[MUSIC]

**OREN ETZIONI:** Whose responsibility is it? The responsibility and liability has to ultimately rest with a person. You can’t say, “Hey, you know, look, my car ran you over, it’s an AI car, I don’t know what it did, it’s not my fault, right?” You as the driver or maybe it’s the manufacturer if there’s some malfunction, but people have to be responsible for the behavior of the machines.

[MUSIC]

**KEVIN SCOTT:** Hi, everyone. Welcome to Behind the Tech. I'm your host, Kevin Scott, Chief Technology Officer for Microsoft.

In this podcast, we're going to get behind the tech. We'll talk with some of the people who have made our modern tech world possible and understand what motivated them to create what they did. So, join me to maybe learn a little bit about the history of computing and get a few behind-the-scenes insights into what's happening today. Stick around.

[MUSIC]

**CHRISTINA WARREN:** Hello, and welcome to Behind the Tech. I’m Christina Warren, senior cloud advocate at Microsoft.

**KEVIN SCOTT:** And I’m Kevin Scott.

**CHRISTINA WARREN:** Today on the show, our guest is Oren Etzioni. Oren is a professor, entrepreneur, and is the chief executive officer for the Allen Institute for AI. So, Kevin, I’m guessing that you guys have already crossed paths in your professional pursuits.

**KEVIN SCOTT:** Yeah, I’ve been lucky enough to know Oren for the past several years. We met after I became CTO at Microsoft, but I’ve been very, very well aware of Oren professionally for many, many years.

The thing he probably doesn’t know is the University of Washington, where he still is a professor, was one of the places that I wanted to go to graduate school. So, the entire time I was in undergrad, I had a copy of the UW computer science course catalog sitting on my desk as an aspirational thing. I was following his work pretty closely back then, which you know, may sound a little bit creepy, I guess. (Laughter.)

**CHRISTINA WARREN:** Well, you know, no, I don’t think it’s creepy. I mean, you’re a fan, and I think that’s awesome that you know each other now, and now we’re going to be able to get into the interview and hear you guys talk.

**KEVIN SCOTT:** Yeah, Oren is – he’s really an amazing teacher, leader, and scholar in the work that he’s doing with the Allen Institute for AI is really incredible. So, I’m super glad to have him on the show today.

**CHRISTINA WARREN:** Likewise. Likewise. Well, let’s get to the interview.

[MUSIC]

**KEVIN SCOTT:** Our guest today is Dr. Oren Etzioni. Oren is chief executive officer at the Allen Institute for Artificial Intelligence. He’s been professor at the University of Washington’s Computer Science Department since 1991. His awards include Seattle’s Geek of the year in 2013 and he’s founded or co-founded several companies including Farecast and Decide.com.

Oren has helped pioneer metasearch, online comparison shopping, machine reading, and open information extraction. Welcome to the show.

**OREN ETZIONI:** Thank you, Kevin, it’s a real pleasure.

**KEVIN SCOTT:** So, I would love to get started by talking a little bit about how you got interested in science and engineering in the first place. Was it when you were a little kid or later on in life?

**OREN ETZIONI:** Well, both my parents are sociologists, actually, professors of sociology. And so, I think instinctively, subconsciously, I ran away from that as far as I could. Like some of your other guests, like Daphne Koller and I think yourself, I discovered that magical machine via the TRS-80 and started playing with it. It was just really fun and really new, and I kinda could sense something very powerful there.

But where I really got engaged was when I read the book *Godel, Escher, Bach*, which many of us did back in the day. I got kind of a whiff of the questions that we could be asking and that we could be studying using computers are really some of the most fundamental intellectual questions of all time. What is the origin of the universe? What is the nature of intelligence? How do we build an intelligent machine?

**KEVIN SCOTT:** Yeah. I mean, it’s really interesting that so many of us have such an interestingly similar experience. I can’t remember – I think I was in college when I read Hofstadter’s book the first time – or tried to read it – because it is intellectually dense. But it really had a remarkable impact on a bunch of people. And I’m guessing it still sort of informs some of the things that you’re doing today, because you know, in a sense, what all of us are doing a little bit with the pursuit of artificial intelligence is not just figuring out what the machines are capable of, but like what does that say about human intelligence itself, which we really don’t understand in a deep sense?

**OREN ETZIONI:** That’s exactly right. One way to think about artificial intelligence or I’ll just use “AI” for short, is really it’s almost a funny pun or homonym. It really, in my mind, refers to two very different things. One is a set of technological capabilities, which I’m sure we’ll talk about some more, but are getting increasingly powerful with speech recognition, with facial recognition, with things like GPT-3, and at the same time, it’s a kind of vision, right, of how do we build human-level intelligence, artificial general intelligence, and so on.

And those two things are actually very different in the sense – in my opinion – in the sense that the technology is progressing very well, but it’s still very, very far off from the ultimate vision of the field.

**KEVIN SCOTT:** Yeah, well, and you know, I think the interesting thing that we all have seen over the past several decades is we keep solving these problems that we think are reflective of high human intelligence, and as soon as we’ve solved them, we decide that, “Oh, well, maybe that wasn’t really what is the core of human intelligence.”

