Bostrom:

If you start to think it through, it's not just human economic labor that would become superfluous in this scenario where AI succeeds and we end up in what I call a solved world, but all kinds of other instrumental efforts as well. And then that raises these questions about yeah, like, could you still live a good life in those circumstances and, if so, what that would look like.

Craig: 0:22

Hi, I'm Craig Smith and this is Eye on AI. Today I speak with Nick Bostrom, the philosopher and author of Superintelligence, Paths Dangers Strategies. In Superintelligence. Bostrom explored the potential future of artificial intelligence and the existential risks it may pose to humanity. His work has been instrumental in shaping the discourse around AI safety and has influenced many of the leading minds in the field. Today we'll be discussing his latest book, Deep Utopia, which asks the question, how will humans derive meaning in a world that is technologically solved, where AI and robotics can do all tasks better than humans. It's a fascinating, if disturbing discussion. I hope you find it as fascinating as I did.

Craig:

Hi, I wanted to jump in and give a shout out to our sponsor, Netsuite by Oracle. A little quick math: the less your business spends on operations, on multiple systems, on delivering your product or service, the more margin you have and the more money you keep. But with higher expenses on materials, employees, distribution and borrowing, everything costs more. So to reduce costs and headaches, smart businesses are graduating to NetSuite by Oracle. Netsuite is the number one cloud financial system, bringing accounting, financial management, inventory, and HR into one platform and one source of truth. With NetSuite, you reduce IT costs because NetSuite lives in the cloud, with no hardware required, accessed from anywhere. You cut the cost of maintaining multiple systems. You improve efficiency by bringing all of your business processes into one platform, slashing manual tasks and errors. Over 37,000 companies have already made the move. So do the math. See how you'll profit with NetSuite. Now, through April 15th, Netsuite is offering a one-of-a-kind flexible financing program.

Head to [Netsuite.com/EYEONAI](http://netsuite.com/EYEONAI). EYEONAI, all run together: E-Y-E-O-N-A-I. That's netsuite.com/EYEONAI, again netsuite.com/EYEONAI.

Craig:

Let’s go ahead with, and introduce yourself.

Bostrom:

Well, I'm Nick Bostrom, I'm currently a professor at Oxford University, where I've been since 2003. Then, before that, in the US. for a bit, did my PhD in London. I grew up in Sweden.

Craig:

You're best known, in the United States at least, for your book Superintelligence, which didn't directly address the question of the meaning of life in a world dominated by superintelligence. It was more about the risks and managing the development of such a technology, but before we talk about your book, can you talk about the recent advances in AI since you wrote Superintelligence? Because it's certainly come a long way and a lot of leading researchers are now talking about AGI within the decade and superintelligence presumably would follow fairly closely on AGI.

Bostrom:

Yeah, it's been a remarkable ride the last several years. Superintelligence came out in 2014 and that roughly was also when the deep learning revolution began, and so we've seen a string of apparently accelerating advances since then. It's a very different era now.

Bostrom:

Back when I wrote Superintelligence, the whole idea of AI eventually succeeding at becoming fully general and being able to do everything humans can do was already a very radical and sort of fringe thought, certainly in academia, mostly dismissed as sort of science fiction or idle speculation.

Bostrom:

And, of course, the further step, that it could then become super intelligent even more so, and I wrote the book to try to draw attention to the set of risks that would arise if we do attain transformative AI and super intelligence. Now, of course, that has become completely mainstream. All the leading AI labs have research groups focusing on AI alignments, trying to develop scalable methods for AI control that could continue to work, even at sort of arbitrary levels of capability. And just in the last couple of years, we've seen a big shift in the public discourse, and even top-level policymakers are now paying attention to the possibility of transformative AI, including existential risks, with statements coming out of the White House and the Global AI Summit in the UK, and really around the world. So yeah, a lot has shifted in terms of how this set of issues is viewed.

Craig:

Has that changed your view on superintelligence or the possible advent of superintelligence and the timeline?

Bostrom:

I guess I gain some confidence just from seeing that things are actually playing out.

