AI and You

Transcript

So there you have, in the words of many of our guests, is where I think the special value of this show is. We raise your IQ – your Interesting Question level. We have some answers, for sure, but questions are where we really suck the marrow out of life, so to speak.

And we indulged in that little self-promotion because this is a special episode, it is our – wait for it - <FANFARE> 100th episode. We will not grow the count by another digit again until the year 2039. That one will be a longer fanfare.

So in this episode I'm going to give a retrospective of the guests and episodes we've had and tell you some of the things I learned from them. Because most of what we talk about is not going to get dated very fast, there are some questions and issues that will have value for another decade or more. So if you're new to the show, I believe there's immense value to going back over those previous episodes. We're feeling our way around this thing called AI like it's the proverbial elephant and we're the proverbial blind men, only instead of six people groping an elephant we have a hundred episodes and 60 guests trying to get a handle on AI. Which is, to be honest, bigger than an elephant, so we're still not done.

So to the episodes, which started in June 2020 when the pandemic was hitting full bore and we all thought it would be over in another couple of months and I started out just giving you an idea of what I hoped to do.

Our first guest, in episodes 2 and 3, was my old friend **Audrey Tang**, who is now the Digital Information Minister for Taiwan. It was so like Audrey to be that generous to lend her name and weight to a completely new show without any idea of what it was going to turn out like. And she is a polymath genius who is at ground zero of one of the boiling points in today's geopolitical world, made only more so now by the war in Ukraine. So she had a lot to say about how she was able to leverage computers and social media, to have what was then the world's best response to the coronavirus pandemic and especially the disinformation that went along with it, since Taiwan is assaulted by Chinese disinformation campaigns constantly. I drew hope from learning that she developed a strategy to defeat that disinformation, even if it was one that I don't think we're likely to replicate in western countries, particularly since one response required the Taiwanese premier to wiggle his bottom at the camera.

Episodes 4 and 5 welcomed science-fiction authors and screenwriters **Judith and Garfield Reeves-Stevens**, who have written novels and screenplays for *Star Trek.* In particular they were producers on the series *Enterprise*, and have developed a number of stories around superintelligent AI. So bringing them in demonstrated how much we were intending to range over this field of AI and how far we would go to find people who could open our eyes to new possibilities. Plus it was amazing how they finished each other's sentences. I called them a "hive mind" and they told me that their nickname on set was the Binars, which you may get if you're a Trekkie.

In 6 and 7, **Ryan D'Arcy** came in, a neuroscientist who spoke at TEDx with me, and he has experience with naked brains while there were still in use. So I asked him what a thought is. I'm afraid that was a bit ambitious. We talked a lot about how brains work because of the parallels we can draw with AI, since researchers who are chasing artificial general intelligence often think we could do that if we can figure out human brains well enough to copy them. Spoiler alert: Not there yet.

In episodes 8 and 9 **Richard Foster-Fletcher** was on the show, he is the founder of MKAI, the Milton Keynes AI group, which is now the premier group, to my knowledge, focusing on understanding and promoting inclusivity in AI and exposing where it is lacking. Richard really thinks big; he gravitates towards concepts like the Kardashev scale, which measures how much a civilization has harnessed the power of a planet, or a star, or its galaxy, and obviously we have a way to go on that scale. But we were talking about how an organization could detect and remediate bias within its AI. Really interesting framing in terms of how does he want to see the world twenty years from now and how can we get there.

In episodes 10 and 11 I was joined by **Kristóf Kovács**, the international supervisory psychologist for Mensa, the organization for people who score in the top 2 percent on intelligence tests. Who better to tell us what intelligence means than someone who has to decide that for a group you have to be intelligent to get into? And we talked a lot about the different types of intelligence, the problems with measuring it, and the general, or g, factor. I learned about the Flynn Effect there – we're actually collectively getting smarter. Hard to believe at times, I know, but it's true.

