpability in the job I do as well um I think as a result though of both the letter and Jeffrey Hinton coming out and saying what he did it seems especially South of the Border where to where we are right now in the United States there did seem to be a heightened awareness among politicians they held a congressional hearing and even during that hearing The Man Behind open AI Samuel Altman basically pleaded with the government there with government representatives to regulate AI more how

(14:44) significant do you think that is well I think it's very significant and again we need to do it quickly and we when we talk about regulation we need to differentiate between regulating the development of AI the research in Laboratories and so forth and regulating the deployment into the public sphere and I think it's it's more urgent and also easier to regulate the deployment there are some very simple rules that we need to make for instance that an AI cannot counterfeit humans that if you're talking with someone

(15:22) you need to know whether it's a human or an AI if we don't do that then the public conversation collapses and democracy cannot survive and it's just common sense the same way for thousands of years we had lows against counterfeiting money otherwise the financial system would have collapsed and even though it's it's quite easy today to counterfeit money people don't do it because they are afraid they will go to 20 years in jail we need the same kind of laws about faking humans or counterfeiting counterfeiting humans and

(15:59) similarly the same way you cannot release uh powerful new medicines of vehicles into the public sphere without going through safety checks and getting approval it should be the same with AI that okay you developed something in your in your laboratory before you can deploy it to the public sphere with potentially immense consequences for society it has to go through a similar process of of safety checks would you Professor bengio describe the deployment side of things as an unreal regulated pardon me area right now well

(16:40) not completely because we we have laws about data we have laws about communication but but they were not designed to deal with some of the problems uh we're talking about so for example there's currently nothing as far as I know against counterfeiting humans and uh it's interesting this counterfeiting example because counterfeiting money is as far as I know punished very very severely because the stakes are so high and I think it's it's similar regarding the regulation for these things um there's

(17:14) um there's a thing that worries me about the way things might be going on in the U.S um I think that is is that we need the regulatory body to have a lot of agility so um here there's a a bill in front of parliament in Canada which has that kind of structure where the law is principles based you know it states like even ethical principles that Ai and companies doing it should follow and it leaves the the details the rules to uh a government agency um and that's good because the the field is moving too quickly the

(17:57) nefarious users that we can think of now maybe you know they'll be different in six months from now um the science moves the technology moves the the market moves and we need a lot of agility from governments which is not what they recognize not their strong suit yeah I know but in in the US my understanding is that in the last few years governments have moved away from these kind of principles-based uh legislation and where you delegate power to some agency and instead try to have everything written down by Congress

(18:29) because they don't want to give any kind of control to some government agency I mean especially the Republicans and that might be a big obstacle for efficient government intervention after us what could be efficient government intervention Professor Harare you've talked a little bit about some simple things they could introduce for deployment what what else yeah so one other thing that they could introduce low-tech uh demands like that if you want to have a social media account you must go to some office made

(19:04) from stones and sign a piece of paper now it's very inefficient but the inefficiency is a feature not a bug we do it with other things a passport a driving license um it's possible to social media could also be you will need to go through this low-tech uh operation and this will immediately get rid of almost all the bots on on social media and going back to the problems of the collapse of democracy I think introducing such a low-tech demand would immediately get rid of most of the Bots and help us save democracy

(19:48) what we should remember is that we are now facing a paradox in countries like the USA also in my country in Israel we have the most sophisticated and Powerful information technology in history and yet people are no longer able to talk with one another people are no longer able to agree on anything even about who won the last elections or whether vaccines are good for you or bad for you so how is it possible that with such powerful Information Technology the conversation is collapsing something is wrong deeply wrong with at least the way

(20:27) we deploy our information technology and and we need to step back and think about it before we deploy even more powerful tools like AI which could result in in the complete collapse of the conversation again if you're having a discussion about the elections with somebody and you can't tell whether it's an AI or a human that's the end of democracy because for a human being it's pointless to waste time trying to change the mind of a boat the bot doesn't have a mind but for the bot every minute it

(21:05) spends with talking with me it gets to know me better it builds even intimacy with me and then it's easier for the bot to change my views we have known for a couple of years that there is a battle for attention going on in social media now with the new generation of AI the Battlefront is shifting from attention to intimacy we are likely to be innocent if we don't regulate it we are likely to be in a situation when you have millions maybe billions of AI agents trying to gain our intimacy because that's the easiest way to then convince

(21:46) us to buy a product or vote for a politician or whatever and if we allow this to happen it will lead to social collapse as Professor Harare is describing that I was thinking professor bengio of the debate in the discussion around just the first battle that he describes social media and its impact on democracy and the pushback that there is to more government intervention versus the you know up against the idea of free speech and the political debate that occurred just on that forget about all the Ali that we're talking about and the

