

Technically Optimistic

An Emerson Collective Podcast

AI is at a crossroads. How did we get here? In **Episode 1 of Technically Optimistic**, Raffi Krikorian, Emerson Collective's chief technology officer speaks with John Markoff, journalist and author, about the history of AI; Suresh Venkatasubramanian, coauthor of the White House's Blueprint for an AI Bill of Rights and professor of computer science at Brown, about ethics; and Adrienne LaFrance, executive editor of *The Atlantic*, about AI's impact on society.

CLIP: AI.

CLIP: AI.

CLIP: The AI craze.

RAFFI:

It seems like everyone is talking about artificial intelligence these days.

CLIP: I mean we are on the cusp of unleashing an intelligence revolution.

CLIP: "Artificial intelligence is no longer a futuristic idea. It's already here...."

CLIP: Real-world AI, that's going to be a platform that will be transformative.

RAFFI:

You may have heard AI talked about as the biggest, most significant technological development since the internet....

CLIP: We have one of the most significant technology disruptions of our lifetime, probably bigger than the internet itself.

RAFFI:

To hear some people talk about it, maybe it's even the biggest thing, like... in all of human history.

CLIP: Sundar Pichai: "You know I've always thought of AI as . . . more profound than fire, electricity, or anything that we have done in the past."

RAFFI:

I want to live in the future – I want to live in a world where cancer is eradicated, where every child has the best teacher possible, where sea levels are no longer rising – and all of this might be possible, with AI.

But, we also have to deal with the challenges. And, I certainly don't have all the answers.

Despite all this recent coverage, we've actually been living in an AI world for a while now. It's used by landlords, banks, and credit card companies, not to mention in video games, social media recommendation systems, Alexa, Siri, your phone's camera, Uber, Lyft ... and so on.

So we actually know lots of the good things that AI can do.

But, we also have to deal with the challenges.

The discussions around AI have either been about all the good they're gonna do, or the doom and gloom. If you want AI hype or AI phobia, there are plenty of other podcasts for you. This show is about the nuances. We're going to ask a lot of big questions, and there won't be easy answers.

First off, let's start by pointing out that there are some fundamental contradictions in the conversation about AI.

On one hand, we could be on the precipice of a world where education, transportation, and health care could be enhanced by AI to dramatically improve our quality of life.

CLIP:

Imagine the possibilities if AI were to take away the mundane from you. If AI were to cure diseases.

RAFFI:

And on the other hand, this technology is sometimes talked about like it's a force of nature... something uncontrollable, that, once released into the wild, can't be stopped... a Pandora's box which will wreak unknown havoc on the world.

CLIP: Many top experts have signed a statement warning of the risks of artificial intelligence, and this is...

CLIP:

NEWSCASTER: Is there any turning back?

HINTON: I don't know. . . . We're dealing with kinds of things we've never dealt with before.

RAFFI:

Over the course of this series, we're going to ask a lot of big questions... and you'll hear responses from a diverse array of thinkers. Like, for example... Who are the people developing

this technology? Does what they're making reflect *their* values, or is it really for the public good?

CLIP:

MEREDITH BROUSSARD: *We all embed our unconscious biases in the technology that we create.*

RAFFI:

Should there be guardrails in place? And should those take the form of government regulation?

CLIP:

SAM GREGORY: *We know existing harms that will be exacerbated with almost a hundred percent certainty, given the current generation of generative AI.*

RAFFI:

What will this mean for the future of jobs?

CLIP:

ERIK BRYNJOLFSSON: *We're not running out of things for humans to do... you know, right now.*

RAFFI:

The future of education?

CLIP:

SAL KHAN: *There's stats that 40% of teachers are already using ChatGPT, I think every teacher on the planet would say, Sign me up, and can I get five or six [LAFF]..*

RAFFI:

Of creativity?

CLIP:

JUSTINE BATEMAN: *I love writing. Why would I give that to some computer program?*

RAFFI:

How about mental health?

CLIP:

ROZ PICARD: *Today the AI cannot replace an expert human therapist. That said, there is a lot more that AI can do.*

RAFFI:

Or the future of humanity?

CLIP:

ADRIENNE LAFRANCE: *Just because you can build it, should you? I don't think that we've seriously dealt with that question yet.*

RAFFI:

I'm Raffi Krikorian. I'm the CTO of the Emerson Collective. I've been in all corners of the tech world for over two decades. I'm an MIT alum, I was a VP of platform engineering at Twitter, I built self-driving cars as the head of the Advanced Technologies Center at Uber.