In 12 and 13 falsely came on, who is CEO of Embodied, Inc., which makes the Moxie robot, which is about the size of a baby penguin and it helps children get past developmental issues. I found it amazing that a robot could be made to establish a relationship with not only a child, but one that needed some kind of therapeutic intervention; and I really wanted to know how they did that. Paolo had worked on the Roomba, and space missions, and he told us how simultaneous localization and mapping works and we discussed all kinds of robots.

Episodes 14 and 15 saw falsely, doctor of philosophy of philosophy, and AI ethicist, at the time with the Centre for the Study of Existential Risk and the Leverhulme Center for the Future of Intelligence in Cambridge, England, come on the show where we started grappling with some of the more abstract philosophical aspects of AI – yes, the Trolley Problem did come into it – but what was really interesting was that she in no way expected to be involved in Ai when she started getting into philosophy, it was obviously a relatively quiet and staid field until AI developed enough ethical issues to overlap with it and suddenly software companies are knocking on Karina's door and she's being interviewed on the BBC every other week and finds herself in an incredibly topical and superheated environment. We talked some about the Extended Mind Hypothesis, which she specializes in, which says that anything outside your body that participates in your cognition, like a smartphone, or a notebook, can be considered part of your mind. And we talked about the Chinese Room and the Value Alignment Problem. So here she was, just learning about the classical philosophers, and within a few years finds herself writing papers on the long-term existential threat of AI. Talk about interesting times.

In 16 and 17, **Roman Yampolskiy** was here, and he's professor of computer science at the University of Kentucky at Louisville, an expert on computer security and AI safety. And we dug into the Control Problem – something that's an indeterminate but probably large amount of time in the future, but some smart people are thinking about it now because it's a huge problem, which is how do we control, or try

to control, AI when it becomes superintelligent? And Roman told us about a paper he had just published on that, and gave us the academic analysis and scenarios of superintelligence, and he's very experienced at talking about this and made it very understandable.

Episodes 18 and 19 saw **Tony Czarnecki** with us, of the London Futurists, and author of the transhumanist *Posthumans* series of books culminating in *Becoming a Butterfly.* So now we got another far-ranging look at the effect of AI, primarily in terms of economics, and from a Eurocentric viewpoint, how the Common Market might evolve or collapse, what's happening and what might happen with climate change.

In 20 and 21 **David Wood** joined us, the leader of the London Futurists, and he's incredibly active and prolific as a futurist, being activist about our need to exercise a duty of care about the future, and what it's like to work at influencing how lawmakers and others shape our future.

Then in 22 and 23 I had **Pamela McCorduck**, the historian of AI, author of *Machines Who Think*, the definitive history of artificial intelligence, and its recent sequel, *This Could Be Important*. She is full of stories about the founders of the field, she talked with most of them, her sense of humor and the way she deflates any balloon of self-importance are simply delightful. And I spent a lot of the time discussing how what she learned could inform and direct what we're doing today. I learned a lot, some of it wasn't too surprising, like the sexism that was rampant in the field in past decades. And other things were illuminating, like how those founding fathers were thinking when they thought they would knock off artificial general intelligence in the summer of 1956. And now with symbolic AI, which had been all but ignored a decade ago while everyone was frantic about deep learning, now symbolic AI is getting more attention again, it was interesting to hear about the early days of that.

Episode 24 was just me, and I was talking about what I call the most important unanswered question around the existential implications of AI, which is *when* we will get artificial super intelligence, and – spoiler alert – no one knows, but I dug into it as much as I can, not to try and narrow down something that no one has been able to narrow down, but exploring the edges of that question, why it's important, and saying as much as can be said about why it's important and what we should do about it to help you get familiar with all those issues and what can we figure out that might help to one day answer that question, and be knowledgeable about those issues.

25 and 26 was **Thomas Homer-Dixon**, founder of the Cascade Institute, an existential risk think tank at Royal Roads University in British Columbia, and author of the book *Commanding Hope,* which was a practical look at how we need to manage our psychology to face what could be overwhelming existential threats from various directions. Really interesting discussions about the inherent problems of complexity, what blind spots in us it exploits, and bringing in one of my favorite topics, Systems Thinking, as popularized by Peter Senge. And how individuals can be tipping points to make global change happen.