(22:20) threat that that poses but just around that there seems to be and somewhat understandably given the pushback a hesitancy for governments to insert themselves too far into this uh how how troubling is that given the prospects that Leia had well there's a kind of uh either Vish I mean Vicious Circle um around democracy itself so if we had a healthier democracy would be in a better position to protect it and you know why is it that democracy is so fragile one of the factors that I think social media and eventually AI might you

(23:03) know really ride on is that humans are quite easily influential and we're not always rational we're you know influenced by as as you've all was saying you know the the relationship we have with a person and if we value that person we'll kind of buy what they say without really checking if it makes sense and democracy has existed for thousands of years as you know as much as democracy has but it could be Amplified by by social media and it could be uh further Amplified with AI um so yeah how do we

(23:44) how do we and so so I think that's what's making it difficult to legislate and regulate properly because there are conflicting interests at stake if you're if you're a corporation you want the least possible uh you know uh barriers that gonna slow you down I like the low tax uh um suggestions very much but even but companies will fight against that because it will mean less Revenue right but even people might because maybe I can just yes go for it maybe I'll just say about the Free Speech issue because this is a small

(24:18) smoke screen used by the big Corporation it's not a free speech issue because Bots don't have free speech both don't have rights humans have rights we have the right to freedom of speech so you know to to ban a human being from Twitter or to ban a human being from Facebook that's a difficult issue but to ban the Bots from Facebook or Twitter that's easy I mean they don't have rights who is going to defend the freedom of speech of bots Elon Musk perhaps [Applause] I got I guess I wonder and I and I

(25:01) completely understand the point around it being a smokescreen driven by corporate interests but it has been convincing to certain subsets of various political ideologies right and how do you say to someone who is used to now having the ease of doing whatever they want on social media that actually you have to go to this building and I mean it sounds amazing wouldn't it be wonderful to know people you're talking to on social media are real people but convincing them to do so like what is the imperative how do we do that that's

(25:29) why we have the conversation we are having today we we need enough people to understand the risks because losing democracy might be losing a lot of the you know uh well-being and freedoms and um uh you know all of the good things that that liberal democracies have brought uh we might lose a lot of that with with uh losing democracy so we need people to understand that and it's it's going to take time it's going to take many discussions but that's the way I can see the threat of losing something dear to

(26:09) us is probably the best way to uh you know bring us to make choices that you know we feel bad like going to this building to get our account approved there are other I don't have time everyone wants making suggestions more extreme suggestions about how to counter the existential threat and the the level to which there needs to be government government intervention even going so far as to suggest some sort of world government and the reason I ask you about that is because I think the same people who will um be upset about any infringement as

(26:46) they view it on Free Speech will also go right to that example to say why government shouldn't intervene so what are your thoughts on that well world government is a terrible idea it's impractical and even if it could be done it shouldn't um there are also ideas about that we have to set up some kind of authoritarian or totalitarian system in order to to prevent to save humankind but I think the only realistic way to save humankind is first of all to save democracy because totalitarian systems will be much much worse than democracies

(27:23) when it comes to regulating Ai and keeping it under control the traditional problem of totalitarian regimes is that they tend to believe in their own infallibility they never make mistakes and they don't have any strong self-correcting mechanisms mechanisms for identifying and correcting their own mistakes and with the totalitarian regime or some kind of super powerful World Government the temptation of that system to give too much power to an AI and then not be able to regulate it it will be almost irresistible

(28:01) and once the totalitarian regime gives gives power to an AI there will not be any self-correcting mechanism that can point out the mistakes that the system will inevitably make and correct them and it should be very clear that AI is not infallible it has a lot of of of power it can process a lot of information but information isn't truth and there is a long way leading from information to truth and to wisdom and if we give too much power to to an AI it's it is bound to make mistakes only democracies have this kind of

(28:45) checks and balances that allow them to try something and if it doesn't work to identify the mistake and correct it why do you think Professor bengio there are people making suggestions that go that far and what is your assessment of them well it it is a simple idea that if we all were some kind of camera that watches everything we do and some central government uh using AI would check that we don't do something really dangerous for Humanity like sending the command to kill everyone or just the bad guys whatever

(29:26) um we would be safer it's an appealing idea if you forget about everything that you've all has been telling us but on the surface it says oh okay so how do we um make sure that no one on earth knowing the recipe for you know getting an AI into this stage where it could be dangerous for Humanity would actually execute the band commands like let's say we all had access to that knowledge how do we make sure that no one will you know do it so I think it it's I don't think it's the right solution

