**kaifu for transcript to be checked by maria**

**CRAIG:** [00:00:00] hi, I'm Craig Smith. And this is Eye on AI.

This week. I speak with Kai Fu Lee, a seminal player in the development of AI in China, having played key roles in the establishment of both Microsoft research and Google there. Kaifu

is a prominent commentator on the development of AI in China, and has increasingly morphed into a futurist using his knowledge to make educated guesses about how AI will affect the development of humanity.

A couple of years ago, I had Kai Fu on the podcast to talk about his book, AI superpowers China Silicon valley and the new world order.

Now, Kai Fu has a new and very [00:01:00] different book coming out titled simply AI 2041 written with the science fiction, novelist Chen Qiufan. The book contains a series of stories about how AI might impact our lives 20 years from now in some of the benefits and drawbacks, the technology might bring.

The stories range from the possible to the improbable, but all give readers food for thought. Kai Fu talked about the book and his optimism, as well as some of the dangers lurking ahead. I hope you find the episode as fascinating as I did.

Why don't we start by talking about your latest book, 2041, a pretty interesting title as you mention in the book, both because the 41 looks a little bit like [00:02:00] AI, but also it's 20 years in the future from now. And so it's in the realm of science fiction, but a lot of the technologies that you talk about are here today.

So could you start with talking about the premise of the book and then which of the technologies that you talk about are closer to where we are today and which are more speculative.

**KAIFU:** Sure.

The premise of the book is to use an accessible language, to talk about AI. So more people can understand what the technology can do today and in 20 years.

And what are things that are likely to be unsolvable and what are the problems that AI would introduce and how they might be addressed. So storytelling would be a best way to do that. And that's why I partnered with Stanley Chen, who wrote the stories based on my technology roadmap. And the chapters were written in the following way.

Each chapter is a story. Written by him, followed by analysis written by me [00:03:00] and the stories were designed to go from basic AI to advanced AI, to futuristic likely, but not certain developments. So that by reading the stories and the analysis that it would become equivalent to getting a training course on AI as well as my predictions.

So the chapters tend to go from basic use into more and more advanced use with things that can almost happen today may take 10 more years in India or 20 more in, in Africa, but it's possible today, all the way to things that still require small breakthroughs, but breakthroughs that are likely. The basic introduction, chapters cover things like externalities of AI, that is an AI could really make us addicted to content or get us hooked to an app or change our thinking and make people who are already leaning towards a certain political view.

More [00:04:00] extremely. So, as we've seen in, in some countries.

**CRAIG:** Could you talk a little bit about which of those chapters that you feel are closest to what's happening now? The one about the Indian girl who can't meet with the boy that she has a crush on because he's lower caste, which then affects the insurance premiums for her family, which are controlled by AI.

That's something that you could imagine happening today.

**KAIFU:** The first chapter goes into how that could develop, it's even potentially more serious than people think, because in predicting a very benign future where regulations were put in to effectively. Not allow that the caste system or in the racism. If you want to draw in America correlation, AI will still unintentionally learn from all the data, the aspects of how racist views could [00:05:00] leave a permanent mark a Country's history. So that negative outcomes as a result of racism could be remembered by AI as financially beneficial or non-beneficial. And use that in this scenario. That's again, benign and insurance company whose goals are relatively aligned with the user, right? The insurance company wants us to be healthy.

We want to be healthy. Unlike Facebook wants to make money. And then it shows there's a lot of newsfeed that we may or may not want to see. So it's not at least not aligned. So in a benign scenario where the insurance company and the insured who ought to be helping each other are yet put in a position of difficulty of being led by the AI.

To, to make it difficult for someone who fell in love with another person because of racist views that are in the remnants of a society and data that causes even an insurance company [00:06:00] to manipulate things. It also points to the concern that if any company became too powerful in this story and insurance company, that was also providing

basically news and content and social media just has too much data within this power that it can do things that appear to be helping people based on the way AI was programmed. Yet it results in externalities that are dangerous. So one could think of this chapter as an extension of social dilemma documentary that talks about how these large companies.

They're not evil by nature, but they're trying to make money. And AI allows them to make money by shaping people's views. And then this takes it one step further by saying, Hey, even if you have a company that is benign and aligned with our needs, what it tries to do can still be negative to the world.