27 was me talking about my TEDx speaking experience, that was for any of you who think you might want to do one of those or are just curious what it's like, especially for someone like me who doesn't have any natural speaking talent and had to learn it all from scratch. That TEDx talk was titled, "How to save us from being left behind by AI," and it focused on the way our pace of life is being increasingly dictated by our technology, how that's getting out of our hands and how to take back control.

In 28 and 29 I had AI in marketing expert **Katie King** on, and – you know, I'm constantly getting requests from PR people and agents pitching someone to come on the show to talk about their company and their product and I turn most of them down because all they want to do is marketing – "what do you think about that Peter, aren't you intrigued now, but wait, there's more!" and that's not where we're at, that's not what we're about, right – but this wasn't *doing* marketing, this wasn't promoting a company or a product, this was about the trends in industry that AI is driving, how they shape up across national boundaries, how education is and should be changing, and how AI is being applied in different sectors.

Then we got into AI and the law, in episodes 30 and 31, with professor **Ted Parson** from UCLA's Program on Understanding Law, Science and Evidence in AI. He's been at Harvard, he's consulted for the White House, and now he looks at how AI can intersect with the law, and we talked about self-driving cars – okay, the Trolley Problem reared its head - social media and disinformation, product liability, the economy, and justice, seen through this lens of law and public policy when AI is driving those things at a level beyond what we've got today. When I listed all those things just now I couldn't believe that we actually gave them all a fair shake, but we did.

Episode 32 was me deconstructing that last TEDx talk, because it had a lot packed into 14 minutes, so I kind of annotated it with a lot of side exploration to make it more multidimensional. And then in 33 I explained some of the terminology and function of neural nets, as a primer for a lot of the things we talk about.

In 34 and 35 **Michael Wooldridge**, head of the computer science department at Oxford University, came on, and he gave a perspective on what it's like dealing with the sensationalism about AI when you're in the position of teaching people how to make it work, how to interpret the sort of things you read in the popular press, so we got a useful dose of realism there about artificial general intelligence, or AGI, and he also told us about multiagent systems and his research into those. One day my Siri may be able to talk to his Siri.

In 36 and 37 **Steve Shwartz**, author of the recent AI book *Evil Robots, Killer Computers, and Other Myths*, talked about the impact of automation on jobs, and the challenge of common sense reasoning. Common sense is a term we love to throw around as though it's something that's just so clear what it means, and it seems to be clear to us and yet we can't define it clearly. Common sense is not common to define. My idea of something being clear is that you can explain it to a computer, and of course we can't.

38 and 39 saw **Beth Singler**, an anthropologist at Cambridge University, come on the show, and she has made these thought-provoking movies on YouTube about our relationships with AI, and you can find one of them by looking up "Pain in the Machine" on YouTube and it'll lead you to the rest. One of those videos contains a memorable scene with a self-aware AI that's embodied itself as a hologram of a girl, and a human is talking with it to try and figure out what its intentions are. Beth had really interesting takes on fiction, especially *Blade Runner,* and analyzing *Westworld* in the context of our culture, and also talking about how the media shapes our views and unconscious attitudes towards AI. I don't know if I'd make a good anthropologist, but I do find it absolutely fascinating and I could listen to Beth talk about these things for hours.

Then in 40 and 41 we had **Peter Asaro,** a professor and co-founder and co-chair of the International Committee for Robot Arms Control, which, after talking about Terminators with Beth, might sound all alarmist, but is very serious and we got into the issues around lethal autonomous weapons. Again, none

of this is getting dated, right, those issues are as current and meaningful now as they were when we talked about them, maybe more so with warfare ramping up in Ukraine and who knows where next. And there are several obvious questions about arms control for autonomous weapons, which I asked, like don't they reduce the human damage of warfare by being smart enough to eliminate collateral damage, and I quickly found out how people like Peter have gone beyond that surface layer and we got this big tour of the different agreements and treaties and laws, country by country, pertaining to lethal robots.