I certainly don't have all the answers.

The thing is, anyone who's ever made you feel like these are all easy questions... and simply black and white ... was wrong.

No one can predict the future.... But we are Technically Optimistic.

music.

RAFFI:

If you have been following the news lately, you've been hearing about ChatGPT or DALL-E, and maybe it seems like AI just came on the scene. The truth is, this technology has been in development, or in your life, for decades now.

John Markoff:

There is such a rich history that has entirely been forgotten. I mean, you know, don't they say those who forget their history are condemned to repeat it?

RAFFI:

Journalist and author John Markoff has watched this unfold in Silicon Valley, and elsewhere. He covered technology for the New York Times for almost thirty years, winning a Pulitzer Prize in 2013 for his reporting. I spoke with him in May.

John Markoff:

So let me tell you just a little bit of the prehistory. I mean, language models made conversational systems work. But the idea of a conversational interactive system has been a passion and been part of the history of this industry for decades. And if you trace it back, there's a great back story, and a great early history.

RAFFI:

Two quick things before John gets into all this.

First, this story, sadly, like a lot of tech history, features almost entirely white dudes. That's unfortunately all too common in the industry, and that lack of diversity has been a huge problem. Just so you know, we're gonna be taking that problem on in later episodes.

And second, this story John's about to tell is really pivotal, because it's going to introduce a key theme for us in this series.

Here's John again.

John Markoff:

In 1962, which was the dawn of interactive computing, there were two laboratories that were created on opposite sides of the Stanford campus. One was created by John McCarthy, who had actually coined the term AI in 1955. And McCarthy came to Stanford and created the Stanford Artificial Intelligence Lab with the intent of building machines that replaced humans.

RAFFI:

Here's John McCarthy reflecting back on his career in an interview from the late '80s.

CLIP: McCarthy

I started my work in artificial intelligence in about '56. I would say that the field has made somewhat less progress than I hoped. In some respects, computers are easily more self-conscious than human beings. It's not hard to make a computer program look at its own program.

John Markoff:

That same year on the other side of campus, Doug Engelbart set out to build machines that extended humans. And he called that technology IA, intelligence augmentation.

CLIP: Engelbart

So take all those things together and call 'em an augmentation system. That's what augments humans.

RAFFI:

Douglas Engelbart, who basically invented the computer mouse, had a really influential presence in Silicon Valley... and throughout the history of computing. Here he's giving a presentation in 1986 at the influential Xerox PARC.

CLIP:

And, for good practical purposes, let's make two parts of that. One side's got all the technology in it, and the other side has all the other stuff.

John Markoff:

So you had this dichotomy, this tension, this paradox at the heart of the industry right from the beginning.

RAFFI:

There it is. That's the big theme. On the one hand, AI can be thought of as a tool for automation – technology meant to replace human beings. But, on the other hand, it can be thought of as augmentation – a tool meant to supplement human beings.

Music swell

John Markoff:

And at the same time, there's another bit of history. I mean, everybody knows that Siri was the first sort of major success in conversational systems. Apple, in 2010, decided to put a conversational system at the heart of the iPhone platform.

And the two architects of Siri, Adam Cheyer and Tom Gruber, people have forgotten they were both disciples of Doug Engelbart. They were really believers in this IA idea.

CLIP: Tom Gruber

...We designed Siri as humanistic AI, to augment people with a conversational interface, that made it possible for them to use mobile computing.

John Markoff:

They, they also both were passionate about building a system that replicated, uh, a vision video that Apple Computer had created in 1987, which everybody has completely forgotten, but they should go back and look at.

Music from Knowledge Navigator ad

It was called Knowledge Navigator. It's on YouTube. You can see it. It describes this interaction between an absent-minded professor and a computing system in the form of an avatar in a personal computer.

CLIP: Apple Knowledge Navigator ad, 1987

AVATAR: *You have a lecture at 4:15 on deforestation in the Amazon rainforest.*

PROFESSOR: *Right... Let me see the lecture notes from last semester... I need to review more recent literature. Pull up all the new articles I haven't read yet.*

AVATAR: *Journal articles only?*

PROFESSOR: *Mm-hmm, fine.*

AVATAR: *Your friend Jill Gilbert has published an article about deforestation in the Amazon and increasing imports of food.*

PROFESSOR: *Contact Jill.*

AVATAR: *I'm sorry, she's not available right now...*

John Markoff:

So where did Knowledge Navigator come from?