Because of all the things we don't know of, what's [00:07:00] buried in the data.

**CRAIG:** Let's stop and Talk about that for a little bit, because it's something in the zeitgeist because of many of the applications, in banking, in insurance, as you say, in a lot of social services, in policing, how do you balance that? Or how do you mitigate that?

Because as you said, the biases in society, the data is a reflection of society. Even if you identify a particular bias or you want to adjust maybe a gender bias, you could unknowingly be introducing other externalities, as you say, other facts that aren't seen for years until the data begins to reflect it.

So is it a problem that can really be solved or is it a matter of just regulation to avoid using AI for particular applications?

**KAIFU:** I'm certainly not someone who would [00:08:00] advocate removing AI. The purpose of going one step further is to really. I think for people who are in the AI community to try to raise awareness that the power that you will, as an AI engineer is enormous.

The benefits are huge, but the negative externalities are also everywhere. So hopefully there will be better tools that will watch how AI is being used, how training data is being collected. And then there would be, I would imagine future training for anyone wanting to go into AI engineering. It would be almost like the Hippocrates oath for being in medicine that you need to be aware of the duty of the profession that you need to cater towards, not an extension of bias, but the removal of it, you're responsible for fairness.

And you need to make sure the training data is properly balanced. [00:09:00] And that when you build an application, don't be, single-mindedly focusing on one objective function, but look at multiple objective functions. So those are some of the directions for the education of the engineers. Secondly, there can be tools just like we have tools compilers.

that will alert us towards a possible problems in our programs. One, one could imagine AI training algorithms that the future can alert us towards imbalanced training datasets, inadequate training, datasets, potential bias in the system, the need to look into potential problems. So there'd be a AI compiler, if you will, that will come out with a warning and then the engineers should heed these warnings.

And then also. Taking AI to the next step, right? One of the problems of AI is a double-edged sword that AI training has is that on the one side is incredibly good in optimizing one objective function such as, [00:10:00] maximizing views or revenue or in the minimizing defaults or reducing insurance rates. So those things are things that AI is incredibly good at, but can we possibly in the future.

Build AI training so that it has multiple objects to functions. In other words, minimize the insurance rates subject to maintaining a social fairness or something like that. This one is obviously incredibly difficult to do, but not impossible. So I think the single-mindedness of AI is what makes it so good, but also somewhat dangerous.

So can we make it a little bit more complex? Just like we humans, when we make decisions, we're not optimizing one thing, why can't AI do the job of a CEO? Because the CEO isn't just there to maximize earnings per share. The CEO also needs to care about employee satisfaction, customer satisfaction, [00:11:00] recruiting, corporate brand, long-term goals, ESG, sustainability. So balancing multiple things is something people have somehow learn to do and something we need to gradually teach AI. So those are just some of the directions. I don't know all the answers, but I think awareness is the first step.

**CRAIG:** But if you leave awareness to the private sector, they're likely to cut corners or do things that are not publicly visible.

Do you see a day when there'll be government audits of AI systems? For example, and do you imagine that extending beyond national borders where governments will get together because the products and their effects certainly cross national borders, how do you see the regulatory environment playing out?

**KAIFU:** That's a good question. I think, first of all, There must be regulations, but then I believe the current environment relies too much on just regulations. So yes, I do [00:12:00] believe regulations are needed the current regulations tend to be extreme, investigate a large internet company for antitrust. And then if they're found guilty, break them up.

I think that is a much too general and too harsh. And doesn't necessarily solve the problem. A patient who needs a surgery and you bring an ax saying, okay, I'm going to chop the person here or there. That's. I think that can't be the only tool anyway. Okay. So I think there needs to be finer tools. And I don't know what they are, but I'm afraid most governments will tend to think of a huge surgery to surgically remove something, but that may be too extreme.

So I think we need to look at other ways of fixing the problem. I do think audits are a good idea because these problems that in social dilemma or in the stories in AI 2041, are not necessarily violations of the antitrust. They may [00:13:00] be inadvertent mistakes by the company, or they might be well-meaning acts that resulted in externalities.