In 42 and 43 we had science fiction author **David Gerrold**, who by the way was just given the 2022 Heinlein Award, and David wrote the classic *Star Trek* episode "The Trouble with Tribbles," plus other Trek episodes and stories and many novels and series, and we primarily chatted because he had written a novel about an AI becoming conscious, it was called *When H.A.R.L.I.E. Was One*, and although it came out decades ago he rewrote it more recently to bring it up to date, and he has some of the best thinking about what it would really be like to encounter a self-aware, supersmart, artificial intelligence acting as a mirror for our humanity, and what would it be like to look into that mirror? David's been doing that kind of thinking and writing for fifty years, we just flew out on wings of speculation to some very far-off places. But at least for me, every time I do that, I come back with something that's applicable right now, like when we were talking about how in his story the first AI designed another AI that would be far more capable, I found myself thinking, that sounds like we could do something like that now, and then in the headlines there are already cases of AI helping in the design of computers and more AI.

Then we really came back to Earth in 44 and 45 with venture capitalist **Rob May** talking about where the smart money is going now, trends in hardware, whether we're in an AI bubble, why he's excited about brain-machine interfaces, and the companies he's following in that area. And I had Rob right after David to make the point that AI spans this huge range of genres and yet if you're interested in AI, it makes perfect sense to talk to both of these people, and everyone else who's been on the show.

46 and 47 saw **Rajiv Malhotra**, author of *Artificial Intelligence and the Future of Power: 5 Battlegrounds*, which is a very useful perspective on the role of technology in neocolonialism and global power structures, and he gave us a lot of ideas about how a new industrial revolution would disadvantage the same people and ethnic groups the last one did. So if you're involved in or interested in Diversity, Equity, and Inclusion, this is a key one to listen to.

In 48 and 49 **Phil Hall** was on the show to talk about his use of AI in performance art; he created something called the EchoBorg, which is a show where people have a conversation with an AI but it's mediated through an actress, Marie Helene Boyd, who speaks what the AI says and so you're talking with her and you have all the affordances a human brings to a conversation, only you're really talking to a chatbot. And that was fascinating, because on the one hand it's kind of a circus act, but on the other you can learn a lot about human attitudes towards AI from it, which is why the EchoBorg made an appearance in front of Britain's All-Party Parliamentary Group on AI and you got to see it talking with members of the House of Lords and the House of Commons. That was memorable.

50 and 51, **Ryan Abbott**, law professor, author of *The Reasonable Robot,* with a really interesting angle on intellectual property – when should an AI own a patent or copyright? And the first reaction I had was, "Why would an AI care about rights, what's it going to do with the royalties, retire?" But that's not the point at all. There are inventions being made right now where AI forms such a part of the process that none of the humans involved believe that their name deserves to be on the patent, and this isn't a matter of overweening modesty; in some jurisdictions it is illegal to falsely place yourself on a patent.

And Ryan has an AI that he's asserted patent rights on behalf of, and taken it to courts, and in some cases, won.

52 and 53 was our one-year anniversary, so we had a special panel about AI in fiction with my old friends, **Jim Gifford** and **Robert James**, experts in various fields of science fiction, and some of my co-conspirators in producing the 2007 centennial convention dedicated to the author Robert Heinlein. And as we learned in the episode with Beth Singler, there is a lot about how fiction reveals our attitudes and unconscious beliefs, and cultural attitudes towards AI that can reveal how we will change as a society as AI gets more and more capable. So we listed all the usual suspects, and some much rarer ones, and talked about their cultural impact.

In 54 and 55, engineer **Tony Gillespie** joined us, and he has written a book about systems engineering for automation, and has a lengthy history in designing automated weapons and other systems so that they can follow international humanitarian law such as the Geneva Convention. And that blew my mind, because I didn't think it was possible to codify anything as social as the Geneva Convention in software or hardware. But I learned how when smart people think about hard problems for long enough, even a problem as seemingly intractable as this one can be at least partially solved.