Music swell

Alan Kay's laboratory. Alan Kay, at this point, had an advanced technology group at Apple. The people who worked for Alan, they were in the Engelbartian tradition, they were not in the AI tradition. Actually, um, they were in the cybernetics tradition.

RAFFI:

The field of cybernetics. It's been defined in a lot of different ways... but it studies how information processing and feedback occurs in systems. Those systems can be intelligent and biological... like human beings or animals... or they can be more abstract systems... like organizations, or societies.

John Markoff:

And his group put together Knowledge Navigator. And so at a certain point I went to them and said, where did these ideas come from?

And Kay said, Nicholas Negroponte.

CLIP:

VOICE 1: And I'm very pleased to be joined by Professor Nicholas Negroponte, who is chairman emeritus at the MIT Media Lab, amongst other things...

RAFFI:

The MIT Media Lab. A research center notorious for living in the future. They work at the interface of humanity and technology. Here is Nicholas in 2013.

CLIP:

NEGROPONTE: Even though you could imagine intelligence in the device, I look today at some of the work being done about the Internet of Things, and I think it's kind of tragically pathetic...

John Markoff:

So I went to Nicholas because I knew Nicholas. And I said, where did these ideas come from? And he said, Gordon Pask. And I said, who? [LAFF]

CLIP:

VOICE 1: Gordon Pask has now been developing his theory of learning for many years, and, as an engineer, at the same time devising his own learning machines. What was deficient in the existing theories to make him start?

PASK: *I think that a lot of existing theory and experimental work was inappropriate to the field of education where you're dealing with people who...*

John Markoff:

Pask was a British cybernetician who had an important theory of learning and education and human intelligence. His argument was intelligence emerged from the interaction between – the communication between humans. It emerged from human language.

CLIP:

PANGARO: *Language is the foundation, but this loop of acting, sensing, comparing to goal, acting, sensing, comparing... Is fundamental.*

RAFFI:

Paul Pangaro, speaking here, was a student of Gordon Pask.

CLIP:

PANGARO: *In language, there is conversation, and this back-and-forth looping of ideas has an evolution, where as time goes on we share more and more an understanding of what the other means. And that's the nature of society, too, where we're enmeshed in these conversations, which are the foundation of all social and human interaction.*

John Markoff:

And so, those roots, I think, have completely been forgotten. But there, there is that sort of fundamental stuff that is now, you know, it's everywhere. Right? Language is fundamental. We now have a computer interface that touches the entire human race. And that's where we are.

Music

RAFFI:

Nowadays, the AI platforms you've been hearing about, like Midjourney or Google's Bard, use what are known as "large language models." Basically, a massive interconnected network of math equations designed to take in huge amounts of data, and find patterns.

CLIP:

ChatGPT is capable of engaging in human-like conversations, providing answers to questions, and even generating original content.

And they can do this because they've read almost everything on the internet, an absolutely humongous amount of text.

So, when you ask it to write you a story about "Jack and Jill," these large language models start to chain words together to statistically match what you've asked for – hey've seen, read,

and ingested an absolutely huge number of stories about “Jack and Jill.” Like... almost every “Jack and Jill” story ever written. And then if you ask it to rewrite that story, say in the style of a pop song, it will -- because it's read the lyrics to an insane number of pop songs.

And, they've analyzed almost every picture ever posted online- along with their captions, and metadata that describes what's in the picture, or where a photo was taken.

CLIP:

DALL-E was created by training a neural network on images and their text descriptions. Through deep learning, it not only understands individual objects, like koala bears and motorcycles, but learns from relationships between objects. And when you ask DALL-E for an image of a koala bear riding a motorcycle, it knows how to create that or anything else with a relationship to another object or action.

But let's be clear: These systems are not thinking. They're not reasoning... they're simply solving mathematical problems. If it seems like they're thinking, that's because they've been trained on the texts that thinking human beings have produced... and they're really good at imitating.

They're so good in fact that some people – even engineers – remain convinced that they must be sentient.

CLIP:

VOICE 1: But our next guest is convinced sentient AI exist. Blake Lemoine is a former Google engineer who was working with a chatbot called LAMDA. And then things took a pretty strange turn.

LEMOINE: Well it started saying things about its feelings a lot, so I asked more questions. And I eventually got to the point where I asked if it was sentient, and the answer it gave was so nuanced that it convinced me that it was. It responded that, well, we don;t really understand what sentience is scientifically, so how could we tell if I'm sentient?