Bostrom:

No matter how, sort of coherent some argument seems theoretically, it's hard to be really convinced in your gut until you also start seeing it around you, or at least see a whole bunch of other people also sharing your beliefs. Increasingly I think that there is more granularity in the precise forms that AGI might take; I mean, with the rise of the current large transformer models, and it's strikingly how anthropomorphic these AI systems are in many ways that it's almost like they have a psychology like a human, like with foibles. Sometimes they need to be encouraged in the prompt to do their best, which is kind of weird from, certainly from what the old school AI used to be, right? It wasn't that you needed to give it, sort of emotional support to encourage it to perform and we also see leading edge systems being very compute intensive, and so that seemed always kind of likely. But it seems more likely now that we see that the frontier systems are kind of gaining in capability as you throw more compute at them. And so it does certainly shift probability mass a little bit, and concentrates it on certain kinds of scenarios. And we might be on the sort of slightly shorter end of the timeline. It was always a wide probability distribution and roughly, you know, we are more or less kind of on track, I guess, of what I would have thought would be the sort of the median of that probability distribution, but maybe slightly faster than expected.

Craig:

Yeah, I mean there's I don't remember whether it was in your book or Kurzweil's but there's this hockey stick graph that, you know, and at a certain point the machine intelligence surpasses human intelligence and it's beginning to feel like we're on that curve, the more acute curve now than we have been. Do you have a view on that, of where we are on that graph, if you believe in the graph?

Bostrom:

Yeah, well, I certainly think things seem to be moving quite rapidly, and I mean, we still have uncertainty but, yeah, I think, non-trivial probability mass on, sort of single year timelines. But with remaining uncertainty it could take longer as well, but it's been, I guess, surprisingly smooth.

Bostrom:

In terms of the growth of capabilities, it seems, kind of, year by year, they have become incrementally better, um, so that might continue, in which case we might have more of a, sort of a median speed takeoff. I don't think we should be too confident in that, though, because that could still be a transition point, when you have AI systems that may be incrementally and slowly improve; but then, at a certain point, they become capable of very substantially driving further advances, particularly, when you can, sort of, fully substitute for the labor of human AI researchers and chip designers, et cetera. Then it might still be that it kind of goes smoothly, but fast up to a certain point. And then you might get an intelligence explosion when you hit this kind of strong feedback loop that I think remains in the cards.

Craig:

Yeah, how do you feel about the threat debate which your book contributed to? I mean, I see the paperclip analogy all the time. That's sort of a nightmare scenario; but since Max Tegmark's group came out with the pause letter, I mean, that seemed to focus regulators and researchers on safety, but at the same time it spread a lot of fear and mistrust in the public, which I found unfortunate.

Bostrom:

Yeah, that's maybe the most dramatic shift since 2014, the mainstreaming of this AI doom concern, which I think is broadly for the good, in that there is now actually research done to try to find solutions to this, whereas back then there were a handful of people scattered around the world communicating on the internet. But now there is more resourcing for that and talent flowing into that, which is positive, I think. But it's reached a point where I'm thinking one might soon start to worry a little bit also about the worry overshooting its target. I don't think we're quite there yet.

Bostrom:

I still think a higher level of concern than we currently have would be optimal, but humans being herd animals, like, once a stampede starts, it's very difficult to fine-tune it.

Bostrom:

You can maybe set it off, but you can't necessarily call it back when it's reached the optimal point.

Bostrom:

So I still think it's unlikely but less unlikely than two years ago that we could end up somehow on a trajectory where AI is never developed because it is, you know, either permanently banned or sort of a moratorium that delays it long enough that we then destroy ourselves in some other way instead, without ever even getting the chance to roll the dice with AI. It's like a fine balance and it would be good if, instead of being super polarized into, sort of, AI throw out all the stops, let's just brush forward. There is no risk at all, at least no existential risk. That's just silly on the one hand, and then, on the other hand, I wouldn't call it doom-mongering, because I do think that the existential risks are very real; but I think at some point there is a certain amount of risk that we will need to take in order to have a shot at realizing the benefits, and there are risks also not to going forward with this. That should be entered into the balance.