In episode 56 I had **Przemek Chojecki** on, an entrepreneur with a mathematics PhD, and what it was like for him to make the leap from academia to business and developing a product that made writing easier through AI.

In 57 and 58, **Charles Radclyffe**, TEDx speaker and the founder of Ethicsgrade.io was on the show. Now ESG gets a lot of air time these days – that's environmental, social, and governance, which businesses want to score highly on these days, and Charles looks at doing that – that's the 'grading' part – through the lens of AI. So for example, they give Skype an A and AirBnB a D. Now – why? And how do you expand the issue of ethics along all its many dimensions in a way that you can do some good and provide a valuable service to enough companies to get paid for it? Really interesting ideas, like the possibility of AI developers undergoing professional certification the way doctors and so on do, that we discussed.

In episode 59, professor **Kakia Chatsiou** from the University of Suffolk came on the show, and she does natural language processing, and has done research on analyzing statements made by politicians. Which certainly sounds like anyone doing that would deserve considerable reward for having to listen to politicians that much. We learned a lot about natural language understanding, word2vec and associations, and even one of the tricks my children play on me.

60 - **Tomas Mikolov**, actually one of the key contributors to word2vec, and a researcher with funding from OpenAI, came to give us an unvarnished evaluation regarding research and claims in artificial general intelligence, and gave us an insight into how neural networks work and helped me get a better understanding of how AIs like AlphaZero work.

In 61 and 62, **Todd Litman**, director of the Victoria transport Policy Institute, joined us to talk about how public transport policy may evolve to adapt to autonomous vehicles. A really interesting conversation about what's really possible for AVs, versus a considerable amount of hype that is of course out there at the moment. Some of his examples involved zombie kangaroos; you'll just have to listen to it to get the context.

In 63, **Sathish Sankarpandi**, chief data scientist for orbital media, came to talk to us about the use of AI in healthcare. And in 64, **Amit Gupta**, the inventor of Sudowrite, was on the show. Sudowrite is a tool for helping writers overcome writers block, and it makes creative suggestions for what they could type next. It is extremely capable and inventive, I love demonstrating it.

65 and 66: Neuroscientist **Olav Krigolson**, another TEDx speaker, joined me to talk about research into brain waves. He is able to monitor brain waves and tell when someone is too tired or distracted to be able to work properly, and this has implications and uses for projects like Mars colonization, which he has actually worked on. I learned some things about neuro scientists in this interview, like how they like to experiment on themselves; Olav likes to demonstrate how an electric shock to the Broca's area of the brain can introduce temporary aphasia, or inability to speak properly. I wondered if this was an occupational hazard of working in his lab, and he said that he only did it to himself. And of course, understanding more about brain waves helps us understand more about how we might create artificial general intelligence, since we might model one after the other.

In 67, **Olivier Caron-Lizotte** came on the show; he runs a stable of contract Ai developers and I wanted to know how that works and what's particularly hot or challenging about it these days.

And in 68, **Daniel DeMillard**, who was once with the IBM Watson project and is now the CTO of FoodSpace, told us how AI can be used for making nutritional recommendations, and really what the state of AI is in the sort of general capabilities that you need to make that happen, when you need some degree of understanding and not just synonym matching or pattern finding. So that took us into natural language processing and self-driving cars. But I was particularly interested in the model for how something like this worked when the data was *here*, the customer was a retailer *here*, and the users were the retailer's customers over *here*, how do you put all those moving parts together.