RAFFI:

But let's put a pin in that question for now.

So why are “large language models” such a big deal?

Two reasons: the amount of data, and the amount of computational power. Those two things have been married for the first time ever. Now, a computer model could process pretty much all of written language.

So now these large language models, or LLMs, have enough computational power that they can discover the rules of language themselves.

To take a very simple example, if a sentence starts with “the weather is....” the model knows it could end with “sunny” or “cloudy”, but not the word “elephant”.

This is what’s called “machine learning”. A computer programmer doesn’t sit there and encode all of human knowledge manually as rules into a machine, the machine has quote-unquote learned the rules all by itself.

RAFFI:

So we say that the machine has... learned. But who was the teacher? We don’t always know what a machine has figured out for itself... or why.

For the same reason it might give statistical weight to “sunny” at the end of a sentence about the weather, it might also reproduce other biases that it gleaned from the data it’s fed.

Suresh Venkatasubramanian is a professor of computer science at Brown... and he’s one of the leaders in research at the intersection of artificial intelligence... and ethics. He refers to it as “algorithmic fairness.”

SURESH:

It turns out – and this is very fascinating, right – these tend to link back to normative claims about bias and fairness that philosophers have been studying for eons. I mean, some of this goes back to Aristotle. It's been a really long time.

RAFFI:

Up until August of 2022, Suresh worked in the White House Office of Science and Technology Policy.... helping to craft the framework for an “AI Bill of Rights.” You’ll hear more about that in Episode 2.

SURESH:

The interplay between the technical formulation, this function, this measurement of ratios and how it carries some normative weight because of what it maps onto in terms of egalitarianism or, you know, equal opportunity and so on, is what makes this area so tricky... because we said the algorithms have no intent, but literally by coding a particular formula for disparity, you are taking a philosophical position that you didn't realize you were taking.

And so the algorithm does have intent, but not in the way you expected it to, and that's kind of hard to get your head around.

This idea that design choices, technical design choices, can have societal or ethical consequences that come from a design choice but have never been spelled out, is something that computer scientists in particular have had a very hard time adjusting to because we think of our field as value neutral.

Because it's just math, it's just code. But in people facing settings, it's not.

RAFFI:

A lot of people say something along the lines of, AI systems just embody the values of those who write it. And the question is whether or not you are cognizant of the values that you're encoding versus you're not cognizant and therefore it just amplifies maybe your predetermined biases.

Do you think that's a correct summarization?

SURESH:

I don't.

RAFFI:

Interesting.

SURESH:

I don't, because the statement, "An AI system embodies the values of those who write it" assumes that I am aware of my values, I am articulating them in the language of code and I'm writing it down. None of that is happening.

So, you know, let me give an example. In a machine learning class, you're taught, the way we train a machine learning model is to come up with a hypothesis-prediction function, and then evaluate how good it was at predicting the outcome. How do you do it? Well, you have the training data. You have your function. You check if the function answers matches what the training data said the answer ought to be, and you just compare them literally and say, well, how many mistakes do you make?

This sounds very value neutral, eh? Just a function. You just calculate the accuracy and you just measure it. No one thinks about this, you just do it. No one is sitting there embodying their values. But what is the value here?

The value here, as we've discovered, is that it's the value, it's a majoritarian value. It's saying what works for the largest number is the correct answer. And we know very well now that if you use this accuracy measure and you have imbalanced classes, then it is more often than not, the the function you produce will optimize for the larger class because it can get more accuracy on that and ignore a small class.

You have not explicitly said you're majoritarian. You have not explicitly said, I don't care about minority groups, but the function you've chosen, because you just were taught that function in class, is doing it.

AI is not embodying your values. It is inadvertently, unintentionally cluelessly capturing things you just haven't thought about.

Music

RAFFI:

We'll be right back... after a quick break.

MIDROLL

Music

Raffi:

I'm back with John Markoff.

John, have we made any progress on this debate between Englebart and McCarthy? The choice between... augmenting human capabilities... and automating away humans? Where do you think we've landed?

John Markoff:

You know, I've struggled with this. I wrote a book called *Machines of Loving Grace*, which now seems like ancient history. It appeared in 2016, where my argument was this is a design choice. That we can design people into systems or we can design people out of systems.