Craig:

Yeah, the latest book, Deep Utopia, considers, I mean, the main theme is, how do humans derive meaning and purpose in a world that's technologically solved by AI? How did that emerge from your work on Superintelligence? Is that, sort of, the next step in your thinking, or is it because the progress in AI has been so rapid that you started thinking about what happens if we really do reach technological maturity, as you call it?

Bostrom:

Both sides have always been there. In my outlook, Superintelligence focused mostly on the downsides, what could go wrong, because at the time that seemed the most urgent thing to bring attention to, to get a clear understanding of where the pitfalls were so that we could then hopefully steer clear of them, and to develop some conceptual framework to allow people to start to make progress on this, to develop solutions to scalable alignment, et cetera. Now I think that there is, like, much broader recognition of that. So I think there's less value perhaps in just repeatedly harping on the risks, because there are a lot of people now actually recognizing that and working on it. And I think there is, kind of, still a certain naivete and blind spot in our conception of what happens if things go right with AI.

Bostrom:

And we've begun hearing some conversations about impacts on the labor market. You know, what if AIs and robots can take over more human jobs? You know, will we need some support for the unemployed, like in the more radical conceptions? Even the question, what if AIs could do all economically valuable labor at some point in the future? You know, maybe with some carve-outs? Then what? Do we need a universal basic income? How would the education system need to change? And I think we will hear more of those debates, but they still, I think, take only one step on a sort of path that allows many more steps to be taken, because, if you start to think it through, it's not just human, economic labor that would become superfluous in this scenario where AI succeeds and we end up in what I call a solved world, but all kinds of other instrumental efforts as well.

Bostrom:

So, if you imagine somebody who doesn't work today, maybe they have entered early retirement, let's suppose they are rich and so they don't need to work, but they presumably still spend a large fraction of their day engaged in various activities which has the structure that they do a certain thing, put in a certain effort in order to achieve a certain outcome, separate from the effort itself, from the trivial. Bill Gates brushes his teeth every morning and he has to do that. That's the only way that he will have, you know, a healthy mouth right, and if he wants to be fit, he would have to go to the gym. If you want to have good friendships, you have to put time and effort into, you know, talking with your friends, understanding them, them being there for them if they hit a difficult spot, et cetera. If you want to have the particular kinds of items that you would enjoy, you have to maybe spend some time going shopping or looking for different things to find exactly. But all of these activities as well could be automated at technological maturity or other advanced technologies would make them unnecessary. So if, instead of having to go to the gym to be fit, you just pop the pill, and it would have the same physiological effects, then would there still be any point in going to the gym? If, instead of going shopping which some people enjoy, you had recommender systems that would be able to pick out better items and then just order them, it might kind of undermine some of the point and purpose in the shopping activity.

Bostrom:

Because if, at the end of the day, you end up with worse items than if you had just let the AI do its thing, it kind of at least puts a question mark about that. It would kind of seem to lose its allure. And if you go through and there's a bunch of case studies, it seems like a very large fraction of what fill our lives today when we are on holiday or retired or when we don't, even when we don't work is of this instrumental character. And at technological maturity, if things go well, we would enter, I think, a post-instrumental condition where, to first approximation, it would seem that there are no instrumental reasons for us to do anything, to put out any effort. Um, and then that raises these questions about, yeah, like, could you still live a good life in those circumstances and, if so, what that would look like.

Craig:

You use the example of Bill Gates or someone that's wealthy and doesn't need to work, but there are a lot of people already who are unemployed or retired, laying on their sofas, flicking through television channels and experiencing either depression or, as you called it, a deep existential malaise. And I mean, this is something that concerns me about the United States, the state of education, and education takes effort. It's uncomfortable to carve those neural pathways until you reach a certain critical mass, when then it becomes more enjoyable. And large swathes of the American public are pretty uneducated and easily manipulated and not content with their lives and all of that.

Craig:

So, if the world is technologically solved, I don't see the motivation for people to go through the difficulty of getting educated, and without education to me it looks like a very dark future. Unless, as you talk at one point in the book about a very dark future, neurotechnological interventions, you know sort of uh, stimulating the brain in certain ways to overcome lack of purpose or lack of motivation, or, as you intimate at the end of the book, using recreational drugs to alleviate that malaise. How, how do you see that, looking at, sort of, the common man as opposed to the educated elite? How do you see that developing in the technologically mature age?