69 was a solo episode with me talking about disinformation. A year before this episode aired I was quite down about what was happening with this information, because it was not talked about very much, and there was little vocabulary for understanding and describing what it was and what it was doing to us. By this point things had gotten better, and research was being done like an excellent project by the Wall Street Journal that visualized how the TikTok algorithm recognized people's tendencies and took them in extreme directions. The same process happens on YouTube and Facebook, and it's not deliberate; it is the result of implementing a machine learning algorithm to maximize the amount of time people spend on a social media platform by seeing what they pause on, and not understanding that that also happens when they are shown something that makes them angry, like conspiracy theories. So as Zeynep Tufecki said, it seems like you are never hardcore enough for YouTube. And it doesn't help that Russia, in particular, realized that they could exploit this through targeted postings that don't violate any law, just exercising freedom of speech, but with the aim of fracturing western society, at which they've been unreservedly successful.

We got into chess in 70 and 71, when international grandmaster **Jonathan Rowson** joined us. He is also a philosopher and author of *The Moves That Matter: A Chess Grandmaster on the Game of Life*. I wanted to understand what goes on in his head when he plays chess, because I suck at it, and then we talked about how AI approaches the game completely differently, because it's using a pattern-based approach, whereas a human grandmaster is constructing grand strategies. What I learned the most about is how humans work with computers in what's called centaur chess to improve how they understand and enjoy chess, even though they will never beat the best computer. And that's extendable to how we will work

with computers at all kinds of other tasks – legal actions, for instance – and Jonathan had the high level perspective from his philosophy background to set this in context for us.

 In 72 and 73, **Kordel France**, CEO of Seekar Technologies came on, and he had done all kinds of things. I learned about autonomous tractors in farming, I learned about rifle sights that are smart enough to lock out the trigger if you're aiming at an endangered species, and we were talking about what consciousness in AI means or would mean. He really illustrated how much this show is about technology and consequences and the sort of thinker that we want on the show, because as I said we get approached by a lot of people just wanting to talk up their marketing book, and Kordel's an entrepreneur, he could easily have done that, but instead he gave us all this interesting thinking about the Turing Test and computer vision and then we got into how he'd used AI to help in psychological diagnosis, like depression and schizophrenia, and then AI being used for COVID diagnosis through X-Rays using a smartphone app. Really covered an amazing amount of territory.

Now, how do you get AI to explain what a decision it's made when the honest truth about that is that a million weights got set in a neural network just so, and that magically causes it to make the right decisions? That sounded to me like a fundamentally intractable problem, until in episodes 74 and 75 I talked with **Michael Hind**, distinguished research staff member in IBM's Yorktown Heights facility. Now this is pretty important because AI can be trained to make decisions in situations that affect people's lives, like granting a loan or adjusting an insurance claim, and it can be shown to be doing as well as or better than humans, but that's not good enough, because when it comes to those kinds of situations, people want an explanation of any decision they don't like. And he explained some of the ways they can now do that, really cutting-edge research, like learning different explanations as part of the model and learning which ones to associate with a particular decision. I'm butchering that a bit for the sake of time here of course.

In 76 and 77 I was joined from London by **Alexandra Mousavizadeh** of Tortoise Media, who founded the Global AI Index, the Responsibility100 Index, and the Global Disinformation Index, which are all instruments with considerable data science behind them that help decision makers assess the levels at which countries and companies are expressing themselves in the areas of AI development and corporate responsibility. And it was also another example of how we've made progress in disinformation since it started bothering me, to not just name it but now have this metric that Alexandra started that names the worst offenders in disinformation using objective numeric criteria and is used to draw the attention of the advertisers on those platforms and ask, "hey, do you want to pay to have your name going out next to stuff like this?" And with the AI Index she had some very useful perspectives on the relative positions and movements of countries on the global stage and we talked about China in particular since they've announced their intention to overtake the US to lead the world in all aspects of AI by 2030.

78 and 79 saw **John Zerilli**, philosopher at the University of Oxford, cognitive scientist, and author/editor of A Citizen's Guide to Artificial Intelligence, joining us. And he has this angle on grass roots political awareness of AI, so the book gets into the areas of Transparency, Bias, Responsibility and Liability, Control, Privacy, Autonomy, Algorithms in Government, Employment, Oversight and Regulation, and we talked about those in the context of their impact on the average person in terms of things like social media platforms and their governance, data anonymization, the right to your personal data, all from a fairly rigorous academic viewpoint.