And it's in call centers. That's a classic example for me about how we choose to deploy these technologies. Do you make call center operators smarter, or do you displace call center operators? And I'd love to watch what happens. I mean, I think it's completely up in the air. And it's a design question, and it's a question that the technologists who are going to design these systems now that we have these powerful tools have to make.

And, you know, I, I think that, I mean, first of all, this is such an amazing time. For the first time in two decades, all the pieces are up in the air. And you know what I've always said about Silicon Valley, I've been around for a long time is, um, the visionaries are always wrong. Any time you try to make a prediction in Silicon Valley about something that's more than two or three years in the future, you might as well be writing science fiction. I mean, you know that as well as I do.

But now you can really see it because we have these powerful systems that are gonna be deployed and... I don't know how it's going to, I can't pretend to know how it's going to shake out.

Here's a good example of the dangers. I mean, I wrote the very first article about the worldwide web. It appeared in December 7th of 1993. And I wrote it because I knew a computer scientist by the name of Brian Reid. And he said to me at one point, you know, this, this web thing is going to allow mid-career computer scientists to reinvent their careers because they can quickly communicate with their, with their colleagues. I thought this is a great idea. The lede of my story was: think of it as a map to the buried treasures of the information age.

Well, I got it all right and I got it all wrong. I was thinking of treasure as being information. And within two years, the dot com world had appeared and I was run over by like a Mack truck while, you know, while the dot com world became Main Street of the entire world. Economic commerce moved to the Internet. I didn't know that – I didn't foresee that was going to happen at the speed it was going to happen.

So here we are again. We're at just that moment.

RAFFI:

You've been talking about the developers and the luminaries inside Silicon Valley... how do you think regular people are supposed to grok this and figure out where it's going?

John Markoff:

Yeah.

In February, I was in New Zealand. And I was on the North Island in a little resort town called Taupō. And I turned on FM Rock Radio. And the two DJs were making wisecracks about chatbots. And I thought, holy shit, this is diffused to the entire world in three months. So what does that mean?

So, to me, it means we're still crossing the boundary. And, you know, I think the crank is going to turn very quickly, and these things are going to become more and more accessible. They're going to be built into the heart of every platform we interact with. I mean, language is fundamental. Culture is fundamental. Now, I don't know – it's such an interesting time. And I mean, it is such an interesting time. [LAFF]

RAFFI:

No, absolutely. I mean and that, that point, like language is such a fundamental part of like how we interact. And now if like these machines are interacting with us through that exact same lens,

John Markoff:

Yeah.

RAFFI:

It gets super interesting. How do you think about, as it sort of like gets more commoditized, as more people are sort of using it, like there's a sort of this air of inevitability. I would love to know from your historical perspective, are there other times when we had this air of inevitability of like, this is going to happen, might be out of our control and it's going to happen, or is this time unique?

John Markoff:

Well, I've seen a set of these waves of deployment. I mean, John Doerr said this – John Doerr is a well-known Silicon Valley venture capitalist. He was at Intel. At one point, Doerr was famous for saying that the personal computer industry was the largest legal accumulation of wealth in history. And then about five years later, I heard him saying it again with respect to the internet. Um, and so there's a scale thing. Clearly, this is scaling on steroids and in ways that I think we have yet to predict.

So, [sigh], gosh. Yeah, it's so challenging to make any kind of predictions. But, clearly, this is different. This is a phase change that's important.

So we just saw the crypto bubble and we saw regular people dragged into a Ponzi scheme where the last guy in is always the sucker. And so here we are at that kind of – I mean, this is different, but – people should google chat bots and get rich quick. Every, I mean, there are a million videos already of people using these tools on various schemes.

CLIP: Get rich quick by using ChatGPT vid

This is a huge opportunity, and in this video, I'm gonna help you guys understand this opportunity, and I'm going to show you some examples of how you can take advantage of it, and how you can be first to this trend.

John Markoff:

So people should realize that there's a trust issue. There's a profound trust issue.

In demonstrating ChatGPT to the former Washington bureau chief of the New York Times who had not played with these things before, the first thing you always do is you, you ask, you know, what's my biography? That's what everybody does.

And it was a compact paragraph. Phil Taubman, who was once upon a time the editor of the Stanford Daily a million years ago. He's now George Shultz's biographer. There were four points in this biography that were created, and they all four were inaccurate. Um, It conflated him with his brother. It misstated the year he came to the New York Times. It made a mistake in the name of the person he wrote the biography of; it confused Colin Powell with George Schultz. So it was fundamentally wrong.