So they needed to be dealt with more delicately. So I think we need to come up with what are the things that companies who possess data and AI capabilities should not do, that needs to be clarified. A lot of the currently publicly described. Good behavior of AI system or bad behavior are too general. We need to be specific like audit rules.

If you break this, then that's bad, if you do this, then there's this punishment. And again, I don't know all the answers, but we need to get to that level of granularity for regulations to be meaningful. And once they are fine-grained regulations, it's impossible for a government to basically evaluate all companies in person.

So as you said, audits would make sense. So, an audit could be triggered by bad behavior by complaints or by signs of problems, [00:14:00] and then an audit, whether done by machine or a human, when problems are revealed will have consequences and corrective behaviors. So I think that is the direction we need to move towards.

The other thing I think is a useful, is to think about what happened to ESG. ESG has generally been successful, slow, but successful in guiding positive behavior incorporations, while companies aren't going to on their own, do anything that reduces their profits in order to quote-unquote do good for AI.

If there are ESG third party watchdogs, that would report how well companies are doing. And the shareholders started to care about the reports of the watchdogs. And then if companies need to voluntarily report their metrics, then I think that creates a self monitoring mechanism that could work perhaps even better than

[00:15:00] regulations. Right? Where we're seeing things like carbon neutral sustainability becoming more and more concrete, becoming more and more measurable by third parties. To give you an example. When I was at Google, there were third party metrics of search relevance. While Google didn't necessarily agree with all the metrics it cared a lot.

Because there were third party watchdogs watching where the search was becoming relevant. One could apply this to fake news in social media and other things. So third parties that watch for these bad behavior and publish them. And then they become credible. These might be funded by a social responsibility, corporations or government or news, media or philanthropy, I think as these gain more credibility.

Companies with a lot of data in the domains of search and e-commerce and social media will have to learn to self behave and then ESG incorporating these elements, I think [00:16:00] will also make shareholders become a watchdog on large companies. So I think a combination of self-discipline, regulations, third party watchdogs, as well as ESG, might be with fine-grained metrics and ways for self-improvement that's probably the way forward to move towards a solution.

**CRAIG:** Yeah. For listeners who aren't familiar with ESG that's Environmental Social Governance criteria.

There's a talk of establishing some kind of public disclosure of what algorithms, what tools, what data sets are being used in a particular AI systems so that people are aware of it on the one hand, but also an audit would show [00:17:00] whether or not companies are adhering to what they're disclosed publicly.

Later in the book, you get very speculative. If there's a story about a Japanese idol who disappears is presumed dead. And it turns out that he was actually a projection of the combination of technologies that you described as XR X reality. And that sounds fairly far in the future. Can you talk about how

far out you feel that kind of technology is?

**KAIFU:** I think one can easily measure my true belief on the timeframe of technologies by looking at what I invest in. I don't currently invest in AR VR. Yes, I make very vivid predictions in the book. I think that shows that my belief is that these are a set of technologies that will become incredibly important, but the timeframe is beyond five [00:18:00] years and perhaps even 10, VR and AR even today still has a lot of issues.

For example, the head-mounted displays are inconvenient and too big and too heavy. And also they make someone look silly wearing it in public. Google glasses alert people that, Hey, there I'm being recorded. So there are social issues, inconvenience issues, acceptance issues that need to be overcome. And then I think the way to make a better prediction in this domain is looking at.

the Engineering of these products. How long will it take before we have a naked eye VR, AR and I believe not within 20 years, how long before we have basically like glasses? I do believe that is doable in less than 10 years. And how long might we have a contact lenses? That would work similarly, probably between 10 and 20 years.

So we can project that based on, if you draw a curve of how much a wage [00:19:00] reduction there has been in head mounted displays that have certain functionality. And then also you can project other things like other feedback suits and the gloves, and also how fast the how many frames a second? How busy do people get in VR environments?

I think we will see that is making progress on a year to year basis. So we can project at some point. This is going to be socially acceptable and light and easy enough. So the prediction is not without basis. The other thing is that immersive entertainment is truly an incredible killer app. We see how people get into these games that are immersive and large tent multi-user and engaging.