Bostrom:

Yeah, so technological maturity is this concept I have for a condition where basically all technologies that are physically possible and that are, in principle, feasible to develop have been developed.

Bostrom:

With super intelligence that might be attained relatively quickly because the super intelligence would then be doing the further R&D, and yeah, you would then enter this condition of technological maturity, you shouldn't just think of it as AIs and robots but all kinds of other things. So, with learning, it's one example of an activity where, right now, you have to put in effort in order to achieve a certain outcome. Like, if you want to understand mathematics, the only way to get there is by first studying mathematics, and you have to put in the time to read the textbooks and do the exercises, and there is no shortcut. You can't pay to know mathematics without, you know, studying it. But at technological maturity, though, I think there would be a shortcut. That would be technologies that would enable, uh, you directly to have your brain altered into a new condition where you possess arbitrarily high levels of knowledge of mathematics. If we're still organic, you could imagine some either like a neural interface or some sort of swarm of nanobots that go around and changing the weights of your synopsis, all sort of masterminded by some unfathomable super intelligence. If you are uploaded at that point, then it's presumably even easier to go in and edit the file of your, you know, the weight matrix of your brain, and this starts to show just how deep this problem goes. Like it's not just that there are a few, sort of, chores that we no longer have to do, but even very fundamental things like learning and understanding. Our own personality also becomes malleable in this condition, like in a plastic world. And so, yeah, I don't think the problem would be that humans, unless they are somehow still finding the motivation to sit through 20 years of education, would remain incompetent and ignorant. I think there would be shortcuts to that, but those wouldn't necessarily involve an extended period of you, yourself putting out effort, unless you chose to do it the hard way.

Bostrom:

So the book basically brackets all these practicalities. It doesn't talk about how we would get from here to there, or what the risks are, or what the policy solutions are, or how society should be organized. All of those are, of course, super important questions, but not discussed in this book. I basically just assume that we reach this condition of a solved world, in order, then, to be able to actually explore the question of what happens then. How is it possible for a human to have a future trajectory that leads there, that still constitutes a good life? And I think ultimately, probably the answer is hopeful, but it is not, there is a sense of ambivalence.

Bostrom:

I think it would require, like, it will feel quite alien, probably to us now, at least, until one gets used to it, and a lot of the things that kind of define the current human condition would likely go by the wayside.

Bostrom:

We are so shaped by these instrumental necessities that we confront in our lives that it's hard to imagine what life would be without them. It's almost like a little a bug that has an exoskeleton that holds these sort of squishy parts in place, and if you imagine taking that off, then you know what becomes of the insect. Similarly, these instrumental constraints are kind of an exoskeleton for our souls, and if we remove that by being able to automate everything and having a plastic world, then one possible outcome is that we become kind of blobs, pleasure blobs, like, kind of, some super drug is giving us immense pleasure, but we have no real character, no structure to our existence, no effort, and that is one possible vision of a utopia. I think it is easy to dismiss it very quickly. I think it's actually a very deep question whether there is something better. Probably there is, and that's kind of what the book explores, the more subtle values that could give shape to our lives.

Craig:

I guess the idea of being able to upload education into your brain, is a hill beyond what I was looking at because, short of that is where I see the dilemma. If you could do that, then you could engineer happiness, you can engineer fulfilment.

Bostrom:

Yeah, subjective happiness, like, any feeling you want. So this is another instance of something that would fill the time of a leisured person today. There's a lot of stuff they have to do, put out effort in order to experience joy and pleasure, and subjective happiness, but there would be shortcuts to this, in fact they already are today. It's just they're not very good, like drugs that could sort of temporarily give somebody a great level of subjective well-being without putting out an effort. Then it's just that they come with side effects etc. But, obviously it could develop super drugs in the future that would not be addictive or have direct ways of manipulating or altering the neural pathways that underlie our, sort of, hedonic experiences. So, it becomes quite a radical quest.