(30:05) but it you know it's understandable you would say well let's make sure everyone doesn't do it because they are you know wired up watched and they are controlled and they can't do their freedom is restricted in very strong ways um it might be a tempting scenario but I think it's very dangerous uh as UL has been explaining and I would like us instead to Think Through what are our options like are there like other ways that we could organize Society so that it would be safe in the sense of safe Ai and that we preserve

(30:40) our human values human rights human dignity and democracy it's a it's a difficult one but I think it's worth spending our best brains time on these kinds of questions what does your intuition tell you uh if I could follow up with you Professor bengio on whether that win-win-win scenario exists well I would say it's a it would be difficult especially as if we don't have enough time but when you get discouraged about something that seems so much bigger than your own little you know human person

(31:23) you have to remember that we can only do our best so I think our moral duty even if the odds seem you know bad is to try is to to think through this so think about um people have been fighting for years decades for us to change our ways to to deal with climate change it it may seem like a desperate cause or at least it has for many years and hopefully it's getting better but but people continue fighting right continue trying and that's what we have to do with the challenge of of um Ai and it's it's worth it because we

(32:07) might bring uh you know good health to everyone uh education uh solve climate change and we might even be able to reform Society completely in ways that um maybe you would call Utopia but but might give us all of these good things um and and even better than the current kind of society we have in terms of democracy and well-being of people do you see Professor Harare a reason for optimism in the same vein do you feel that the worst case scenario is avoidable wait wait I didn't say that I was Optimist I said we have how can I

(32:44) characterize it as such I I said that I don't want to be either Optimist or or pessimist I said that we have a duty a moral duty to go for The Optimist solution and try to make it happen and so I asking based on that do you think that is possible Professor Harare yeah I think basically we created the problem we still have the power to to solve it not for many years because the AI will take power away from us but at present we still have the power and maybe one more General thought about what we just heard is you know we about

(33:24) intelligence and all this talk about Intelligence being such a good thing we are the most intelligent beings on Earth so far and look where our intelligence has brought us that we are now discussing that because of the things that we invent we are now discussing these scenarios like putting a come around every person and destroying human freedom and human privacy in order to protect us from something we invented with our intelligence allegedly in order to provide us with Better Health Care and other stuff I think one of the

(34:01) things about intelligence is that the most intelligent agents at the same time also tend to be the most stupid agents around the frogs wouldn't do it something like we just did so also when we think about artificial intelligence it's also probably going to be incredibly stupid in some of the things it does it's not necessarily optimistic but it's just to kind of frame the discussion about artificial intelligence and intelligence in general uh uh for a moment and again ultimately the problem is us not the AI like what we've just

(34:43) discussed is for instance the collapse of democracy it's mostly bad human actors who are using AI in a malevolent way so I I think if if we spend as much time on developing on understanding on and developing our own minds as we spend on developing AI we would be safe I I completely agree um and I would go farther than that I think it's the only way if we want to if we want to you know avoid the Big Brother totalitarian regime that we talked about the only way to converge to something that uh would preserve

(35:28) um our values is by changing humans and I I think it is possible so there are subsets of humans that are more ethical wiser uh better understand themselves uh more compassionate more rational so we know it is possible how do we how do we bring that to the whole of humanity that's the Utopia I was talking about do you just to follow up on that as you describe that I think of a lot of people who maybe have just seen the headlines about Ai and who hear that and think that this is just a way to completely reorient or change society

(36:16) and they get scared off by that is there is there a risk uh to talking about that as the solution and then alienating people who need to buy into whatever solution is actually implemented yeah and I understand um people are close to me who are concerned that by focusing the discussion on these catastrophic risks to democracy or Humanity we are not focusing on the current issues with Society the current issues with AI with with buyers and um and discrimination for example and the injustices that exists so we we need to do all of those things at

(37:06) the same time um but but but maybe there's a convergence of the solutions Professor Harare what are what are your thoughts on how to navigate getting buy-in both politically and among the public to making sure we are committed as a society to trying to solve the issue and first of all we need to focus society's attention on these problems it's not uh in order you know it's not alarmists it's not in order to hide from us other problems it's but it's really again aside from the long-term existential risk

(37:49) many of our most immediate problems uh in the economy in society can get much much worse because of AI we didn't talk much about the job market but this should be a very Central concern for everybody and I don't think that AI will destroy all jobs it will destroy some jobs it will create a lot of new jobs but a transition again retraining people this will be very difficult again remember that Hitler Rose to power because of three years of something like 25 unemployment in Germany so um even if in 20 years it will be