The quality of games are improved to increasing people's interests. And if you want to call it, addiction in games is huge. And we are beginning to see movies and games. [00:20:00] And multiple endings, all evolving. So I think that is a clear direction, and I actually worked on VR in the nineties and Silicon graphics.

It was very primitive, but even then we could project that this was something people wanted to do. And also it, it we've learned once again, the history is truly a predictor of future. 3D technologies, how did Silicon graphics gain his notoriety? It was because of 3D graphics. How did it lose its luster? Because 3D had become pervasive.

So the killer app for 3D immersion is absolutely entertainment. It might be social like Facebook is predicting. It's probably not in a business in that kind of minority report way, that's probably maybe later. So that's why the chapter that I wrote on XR is all about in the projecting [00:21:00] and future trends.

This is definitely going to happen. Glasses and contact lenses and focused on entertainment as the key. Breakthrough domain. And also focusing on the kind of entertainment you can do with XR that is undoable today with games. So those are the elements behind the story. And then of course, Stanley made it a very interesting, a story that hopefully becomes compelling and makes people think, wow, that's the kind of game I would too one day I want to play.

**CRAIG:** Are you hopeful for society when you talk about the entertainment being the killer app, there are a lot of good things. Even the Minecraft now is embedding educational elements in that world. It could be very powerful for education. Are you over all optimistic?

**KAIFU:** I think, it will go in sequence. If you just say what has led the development that's 3D [00:22:00] technologies, because 3D really is precursor to XR.

And we'll see that even today, 20 plus years after Toy Story is still primarily entertainment for 3d technologies. And yes, it's 3D used in education and training, certainly, is it used in 3D web graphics? Of course, but, but mostly still driven by entertainment. So I would say that XR certainly will help education and training.

One can point to a lot of scenarios. It's already been demonstrated that a head-mounted display XR environment can help children with autism, for example, because it is a safe social environment in which they can learn to engage et cetera. So of course, and then you can use it for anything you have to learn about your skillset.

You're training to be an auto mechanic or aviation repair or whatever. I think AR can be a great way [00:23:00] to enhance training. Having said all that, I believe the killer app for some time to come for 20 years to come for XR will be primarily the 90% plus in entertainment space. And there will be these uses in training and education and real estate and business that will also be useful.

But primarily the business driver will be that of entertainment. And as you said earlier, the social implications, if we are in the future in a world with many jobs being displaced and games become so engaging, there could be a tendency for many people who lose their jobs and potentially lose their meaning of life to drift in the virtual world, to find a parallel life in which they can be satisfied.

And that could be a very bad thing for society. If a large enough percentage essentially become the useless class that live in a parallel life that has no [00:24:00] meaning for this world. But I think we also need to be aware that XR is coming. It will become incredibly engaging and addictive and it will be mostly applied to entertainment.

And then with those as givens, I think we have one or two decades to avoid the problems.

**CRAIG:** Can you talk about the Isle of happiness?

**KAIFU:** The Isle of happiness, is based on a story of a middle Eastern sovereign and a small island doing an experiment with this inhabitants to deeply measure everything that they do in their public activities and estimate whether AI is making them happier

and trying to build an environment that continues to make people happier. So it's an extension from the earlier chapter of what happens when you have a single minded basic objective function. It's very good in optimizing something, but it's also. Too [00:25:00] single-minded and not able to look broadly. So one way was to make a multifaceted objective function that caters for many things while another direction is why don't we measure something that is truly innate and universal, that makes everybody better.

But then we don't have the issue of objective function being too narrow minded about improving revenue or something like that. So what can we measure about us? There are ways to measure or approximate our happiness, right? Our expression and micro expression. Tell a lot about what we're feeling and also are in the hormones.

Endorphin and other hormones can potentially be continuously measured. Right? There is continuous glucose monitoring. There can be with some research that we did continue as hormone approximate measuring, and those correlate also with happiness. So it should be possible to build a universal [00:26:00] observer of individual

actions and speech and expressions and hormone changes. And then based on the things that people have done or things that have happened in environment, what made the person happier? When they, the endorphin level go up, what made their pupils dilate? What made them smile and then try to optimize that. So that was the premise of the story.

And without giving away the ending, I would say that it is basically pointing to the future. When we can look at higher level things to use AI, to make us happier, that seems like a feasible direction. These things can be measure. But there are some obstacles, whether people are willing to give their personal data to some AI.