Bostrom:

It really starts to, when you sort of think, take this to the extreme, you really confront these very old questions that philosophers have wrestled with. You know, about the ultimate value that we have, the meaning of life, the purpose of human existence. You see them, though, from a very different angle when you look at them under these extreme conditions postulated in this book. It's almost like a particle accelerator where you smash nucleons together at extremely high speeds and these immense energies. You then see their constituents, you may see quarks flying around and stuff like that and then you can infer that those same constituencies are present in ordinary matter at other times as well.

Bostrom:

It's just, we can't observe it. And so, similarly here, you can think of it as a philosophical thought experiment. You consider this very radical condition where the world is completely plastic, and then you sort of see what happens to our values in that scenario. It's like smashing them together and seeing that we are forced to look at their basic constituencies. Now you could read the book as a philosophical thought experiment like that, which could be of interest, even if this was guaranteed never to happen. To me, though, it is also not at all impossible that we will eventually reach something like this if things go well. So there is also a kind of more practical motivation for thinking about these questions.

Craig:

Yeah, although if, sort of, neurological engineering, if that's the end point, then a lot of these questions are kind of moot because you can solve them through engineering of the brain. But short of that is when it seems to me there's a crisis. And also I want to ask, why do you think of this as a hopeful or optimistic book?

Bostrom:

Well, I mean, the reader can decide for themselves. It's not a book of conclusions so much as it's a book about, that tries to help the reader to bring certain questions into focus. And, yeah, I think it's ultimately a hopeful book because I think there is something extremely valuable that could be constructed in this condition of a plastic world that will be very worth having.

Bostrom:

Um, but I'm not pretending it's very obvious. I mean, there is an aspect to it that is obvious. Which is the most obvious is, like, all the horrific negatives that pervade the current human condition, properly, and bombs blowing up in neighborhoods and ruining people's lives, and somebody getting a call and their spouse has died in a car accident, like that's just kind of. As well as the smaller things that we shouldn't forget that just makes a lot of ordinary existence a drag, like the little headache that you have or the little, like, all these kind of things that just um, the boredom of many jobs, eight hours every day, feeling tired and doing something pointless that you don't enjoy. Just getting rid of that all in itself, I think, would be an enormous boon, um, and then there's the extra on top of that, which I think, could be immensely valuable as well.

Craig:

At some point in the book you were talking about Thaddeus Metz's philosophy or his theory of meaning and one of the things that he talks about is striving and then achieving a goal and and a lot of things you just mentioned. You know, certainly there's a lot of pointlessness to it, but maybe happiness and fulfillment exists in contrast to despair, and disaster, and frustration, those things - then fulfillment and happiness may lose some of its value on a relative scale.

Bostrom:

I mean, I know some people who are just, kind of basically happy all the time. They have personalities that are naturally sort of cheerful and uh, they, I don't know, they seem pretty happy and I'm not sure it's taken away from them. And other people who are, like, depressed all the time, their suffering is not less because they might not have the bright moments to contrast it with - I think so I don't know about that and certainly the subjective aspects would be completely up for choice in this world.

Bostrom: 34:12

Now I should say, at the first approximation, like I said, the challenge becomes It looks like there is nothing we need to do for instrumental reasons in this condition. Now I think there are important qualifications to that. If you look more closely, there are various activities where there might still be a need for human effort. It might be most, like, obvious in cases where you have consumers that have a preference that a certain thing be done by humans and we already see this. Maybe people prefer to see humans compete rather than robots in an athletic race or something like that. Sometimes people might pay more for a trinket if it was done by a certain category of person that they like, like in some indigenous community or like, rather than in a sweatshop in Malaysia. And so already we sometimes have sort of products that achieve higher value, not just because of what they are but because of the specific preference that generated them. And we can have preferences over that specific preference that generated them. And so to the extent that we do, even at technological maturity, a preference that certain tasks be performed by humans, that could then create instrumental reasons for other humans to perform those tasks.