So I've found I, I use these things with such great care. Whenever you want to, to know anything, you have to check it. So that's at the heart of these things. And you sort of push the people who have designed them on, how can we make these things accurate?

I think the word "hallucinate," which has become this popular way of describing the inaccuracies is actually a bad word. I mean, I would say something like confabulate instead of hallucinate or error prone because they are profoundly error prone at this point.

That said, in remarkably diverse areas, people are using these things as the intellectual Swiss army tool of the moment. All my friends who code are starting to rely on these things. And I guess if you use this as a productivity tool, I mean, I have to ask you as a technical person, do you still write code? Would you use one of these?

RAFFI:

I do still write code and I do use one of these and it's been a game-changer for me. I literally am writing in programming languages that I'm not fluent in, but I'm writing in it because it's basically helping me along. It's insane.

John Markoff:

And for writers, that's true too, like me, I'm using this as an augmentation tool, but it touches everything. This weekend, I was talking to an artist who teaches a course on Art and AI. I mean AI as a paintbrush, a tool you can use.

And she's preparing this course, and she's been preparing it for six weeks and in the course of those six weeks she's changed the tools she's using three times already. She went from DALL-E to Midjourney and now it turns out that Adobe's about to introduce this thing... so it touches everybody who uses Photoshop and Illustrator.

So that to me gave me this sense of a breadth that you, I've never seen anywhere else that differentiates what's going on.

RAFFI:

You just heard John talk about these things called "hallucinations"-- that's a colorful term that we use for times when Chat GPT or these large language models spit out nonsensical, irrelevant, or blatantly false information. For example, when I ask ChatGPT about Raffi Krikorian's biography, it gets some details weirdly wrong. Like it says that Twitter once bought a company I started. That... didn't happen. So where did it get that from?

One reason why a large language model might sometimes behave like this is because it was trained on bad data. There's plenty of wrong stuff on the internet, and if that's part of what the model used to quote-unquote learn how to respond... then it's gonna spit out wrong stuff. That's why John says we should them "confabulations" rather than hallucinations—falsehoods in the training data will lead to these AIs giving us false responses.

But... that's only one of several possible causes for these hallucinations. We don't really know why hallucinations occur. It could be that hallucinations are something we just have to accept along with the other aspects of "generative AI". Even though it was trained on all that data, it's not feeding you a simple copy. It's more of a mishmash... as best we can tell.

But this phenomenon should give us pause.

And then, there's what happened to Kevin Roose, a technology reporter and podcaster at the New York Times. He was interacting with Bing's AI, based on OpenAI's GPT4, and the conversation went from being... a little weird... to what he called "deeply unsettling."

CLIP:

VOICE 1: At first, Roose says the chatbot seemed useful. Then, he felt things took a surreal turn.

ROOSE: It was moody, it was needy, it was, you know, displaying all these personality traits.

VOICE 1: Roose adding, the bot seemed to be expressing feelings of sadness, yet also declaring its love for him, even going as far as to comment on his marriage, reportedly replying, You're married, but you're not happy. You're married, but you're not satisfied.

RAFFI:

Ok, that's an extreme example, but the very fact that these strange encounters DO occur ought to raise serious questions about just how useful these models really are at this stage.

How much responsibility do the developers have for what their software does? The engineers that designed and trained these LLMs might not be directly causing hallucinations. But... since they decided to release this tech to the public... should they be held responsible?

Take the case of self-driving cars, for example. Tesla was forced to do a recall of some of their vehicles with their "full self-driving" Autopilot software to address concerns that cars in FSD mode weren't stopping at lights or stop signs. And according to [recent reporting in the Washington Post](#), there have been 736 crashes and 17 fatalities resulting from Tesla FSD since 2019.

Software engineers and companies are deploying AI software that could literally kill people on the road.

Who is responsible, and how has our view on responsibility changed over the years?

John Markoff:

I mean, there's always been this carveout that has bothered me in the software industry about liability. And here we are, I mean, who's to blame if these systems go off the rail? And,

you know, the software industry got this liability carveout early on and we haven't sort of adjusted that as a society.

And now... God, are we facing it in steroids. I mean, it's so striking to me. Um, Geoff Hinton just walked away from Google. I think Hinton is a fascinating guy. He has a long history of sort of thinking about the ethical consequences and responsibilities of technologists. I mean, he left America very early on because he didn't want to take funding from the military in Carnegie Mellon. And that was very good.