And that's why I pointed to an experiment on an island, not a large country. And also there are possibilities of using a federated learning and things that can protect your privacy, but at the same time on learn globally across multiple people. [00:27:00] It also points to that I think these kinds of experiments are more likely to be done in very small contained environments to prove its efficacy and then, more countries may be able to, or more companies may be able to adopt it.

If they work. Of course, the technology is still difficult. One difficulty is, as I mentioned is whether people will allow this data to be collected. Another difficulty is how do you build an AI objective function based on happiness, which is not instant gratification, but really long term. And the longer term you look at happiness, right?

The feeling of watching your children grow up to be the best that they can be. That's the actions that you took perhaps five or even 10 years ago that helps that outcome. So how do you trace long-term actions that resulted in something that helped your happiness? That's a big question that we don't know, but it kind of points to [00:28:00] the importance of that.

Just like today, I have technologies that like GPT-3 can look at long-term things that I said a million words ago that contributes to my next word. Why couldn't we do that with actions. So theoretically, these are all pursuable topics. And then the final, big question that is asked in the story is what is happiness?

And it turns out in the story by merely measuring instantaneous happiness dilation of our pupils, a surge in certain hormones, a smile on the face. It may be too short term. It may not be measuring our true happiness in the longterm because if we look at Maslow's hierarchy, the basic kind of happiness is sustenance and then followed by things like love and companionship.

Respect and success and then followed by self-actualization. Then the higher level you go in the Maslow [00:29:00] hierarchy, the more difficult it is to measure. So that again, points to a challenging research problem. That will be very difficult to do. But it also answers the question I raised. Can we measure people's happiness?

The answers measure instantaneous, happiness, maybe, but if true happiness is about self-actualization that's so long-term how do we measure that? That's a really long-term problem that is worthy of working on, but probably cannot be solved in 20 years.

**CRAIG:** You mentioned education earlier and the promise that AI holds in various ways were improving education and education is the foundation of self actualization of that kind of long-term happiness. I would argue [00:30:00] that AI will go a long way toward improving the education system with personalized education and optimized education.

So that may be part of the answer. You mentioned climate change a few times in the book and climate change is a multifaceted problem and AI can be applied in so many different ways to either answering questions about climate change or creating or optimizing energy usage or production and that sort of thing.

Do you have a view on whether AI is going to be critical in solving climate change, which is an existential threat.

**KAIFU:** In the book, I'm making a very optimistic prediction that we are able to fix the climate change problem. Is fixed partially by AI, but also by a lot of other technologies. So the largest contributor to the problem today is, is fossil fuel.

And [00:31:00] there is a natural, predictable trend when alternative energy will and must take off alternative forms of energy. Have had difficulties in the past because economically they were not the lowest cost energy. So fossil fuel was significantly cheaper and more convenient, but the things have changed. If you plot the cost of solar energy, it has come down over 80% in the last 10 years, and also a lot

in 10 years before that. And there's no reason to doubt why the continued lowering costs will keep happening for the next 10 to 20 years. So another factor of three to 10 appears by experts in the space at the same time. Battery storage currently primarily through lithium ion batteries, but there are new materials being researched.

This has become a hot area for research, and we know that [00:32:00] historically, whenever an area becomes very hot, very lucrative, the best minds jump into the area and breakthroughs happen and we are seeing substantial progress and battery storage has also become a lower cost over the last 10 years, again, by 80 plus percent reduction.

And that's likely to happen. It will require some breakthroughs, but I take the leap of faith that will also happen. So if you put the two together, we see a future of a new form of energy becoming the cheapest solution by using solar panels to collect energy when the sun is out and then storing it in battery and then using it when it's

night or lack of sun. And if you project both of these improving, they will become by far the lowest costs in a 10 year timeframe. Of course, there are still many issues. Battery technologies need to be improved. Transportation of energy is a problem. You can't collect it in [00:33:00] a Wyoming and send it to New York.

But then you can do it in a distributed fashion. You can collect it for initially for factories or businesses or buildings or communities or districts. And that distributed form of energy will become so cheap that it will become the dominant form throughout the world at a fraction of the cost with potentially the electrical grid still existing, but as a backup solution only to be used as needed.