Bostrom:

And I think this generalizes. And there are many more subtle forms of this where there might be cultural traditions that we want to uphold and the only way to do that is by doing a bunch of stuff ourselves. It wouldn't count as carrying on the tradition if we just programmed a robot to do the ceremonies and to, you know, do the mass or whatever, the spring, rite of spring, whatever the different things. So as you start to zoom in, I think there are possibly a lot of reasons we would have for doing things ourselves. And more generally, I think there might be a bunch of quiet values that we can't hear very much right now because we're living in this den of screaming moral imperatives all around us. There's, like, you've got to make a living, otherwise you've got to be kicked out of your apartment. You've got to help this needy person because they're really suffering, you've got to do this, that, and the other, and these should, I think, have priority in our current condition.

Bostrom:

But if we imagine somehow all of that went away, then I think there are these more subtle values, like some of them quite almost of aesthetic quality, that then ought to have a bigger influence on how we spend our time and our behavior. It's almost like when you go out at night, your pupils dilate and you realize that there is this big starry sky up there and it was there all along. You just couldn't see it because there was the blazing sun right, like, the kind of, just took up all the focus of your visual system. But in this different situation where, yeah, the sort of loud moral imperatives quiet down and die away, then I think we should sort of sensitize ourselves more to these subtler reasons for doing things, and that might reveal a whole iridescent sky of, sort of, values that we are more or less oblivious to at present. That could give structure in utopia.

Craig:

Wow, that's, yeah. I mean that you remove all the imperatives, that there will be motivation and that there wouldn't just be mass depression.

Bostrom:

Yeah well, I mean, if people don't want to be depressed in this condition, that would be trivial to change, unless they also had a preference for not being changed. So that's another thing that could create sort of instrumental obstacles, is even if there is a shortcut but there's some other value, that sort of blocks you from taking advantage of it. Um, but again, and I think this is like a little bit of a Ponza Sinora, like that the idea that we would be subjectively depressed or unmotivated and wishing that we had more motivation or felt better, but not being able to get there, I think is implausible in this scenario, because we already have crude ways of manipulating human mood and motivation. And clearly, if you had complete control over the entire neurochemistry of our brains with the kind of super-duper neuroscience that would be developed by artificial super intellects, then that would be amongst the easier things to manipulate and it would be more like a kind of, you know, like whatever, a control panel with different knobs you could choose to fine-tune, to feel however much motivation you wanted to.

Bostrom:

And so the question then becomes rather, the values that guide our selection of A. Do you touch this control panel and manipulate your own internal psychology? If so, how or with what constraints? Because if your only criterion is, kind of, do you actually enjoy it at the end, then it would be easy to crank up, kind of, pleasure dial to the max, and then we would have this scenario where we would become, kind of, blobs and, as I said, it's not at all trivial to say that is not a good thing. But if one doesn't ultimately think that that's the best thing, then this is a much more complex discussion about the various subtle values that, kind of, come into view.

Craig:

Yeah, you talk about population growth, of reducing individuality and compelling life stories. Why do you think there would necessarily be a population growth?

Bostrom:

Because, you know, it wouldn't necessarily be a population growth. I mean, right now there's a population, I mean it's still growing, but kind of set up, right.

Bostrom:

I mean no, that's a very open question. I think if you did have a scenario with a lot of uncoordinated replicators, eventually you would have replicators that figure out the way to make a lot more of themselves, like with, kind of, the history of life which fills every niche and nook and cranny. You know, high on the mountains, deep in the ocean floor, in the desert, in the polar caps, like everywhere where there possibly could be life, there is, kind of, life. But you might imagine certain scenarios into the future achieving a, kind of, form of global coordination which would have the option of instead having, you know, slower population growth or no population growth and then having, sort of, per capita resource endowments increase instead. You could have both as well, because the future is very big. You could have a fair amount of population growth and still have average income rise enormously, but at some point obviously there are like ultimate physical limits.

Craig:

Yeah, I mean the thing on population growth; I've spoken to people who are very focused on longevity studies and engineering, you know, different forms combating aging, and that's something I always ask. Well, what happens if everybody can live to150 or 200, 500, that it's just going to overcrowd the earth, because even with a declining birth rate, any added birth is going to increase the population. I found it fascinating when you talked about the reduction of individuality or compelling life stories because there would be a, kind of, a reversion to the mean of what people are like or what they do. There would be not as many choices because technology would take care of so many things that needed to be done. Can you talk a little bit about that, about your thoughts on that?