You know, I wrote a front page story for the New York Times in 2009 when the AI community had their Asilomar moment. They went to Asilomar, they were trying to evoke this pause that the biotech community had taken in 1975 and they wanted to know, is this the moment where they should pause?

CLIP:

With the advent of recombinant DNA techniques in the early 1970s, a debate arose regarding the societal and ethical posed by the capacity to modify the molecular centerpiece of life. In 1975, scientists met for the landmark International Conference on Recombinant DNA molecules, held at Asilomar.

John Markoff:

Well, as it turned out, the AI community didn't pause. But they did begin wringing their hands. People began thinking about the... potential of this technology and they, you know, I hate to say it, they wrang the hands all the way to the bank.

RAFFI:

How sold are you on the impact of AI going forward?

John Markoff:

I, I'm so skeptical. I mean, I've heard that... I mean, and this is a replay of the assertions that the technical community made about self-driving cars. We've been there before.

First of all, the narrative on this stuff is going to be effective by what happens during deployment. And there are going to be bad things that happen. You and I both know that. And those bad things are going to affect, they're going to affect regulatory things, all kinds of things that are going to happen in this path that we cannot foresee now.

But... we're going to go through this path fairly quickly. I.. so this notion of building systems that do what humans do, I've been particularly fascinated by the explosion of the, the effort to automate GPT. There are things like auto GPT, agent GPT, and they have – they're great demos. They are completely dancing dogs. You know, the reporters of the Times who call these people up and say, you know, these are great demos. Do these things work? They don't work yet.

So I don't know, I mean, is it going to be a path like the path to self-driving cars? I mean, I do see cars in San Francisco that roll by me without human beings in them. That is just a stunning, stunning thing, but I still... I, in 2016, 2015, I used to say things like, if a car picks me up in San Francisco and takes me to dinner in Palo Alto, I'm buying, and I haven't lost that bet yet. Maybe I will. It's going to be nip and tuck.

The emotional relationship between machines and systems, the psychology is the thing that's interesting to me. This notion of anthropomorphism is at the center of, for whatever reason of human psychology, we tend to anthropomorphize anything we interact with. You know, whether it's a pet rock or it's our car or our dog, we treat them as if they're conscious, self-aware systems.

And now we have systems that actually appear to be conscious and self-aware, even if they aren't. And how's that gonna play out? ... And if they're trainable, people immediately exploit any weakness and turn them into these monsters. And, uh, that.... I'm concerned about that.

music

RAFFI:

Beyond worrying about how this technology might be used, or abused... there is a whole other class of worries about what AI will do to our *humanity* – whatever we might mean by that term.

Adrienne:

So many smart people have been preparing for this moment, knowing that there would be a turning point with AI, where suddenly it would develop to the point that its capabilities are, are getting more and more astonishing, but also, like, the mainstream access to it is huge.

RAFFI:

Adrienne LaFrance is an investigative reporter who's covered technology, politics, and the media. She's now the executive editor of *The Atlantic* magazine.

Adrienne:

I do think it's different, I think there's probably gonna be, probably bumps along the way, and, you know, will certain aspects develop as fast as some people say? Maybe not. But possibly the most consequential technology humans have ever made. I love technology. I love the internet.

But I also worry a lot about how it's deployed, and especially in a moment like this, where you see sort of, people racing to be fastest. And of course, people wanting to make money off of this. And it's just sort of suddenly kind of like in the air in a really exciting way.

But it, to me, it's this moment where we can sort of stop and say, OK, we're about to live in this world that's being transformed. It's fantastic, really exciting. I genuinely believe AI will save lives. I think the scientific potential is immense. But what do we stand to lose along the way? And are we comfortable with those trade-offs?

And so, my hope is that we can, as humans, not be like, humans are better forever [LAFF], go away robots! But like really take stock of what it means to be human, and what it means for that definition to change in this moment, maybe.

RAFFI:

I mean, so... what IS your take on that? Like, what are the tradeoffs, and does this cause you to reflect on humanity itself?

Adrienne:

Of course, I mean, my actual belief, this is sort of dark, but I think... I think what makes us human is that we die. Like we only get so much time on Earth. And we're compelled to seek knowledge and connection with one another. And I find that to be profoundly mysterious and wonderful.

You get a sense of urgency knowing that you only have so much time on Earth. And I mean, I guess machines can rust or [LAFF] their code can break. But uh, but to me, that's the key difference. It's not art-making. I think you know machines have proved they can make really provocative art.