Episode 80 was the last one of 2021 so I had a panel on for a special episode predicting AI trends for 2022 with returning guests **Katie King** and **Richard Foster-Fletcher**, and new guests **Ben Goertzel**, author of *Ten Years to the Singularity If We Really, Really Try,* and **Prashant Natarajan**, VP of strategy and products at H2O.ai. What's going to happen in AI in 2022? Listen to the episode and find out, or see how close we were!

In episode 81 **Tannya Jajal**, author of the new book , Thinking Machines: AI and the Intelligence Explosion, joined from Dubai to talk about her work with high school students through something called Awecademy, and being the UAE Chapter Lead for the Global Women in Tech Movement. So we talked a lot about messaging, how to get through to audiences effectively, and how important education about AI is.

In episode 82 we went back to IBM Yorktown Heights to talk to distinguished researcher **Kush Varshney**, author of the book *Trustworthy Machine Learning,* and that was what I wanted to talk about. I thought it very noteworthy that AI is maturing to the point that there are people in roles specifically named for AI trustworthiness, and I was very interested in how Kush had some perspectives about how the field may need to mature even more, like with black-box-type instrumentation. Lots of parallels with commercial aviation being drawn in this episode.

In episode 83, **René Morkos**, who had applied an AI PhD to construction, came on to talk about how you could optimize the utilization of a construction site through AI.

In 84 and 85 we had science fiction writer **David Brin**, author of *Earth,* the *Uplift* series, and a *Foundation* sequel, really expanding our minds with incredible ideas – you might think a science fiction author would only be talking about things in the distant future, but he had novel ideas for how to solve immediate problems like privacy and advertising on the Internet. We talked about so much - Jefferson, Einstein, Periclean Athens, prefrontal lobes, and that was just in the first five minutes. To pick one thing I learned from that would be damning the interview with faint praise.

In 86 and 87 **Stuart Russell** joined us, AI professor at UC Berkeley, coauthor of the most widely used textbook in artificial intelligence in universities, spokesperson for the famous *Slaughterbots* video, the 2021 BBC Reith lecturer in the UK, and an activist on the world stage in the conversation on lethal autonomous weapons and the public narrative and understanding of AI. No one has a wider reach in gauging how to communicate about AI to mass audiences and I was keenly interested in what he had to say about how that narrative was evolving and how people are perceiving AI. I would say that most teachers of AI do not take a position on existential risk or they denigrate that possibility, so when Stuart does warn us of scenarios like the Control Problem and the Value Alignment Problem, he has credentials that make it impossible to ignore him. And we talked about that of course. I could listen to many of these interviews multiple times and this one was certainly in that category. Stuart is to AI as Neil DeGrasse Tyson is to astrophysics.

In episode 88 we were spanning four time zones talking about AI in Music with **Bob Sturm**, professor in Sweden, **Dorrien Herremans**, professor in Singapore, and **Henrik Vincent Koops**, researcher at Utrecht University, and co-organizer of the AI Song Contest. If you've heard of the Eurovision Song Contest, okay – this is nothing like that. No one getting dressed as a giant butterfly to belt out a song here. Of course there are introspective questions about creativity raised whenever AI gets into music, which it is doing more and more ably now, and my guests opened my eyes to how much more that's being

commercialized now, how good it is, some of the applications and companies, and the sort of strange things you can do like make up a new Beethoven piece in the style of Gershwin played on hammered dulcimer and bongo drums.

In 89 and 90 we met **David Danks**, professor of philosophy, psychology, cognitive science, and machine learning at UC San Diego, who's really good at explaining AI, obviously through a lot of practice teaching. And he thinks a lot about how policy should evolve to help us with some of the problems that AI can cause, like disinformation and other ways that it can change our preferences. One of the things we talked about was the A-Level algorithm disaster in England the previous year.