Bostrom:

Yeah, well, I mean, I think there could be a lot more diversity in a technologically mature condition than there is now, just because, well, for many reasons; first, we could share the world with all kinds of digital minds that don't yet exist. They could come in an enormous variety, a much bigger variety than the space of biological minds that we have. Then, hopefully, we would take better care of the non-human animals that we are currently often mistreating and abusing horribly, and some of those might get uplifted in various ways so that they could, sort of, communicate in a, sort of shared animal-human AI society. Also, the set of forms that the human being could take could increase greatly with various kinds of enhancements or modifications. That could give us much more ability to shape not just the world around us but ourselves according to our deepest values. And then, yeah, just being able to continue to develop for longer than is currently vouchsafed to us.

Bostrom:

We kind of grow up for 20 years and then there's kind of stasis or slow growth for another few decades and then we start to rot, just as, maybe, we have gone to acquire the rudiments of wisdom. It's all erased by old age and it kind of just seems sad and I don't think we have any particular reason to think that 80 or whatever it is, is like, the optimum for a human. I mean, it might be, given the constraints, that if you are going to biologically decay, maybe it's just as well at some point that, like, you put an end to it, so that, like, if it were like leaving to 200 but you were, like, on a respirator in a sort of care home for the last 100 years, I agree that that doesn't seem great.

Bostrom:

But if, instead, you could continue to grow and develop with full vitality and exploit new careers and build on your learning, I think that might be very great. At least that people should have the option to do that. And in this scenario, yeah, like our lives wouldn't need to be truncated by this kind of artificial cellular decay process that currently cap what we can do. I mean, I think it would be great if, if some people of the past were still with us, that, like you read about and you read their works and you wonder, like what would they, what would they think about the world today if they were still here like with full vigor and they could have continued to be active. And I think, yeah, it would be nice if everybody had that option.

Craig:

Yeah, although there are a lot of people who it might not be nice to have around.

Bostrom:

Yeah, but then there are also a lot of new people coming into existence, some of whom will turn into jerks and tyrants and stuff. So I'm not sure the churn changes the ratio of good to bad people. Yeah, it would be interesting to see. We don't know what the human would like, what level of kind of maturity, of personality, or wisdom even, or a spiritual, I don't know, growth that some people could attain if they were 200 years old or 500 years old, or if they had seen history over a larger time span, I mean, but that's even setting 500 years old. Or if I had seen history over a larger time span, I mean, but that's even setting aside the upgrades that could be accessible to humans in that scenario.

Bostrom:

Anyway, I think we are, kind of, rather blinkered in our view of the possible. And if there is this future, I think, people will look back on our current time with both wonder and horror, like a horror at what we had to go through, what lives were like in 2024, even in the most privileged places, let alone in all the horror spots of the world, and then wonder at us living at this critical juncture in all of human history, if maybe, the long-term future for humanity gets shaped in this century, in the coming decades. Imagine being alive at that time. These people living maybe millions of years into the future, in this technological maturity, more or less static condition perhaps, where everything has been kind of developed to its maximum, and then thinking like, wow, they were right there when it happened, like it was their choices that shaped this whole future that would not last until the heat death of the universe. And lo and behold, they mostly didn't care or paid attention. They were too distracted with their social media feeds or what they were going to have for dinner.

Craig:

In going through this thought experiment, what interests me is not technological maturity, but maybe the period short of that, when there is presumably universal basic income and a lot of people are unemployed. I mean, how do you give meaning to people during that period, before you have the ability to alter them neurologically?