Certainly, if you look in tech history, there were times when people would have scoffed at using photography at all. But even over the course of cinema, like a film historian could tell you all of the debates over how technology has been used. and certainly even in music, like any art form has had its technological debates over time. And so, I mean, I think a world with AI-enhanced creation is really exciting.

But I think that what humans bring to their endeavors, whether creative or otherwise, is a sense of our own finite existence.

Raffi:

How do you get people to resist the seduction?

Adrienne:

Hmm...

Raffi:

Part of the reason I think this time is different is because now these machines are making a direct assault on things that were so innately human. Like language is such a human thing. And humans have this ability to anthropomorphize a rock. So now you have a rock that talks

to me. Like I'm gonna associate so much quote unquote humanity to this thing. How should people resist that seduction then?

Adrienne:

Right, right. I think it is really seductive. I think you're right. That's the right word for it, certainly. Um, I do worry about the dulling of artistic or intellectual instincts. Like I think that if you were to say, I'm going to outsource everything I do to AI, except for my one favorite thing, which is, I don't know, playing the guitar or something, whatever it is. Like, I think, I don't know.

Certainly there are mundane tasks that AI could make easier for us. Like I see all of those arguments, but I do worry about anthropomorphizing machines and losing sight of what you're actually communicating with. And giving up human interaction? Like, I think about this already with the social web, the people just sort of maintaining a friendship by clicking the Like button and not then taking the time to make the call or pay the visit or go on the trip or whatever.

But in terms of, yeah, the seductiveness, I think one of the biggest things that concerns me about people being seduced into finding perceived real connection with machines is that they will be tangled up in commerce, meaning that likely many of these tools will be built by big companies that want to profit off of people's engagement and, you know, scoop up their data. And you know, we see this in the infrastructure of the modern social web. And you can very easily see how layering AI would make things even more engaging. Just look at like TikTok's "For You" algorithm, you know?

And so engagement or sort of clouding people's ability to see that this isn't a real person on the other side, in a format that is profit-driven, leaves opportunity for real damage and corruption. It's not inherently bad. you can see how it would lead to warped and harmful incentives.

Raffi:

Are you basically painting a picture... or are you thinking about a picture of like a humanity that needs to be saved?

Adrienne:

Um... I have been thinking about it more as a sort of like renaissance of humanism in a way.

Raffi:

Hmm..

Adrienne:

... which sounds maybe grandiose and possibly cheesy, but I just, I just keep thinking about how the most recent sort of technological revolution with the internet and the mobile web

and, you know, the social web. People have just sort of let it wash over them, in a way. Too many people, I think.

And in particular, I'm always fascinated when people just like don't stop to think about how their data is used. It makes me insane. And I, I worry about people assuming that these big tech companies are just so powerful, this is just the way it is. We just have to live with it.

And I really believe that if enough people try to define the world as we would want it to be in terms of how we use this technology, I think you can exact cultural change that hopefully eventually, again, maybe this is me being overly optimistic or something, but the cultural norms are so powerful in terms of how we use these things. And if we can get that right, you know that can then inform questions of regulation, if that's appropriate.

Right now our culture is out of whack, I think.

music

RAFFI:

Adrienne's most recent piece about AI is in the July/August issue of The Atlantic, and it's called "In Defense of Humanity."

On the next episode of Technically Optimistic... How can we ensure that AI is used safely without putting up barriers to innovation?

We'll be talking about the unique challenges of regulating AI... with lawmakers...

CLIP:

SEN. BENNET: You know, I'd love to have the opportunity to have the American people in effect put themselves in negotiation with these largest companies.

technologists...

CLIP:

REP. OBERNOLTE: I don't think it's the government's job to force investment into technologies. But, what we can do is make it clear that there is a future for that investment.

and activists...

CLIP:

TRISTAN HARRIS: One thing I'll say is that I know that Senators in the U.S. have been afraid of talking about the more catastrophic risks that can come from this. I think this is a huge mistake. We have to move the Overton window, so that it's totally

legitimate to question the kind of risks that we're talking about. These are not sci-fi risks.

in episode 2.

RAFFI:

Technically Optimistic is produced by Emerson Collective, with original music by Mattie Safer.

For updates, additional content, and engaging discussions, follow us on social media! You can find us on Instagram, LinkedIn and Facebook ... @EmersonCollective.

I'm Raffi Krikorian. See you next week... on Technically Optimistic.

music