(38:28) okay we don't have 20 years what do you do with 20 unemployment for three four years and with regard to politics I think there is also a deficiency of people within the political system with a very deep understanding of the technology and its impact because most of the people who understand the technology they want to be the next Zuckerberg or the next Elon Musk they don't go to politics in order to raise the awareness of the public and regulate it they go to business to make billions we have some few exceptions for instance

(39:13) Audrey tongue who is the digital minister of Taiwan uh she was a hacker she could have started she could have went the business way to to have a startup and make billions and she decided no I will go to politics and I will help the public and the political system understand and regulate these explosive potential so we basically need fewer zuckerbergs and Elon musks and more Audrey tongues that's definitely one way to put it um I I wanted to to pick up Professor bengio on a point that Professor Harare made around jobs because I do think the

(39:56) discussion around what AI means for my job for your job for the audience's jobs is a very tangible way to ground the thread it poses just like with the existential threat climate change poses looking at the floods in Quebec the fires in Alberta certainly grounds it for for us and ultimately I think changes the impetus for politicians to act um your your what is your how would you characterize the threat or lack thereof AI poses to our jobs uh well I'm not an economist um so I've tried to read a little bit on

(40:29) this and I see that the you know various camps as usual in the economy um make very different claims so so on the one hand there are people who are like uh for example there was a recent study coming out of openai and and academics suggesting a large fraction of the jobs would be say modified uh and then would you you get a lot more productivity which means we would either have less people doing those jobs or we would do more of those things so think about programming is a good example if it was if we could do our programming

(41:07) job twice as fast because of AI tools do we are we going to have half as many programmers are we going to have um you know the same number doing twice as much good so so that's uh you know it's hard to to predict those things and also one of the arguments I've heard on the side of that's not worry is um societies change slowly and even if we had the technology for something it might take years or sometimes decades for people to really bring it in society and and have a big impact on the job market

(41:46) now okay they're also the other arguments that are saying AI is different once you have a system that essentially uh does their work better so think about right now we're talking about all the jobs that you can do by just manipulating language like you just you know uh email and then social media and uh and so on um and and working with databases and stuff that it's it's likely that those kinds of jobs could be done better fairly quickly in many sectors and whether companies are going to be able

(42:24) to do this quickly or not it's really hard to predict but if it if if they do um we could really have these transition problems that Yuval is talking about do you think Professor Harare that the jobs part of the discussion is a way to ground this for people who are maybe agnostic at this point absolutely because this makes it very very clear what if a bot or an AI is coming from my job this grabs people as people's attention immediately and again it's not a simplistic idea that there will be no longer any jobs for humans

(43:00) there will be a lot of new jobs but the transition is difficult how do you retrain people especially if you need to take into account the the global uh considerations because the AI Revolution is being led by a very small number of countries who are likely to become extremely rich and Powerful because of that whereas it could destroy the economy of less developed countries even if you think about something you know like textile industry what happens to the economy of Honduras or Bangladesh if it becomes cheaper to produce textiles

(43:41) in Canada in the US than in Honduras Guatemala in Bangladesh so do you really think you'll be able to retrain millions of textile workers in these countries to being I don't know designer of virtual reality Games and who will pay for the retraining I mean in in the advanced developed countries the the gains from the AI Revolution will enable the government hopefully to cushion the blow for the people who would lose their jobs enable and will enable them to retrain but I don't see a U.S government leaving

(44:21) taxes on uh Google and Facebook and Microsoft and sending them money to Honduras or to Bangladesh to help the people there retrain and cope so we could see a kind of new like with the Industrial Revolution in the 19th century that led to very few countries basically conquering and dominating the whole world this could happen again within a very short time due to the automation Revolution and the AI Revolution and again if you add to that it's not just the economy it's also the type of political control that you can get

(45:04) from a harvesting all the world's data and analyzing it uh you know previously to control a country you needed to send in the soldiers now increasingly you just need to take out the data what happens to a small country somewhere when the entire personal records medical records whatever of every politician and journalists and judge and military officer is held by somebody in Silicon Valley or in China is it still an independent country or did it become a kind of data colony so these are the the kind of immediate

(45:48) dangers that I think uh should be clear to to any citizen no matter what their views are on the long-term existential risks of AI well on that that's probably a perfect note to finish our discussion because I think we are all a lot more clear after this discussion on exactly what you just laid out what both of you have laid out I can't thank you both enough for all the information you provided us with today thank you it's been a pleasure thank you very much it's been a real pleasure and I do feel

(46:27) like instead of crying ourselves to sleep we're armed with a bit more knowledge now so that's a good thing thank you all very much for your attention today thank you thank you professor