So what this projects is several key changes. One is that the cost of clean energy will come down dramatically. So that is not just the right thing to do to shift, to clean energy, but the economical thing to do. And secondly, it does require a certain infrastructure to be built in. But it looks like the major countries are in agreement to build these infrastructures.

The third interesting observation [00:34:00] is that energy will go through a transformation from what used to be natural resources driven. Like who's got oil and natural gas and coal into manufacturing driven. Namely, can we produce enough solar panels and batteries that can store and transport. So those are, I think, natural revolutions that will take place that has basically all the wind behind the sails because smart people are going into it.

Countries are becoming more committed, costs are coming down. So it's economic to do not just socially rights to do. So, that's what makes me an optimist. The AI can play a role in improving the distribution. Improving the manufacturing, improving the scientific process, but frankly, it's not the major thing.

There are many other things that will come into play. But the other interesting thing that is even more speculative in the book is [00:35:00] that as the cost of energy drops dramatically and the cost of materials also drop because rather than, taking things such as plastic from oil, we can potentially make even better materials that are lower costs and I'm not necessarily limited in quantity and then followed by automation in the factories.

So effectively, the cost of goods today are largely the combination of materials and energy and labor. So we're looking for a 20 year out future where the cost of energy might be 10%. What it is today, cost the materials. Similarly, cost of labor, similarly through robots and AI replacing the workers. So we will actually enter what in the book, what I call the age of plenitude or abundance, which is the last chapter that says if cost of most goods can drop by a factor of 10, then we can potentially.

Wipe out poverty and hunger, [00:36:00] and we can potentially make a good life accessible to all people on earth. What possible things will need to change by economy? That's based on scarcity. There's no longer a scarcity money is needs to change. People's goal in life needs to change because accumulation of wealth may be no longer that important.

So what will the future of humanity become? So the last chapter is a set of speculation based on these three major things that essentially makes the age of plenitudes arrive in a matter of twenty years. Yes,

**CRAIG:** 20 years is certainly optimistic but the age of plenitude itself is optimistic and your concerns. Last time we spoke about what various people have called a useless [00:37:00] class. People who are displaced in the workforce can enjoy those age of plenitude, but have lost a sense of purpose. You were just talking about the isle of happiness.

How AI can possibly optimize different levels of happiness? What's your feeling now instead of evolve, do you remain optimistic about the state of humanity in this future? Again, my view is that education has to become the dominant player in human development for everybody because the more educated you are. The more likely you'll be able to find some way to solve actualize as you were saying, but how do you feel about it?

**KAIFU:** I agree. Education is a very important ingredient. There are other ingredients that are probably all necessary at the same time. I'm optimistic that we will have the tools to solve the problem. I'm not sure if we are wise [00:38:00] enough as a species to actually apply the tools correctly early enough. And the book intends to

signal the problem and hope people will be motivated to move towards the solutions. So the types of solutions I talk about in actually multiple chapters about this point of the job displacement by AI, I think one chapter talks about the possibility of encouraging more people. Into connecting with human beings and service jobs and social connectedness as something AI cannot replace.

And only the service industry is large enough to accommodate a lot of, a lot more people who may lose their routine blue collar or white collar jobs. So service jobs in the area of healthcare services in jobs like tour guides, concierge, elderly care is one class that I talked about as being large enough new jobs, a [00:39:00] second suggestion implicit in another one that the stories is that there may be jobs that don't have direct economic contribution to the society, but people should be paid nevertheless for those jobs because they lead to good things.

And that includes things like voluntary services, foster home, hotline, homeschooling of the children could count as a compensated activity. And also what happens when entry level jobs are wiped out, but senior jobs are still needed when entry-level reporters are replaced by AI, but we still need columnists.

Then maybe we do need people to be, to have entry level jobs and to write entry-level simple, quarterly report. And the sports reports, a simple news so that it can be practiced and be selected for upward mobility. Even though AI could do as good a job, we might still want people to do those jobs. So we can [00:40:00] group people who are very talented.

And another one that the stories talks about economic redistribution, because one of the other outcomes is that as large numbers of people whose jobs are replaced by AI and automation, there are ultra wealthy billionaire as being. Who build great technologies and use AI to displace jobs and make money and et cetera.