In 91 and 92, **Ben Shneiderman** was on the show, professor at the University of Maryland – I guess we have a lot of professors on the show, don't we? And he's the author of the classic book on computer user interfaces, the bible in that field if you will, and now a new book, *Human-Centered AI*, which extends the idea of user interfaces, which used to be bound up in things like affordances and menu design, and button layouts, to how we interact with AI. And we learned that this is a very hot field, and again, one of the goals of the podcast is to give anyone looking at a new career an idea of what it's like in different fields, and if you want to get into one that's exploding right now, human-centered AI is one of them.

We went to the other side of the classroom in 93 and 94 when I talked with **Shea Sullivan** and **Hannah Grubbs** who are students at North Central University in Minnesota but also members of the student-run Institute for Digital Humanity, which they started after one of their professors told a class, here's what's going on with robots in the world today, what do you want to do about it? And I love students for how they just turn thought into action. The energy emanating from their web site just crackles, from the brash headline, "The '80s called. They want their dystopian robot overlords back" to the list of schools and companies they've partnered with, like working with Netflix to construct a learning module around the documentary *Coded Bias,* which is about racial bias within AI. So this interview was a total blueprint for student activism on current social issues triggered by AI.

In 95, computer historian **George Dyson** came on the show, author of *Darwin Among the Machines* and *Turing's Cathedral*, and *Analogia*, to give some perspectives from the dawn of the computing age, because he's made extensive studies of and written about pioneers such as John von Neuman. The modern computer architecture didn't just emerge one day from Zeus' forehead, nor was it the only way that computers could have been designed. We had a bit of VHS vs Betamax discussion, and talked about how much the early history of computers depended on the atomic weapons program and the space program. And another of George's interests, the Orion Project to propel spacecraft using atomic bombs. Didn't happen. So I was not expecting to hear someone make the case for analog computers these days, but guess what, that's exactly what George did, and it may well be a direction that could give someone a strategic advantage today because no one's exploring it.

What do babies think? In 96 and 97 we had **Alison Gopnik**, professor of psychology at UC Berkeley, and she knows, better than anyone, how babies think, and tells us how that can help our planning on developing artificial general intelligence. What I learned that surprised me was that there are things babies, young children, are better at than adults are, at certain kinds of thinking, and that gives us some useful directions on improving our own thinking. I made a connection with the way some genius scientists behaved.

And in our most recent two episodes, futurist and speaker **Calum Chace** came on, author of *The Economic Singularity*, to explain what the economic singularity is, how it contrasts with Ray Kurzweil's singularity, and w had a fascinating discussion about how AI could be used in the metaverse, which is increasingly a thing, even if I haven't figured out what to do with it yet.

And that brings us up to today. 100 episodes, started when we thought we were nearing the end of the pandemic and things would go back to normal really soon. Ha.

One of the goals of interviewing people from such a diverse set of fields is to expose you to so many different ways you can work in AI and get an idea of how you would fit, if you're thinking about careers. You've just heard how incredibly diverse the experts we have on this show are and how much you can learn from them. I hope you'll go and pick out the ones that caught your interest because just because they weren't recorded yesterday doesn't make any difference to how useful and relevant they're going to remain for years to come.

In today's news ripped from the headlines about AI, you know what the GDPR is? It's the reason that every site you go to now has a popup saying click here to enable cookies or this site won't work for you. Long story about that. Well, now there is a new AI tool that might make a difference to that experience. It's called CookieEnforcer, and is was built by researchers from Google and the University of Wisconsin-Madison. It analyzes those popups, figures out what actions will disable all unnecessary cookies, and selects those settings. I would love to tell you what it was like to use that tool, but it isn't available to the public yet. But they're planning on making it available as a browser extension. So keep an eye open for CookieEnforcer.

Next week, my guest will be Bryant Cruse, former Naval Aviator, and then system engineer on the Hubble Space Telescope, now founder of the AI company New Sapience, which is developing new technology in the artificial general intelligence space. That's next week on *AI and You*.

Until then, remember: no matter how much computers learn how to do, it's how we come together as *humans* that matters.

http://aiandyou.net