Bostrom:

Uh, I think it's personality, and attitude, and your cultural, and spiritual outlook, because there's so much that needs to be done in the world today. There's like no shortage of things that would be worth doing. I mean, if there is one person you could help somewhere, or be kind to somebody who is down, or something that's already worth doing, and then various people in various places have other opportunities to do things as well. And so, if somebody's feeling a lack of motivation, I don't think it's like the world that is lacking opportunities for useful action. I think it's that it somehow doesn't get their hooks into their motivation system. But I think if you want to think of this, kind of intermediate condition, I don't know how long that would last, but if we imagine such a thing, I think, for example, the education system would need to change, which right now is a sort of factory for producing workers. So you take kids and discipline them to sit at their desks, and do what they are told, and to be able to tolerate boredom so that they can then, you know, enter either the factory or, with some more processing in the education system, can become apt for office work, and that, I think, is very sad. I mean, maybe it's necessary today because we do need a lot of workers to actually make the world go round, but it would be much nicer if we could have an education system optimized for training people to have a great life; like to cultivate the art of living well, teaching sort of conversational skills, appreciation for art, creativity, taking enjoyment in the outdoors and in nature and in sports, and in developing deeper relationships and understanding themselves through spiritual practice, like all of these things that are the things you would need if you really wanted to squeeze the last drops of value out of a human life need. If you really wanted to like squeeze the last drops of value out of a human life, uh, those could be taught instead of a lot of stuff that is currently the focus of the education system and I think that all should and would happen.

Bostrom:

Um, I'm not sure we are quite yet.

Bostrom:

I mean, it's an interesting question for kids who are in school, like young kids.

Bostrom:

So, suppose you have somebody who's just like seven, or something who enters first grade, so it's got to be over a decade until they enter the labor force and their lives, even without any life extension or advances right, might continue for another 80 years or something. So that's a long time for any of this to start happening. And so, it does make you wonder a little bit if you have young kids, if you think about what does it actually make sense to teach them? What kind of person do you want to encourage them to become, given that the world that they will enter when they grow up is either already so different or is soon about to become so different than it is today? But yeah, anyway, I think there would be a huge transformation in culture and education that would be called for, if we imagine this intermediary stage where robots can automate most human economic work but where we still don't have a fully plastic world.

Craig:

How do you personally maintain a sense of purpose? I mean, you're very motivated and active but we all face those moments when we ask why am I doing this? We're basically just biological hosts for these blind and dumb genes whose purpose is just to replicate. And so we're these confused, we've developed consciousness, and we're trying to figure out why we're here, and the fact is, that's the only reason why we're here, and that leaves us without much purpose on an individual level. But how do you think about those things personally?

Bostrom:

Well, I mean, I have a lot of stuff to do and so, um, it's more. It's like, the shortage is not, sort of, in things that seem worth doing, but more in the time available to do them, and there has to be a sort of, brutal triaging of things.

Craig:

But what would you say to someone that doesn't have that motivation? As I said, I know people that are sort of living in this existential malaise.

Bostrom:

Yeah, well, like, yeah, figure out whether it's the problem that you're sort of in a life situation that is bad, like stuck in a dead-end job that you don't like, in a relationship that gives you no fulfillment, and if those are the problems, then I think, if possible, change them.

Bostrom:

Or for another set of people, it's not that their lives are objectively bad in their circumstances, but the cause is in themselves, in their brain and then trying to figure out the way to fix that; you know, whether it is some underlying health issue in many cases, or some hormone that is not at the optimal level, or whatever. And then, yeah, in either of these two cases, there is also, and it's like, in many secular communities this has kind of fallen a little bit by the wayside, but the idea of community and spirituality as a sort of meaning. And if one is missing that, then one, I think, is more likely to fall into this malaise of feeling sort of unanchored and demotivated. So finding some group of people and or some set of spiritual values that you could devote yourself to could be another place to look for those.

Craig:

Now, through April 15th, Netsuite is offering a one-of-a-kind flexible financing program.

Head to [Netsuite.com/EYEONAI](http://netsuite.com/EYEONAI). EYEONAI, all run together: E-Y-E-O-N-A-I. That's netsuite.com/EYEONAI, again netsuite.com/EYEONAI.

Craig:

That’s it for today’s episode. I want to thank Nick for his time. If you want to read a transcript of today’s conversation, you can find one on our website, eye-on.ai

In the meantime, remember, the Singularity may not be near, but AI is changing your world. So, pay attention.