The distribution needs to somehow take place through mechanisms like universal basic income, but also encouraging people to be retrained into new jobs. So the idea of retraining is also embedded in another one of the stories. And finally, your favorite topic education. Is another one that the story is called twin sparrows. Twins.

Both got the help of a personal AI assistant. That is like a superhero friend who knows a lot about each individual [00:41:00] and tries to make them the best that they can become, pushed them into professions and choices. Individualized success. Is achieved by making someone who's very good at and very passionate about something, directing them in that direction.

And that AI companion slash teacher is programmed by people who are very wise and who also act as mentors. So it points to a future where AI can be used for learning basic tools for the kids and also targeting the kid's shortcomings to make sure they have a strong foundation and also develop their interest in skillsets and encouraging them by making education fun while the human mentor will still help the children in grounding

there are values knowing what's right and wrong, learning how to make friends and work [00:42:00] as a team and communicate and learning to be creative and build critical thinking skills. And that is the kind of education revolution that I point to that could make the young people of today. Not only ready, but also thrive in the age of AI.

**CRAIG:** Is there anything that you see that could interrupt this progress? What we're talking about is really a fairly linear progression toward that future and it's most certainly won't be, is there anything that concerned you that maybe the promise of deep learning will fall short at some point, or there'll be external factors that interrupt research.

**KAIFU:** In the book, I talked about many things that could go wrong. One is certainly. Bad guys who use AI for evil and that has happened and will happen. And hopefully that can be contained. The second [00:43:00] is that some of the issues and externalities give AI a bad name so that before AI has a chance to succeed and prosper and help us become better.

It becomes a negative word, but just like early days of electricity, right? There were a lot of controversy that was probably not ready to kill off electricity, but at least made the controversial. I hope that doesn't become too bad. And I also think there are fundamental human problems and frailties, like fear and greed that could get in a way.

A lot of what I talked about optimistically, do not necessarily address the issue of fear and greed. What if products are no longer scarce, cost of goods plummet, but corporations continue to create artificial scarcity, like diamonds are not really that scarce anymore. We can also make them and also mine, more diamonds, but de beers chooses not to.

And there's all [00:44:00] these commercials that makes people think diamonds are the girl's best friend when it's really that scarce and should not be priced that way. So humans have a bad history of letting fear and greed create fake stories in the Yuval Harari says a lot of what we do is stories, but these stories are used to brainwash people.

If people choose to believe that there is no scarcity and continue to hoard the money as the only goal that I'm not sure we can control and contain because the idea of accumulating money for safety or for greed, or for competitive reasons is a thousand year practice that I don't know this optimism can overcome.

And then finally, I think there are some serious, dangerous, I point out in particular that of autonomous weapons, that low costs that a terrorist can make a dedicated assassination drone by simply taking commercial products, a drone [00:45:00] from the internet add face recognition. Just download open source software and then add some dynamite or leveling of a bullets.

You can make a mini miniature assassination drone that way, and they're very hard to control or contain. And also a terrorist group can make a hundred thousand of them, put them in a airplane or a large cargo truck and just unleashed upon the city. So those I think are much harder to control, even more difficult than nuclear weapons because nuclear weapons are at least contained by the principle of mutual assured destruction that if you unleash nuclear weapon on other countries.

The counter response can wipe you out. So that keeps everybody in a deterrent situation, but autonomous weapons may does not work that way. It's very effective and also very low cost [00:46:00] and accessible to terrorists and non-state actors. So those point to a direction that needs some kind of regulation, I do think containable.

Humans have contained nuclear weapons, chemical weapons, biological weapons, but many large countries are choosing not to, to regulate or even discussed the possibility of regulating autonomous weapons, which I think is something that a lot of AI scientists have spoken up. And something that I think deserves more attention because of the seriousness of the, uh, the outcome and the consequences that could follow.

**CRAIG:** That's it for this week's podcast. I want to thank Kai Fu for his time. I encourage those interested in the future of AI to read the book. Which will be published in September by random house.

Remember

[00:47:00] the singularity may not be near, but AI is about to change your

world. So pay attention.