You know, I remember when it was chess playing and then it was Go playing and then it’s these feats of perception that we’ve been able to do and, you know, like, there are still these things that are – for human beings, like, very, very easy that I don’t think we think of as being cognitively sophisticated that completely elude the ability of machines – common sense reasoning, navigating physical environments, and all sorts of fun stuff.

So, you know, I do think we have been confused about what the definition of intelligence is for a very long while now, which makes going after it interesting. (Laughter.)

**OREN ETZIONI:** Well, what you’re saying is so true, right? There are writings in the ‘60s and the ‘70s and later talking about, say, chess, right, as the pinnacle of intelligence. And you can sort of see why, right? For most people, playing Grandmaster Chess is an incredible feat of intelligence. Very elusive. And then to find out that the machine can do it is startling.

At the same time, there is something called Moravec’s Paradox, right, that says things that are easy for the machine are hard for people – like playing champion-level chess or Go. Conversely, things that are easy for people – like you said, common sense, crossing the street, driving a car – most people can do it with reasonable success, and machines are still surprisingly far away.

I would say, however, that one remarkable thing, which of course you’re familiar with, is the articulation of “How can we tell if a machine achieved intelligence?” by Alan Turing in the ‘50s, and he devised the Turing Test. And while the Turing Test sometimes is misapplied, it becomes – so, to define it, right, the Turing Test is you’re communicating with somebody, let’s say over, I don’t know, Twitter or Slack or Microsoft Teams, I should say – whatever it is – and you don’t know whether it’s a person or a machine, and you have to try and guess.

When you can’t tell them apart, then the machine is said to have passed the Turing Test. Now the thing is, if it’s misapplied, it becomes a test of human gullibility, right? We can trick people into thinking, “Oh, yeah, I’m talking to a person.” But if it’s done right, if you subject it to a rigorous test by, say, a panel of experts, then it’s actually quite an achievement, right, to have a machine that behaves intelligently, et cetera, et cetera.

So, I think we have that notion. The problem is that we take these narrow slices, like playing chess or like, I don’t know, solving the Rubik’s Cube. And it’s not surprising, by the way, that they’re often in artificial domains, that they’re often games and such. And we say, “Oh, that’s intelligence.” And then we find out, well, not quite.

**KEVIN SCOTT:** Well, and you know, I think yeah, you pointing out that they are in these artificial domains is an interesting thing for all of us to recognize about the state that AI is in right now, and where it is likely to be in the near future. So, I think some of the things that we may think are insulated from encroachment by AI, like a bunch of white collar work, for instance, is actually far more likely to have AI impact it than things like manual labor, for instance. Where, because we haven’t solved these problems where the AI has to interface with the physical world, it just – like, we haven’t even figured out the basic experimentation loop for solving that problem.

So, like, we can’t iterate there nearly as fast as we can in these artificial domains, where progress is still pretty good. I don’t know whether you would agree with that or not.

**OREN ETZIONI:** I think it’s a very rich topic, but I definitely agree with what you’re saying, that the speed of experimentation and iteration makes a huge difference. So if you’re dealing with a robot, say, right? And every experiment that goes awry, you can break a robot arm, or god forbid, hurt somebody, obviously, that slows things down so people work in simulation. Then, the simulation doesn’t necessarily map to the real world.

Whereas in games, right, with self-play, the computer can play each other and very quickly gather millions of training examples.

You know, the way I think about this topic is if something is very rote, you do the same thing every time – like collecting tolls or operating elevators, those things are not great jobs, right? I mean, is that really – you know, hopefully we can find people better jobs than those. And those are easy for the machine to do.

Then, the next level is making binary distinctions, right, or categorizing things. For example, take email, is it spam or is it not spam? When you have a huge number of emails – literally billions – you can train up a really good model that says yes or no.

But when it’s much more complicated, like how do I design a good podcast? How do I be the CTO of a major company? I think, Kevin, you’ve got job security. We’re not going to be replacing you with a program anytime soon.

**KEVIN SCOTT:** I don’t know, maybe that would be nice. (Laughter.)

**OREN ETZIONI:** Well, what would be nice is to have a program that helps you, right?

**KEVIN SCOTT:** Yeah.

**OREN ETZIONI:** So, a lot of the work that we do and a lot of the work elsewhere really can think of AI as “augmented intelligence.” Would you be more effective at your job if you had a program that helped, I don’t know, prioritize your emails, right? That helps you author emails more efficiently or even papers, right, all that.

**KEVIN SCOTT:** Yeah, so let’s go back a minute. So, you found the TRS-80 when you were a kid, you read, you know, *Godel, Escher, Bach* like was that when you were in high school or in college?

**OREN ETZIONI:** In high school, yeah.

**KEVIN SCOTT:** Yeah, so as a high school student you read this, like, really formative book. And then did you major in computer science when you went to university?

**OREN ETZIONI:** Well, actually, it was – I’ll share with you a quick little anecdote. So, when I went to college, I went to what I like to call a small community college in Cambridge, Massachusetts, but it has the august name of Harvard. They didn’t have a computer science major. They had applied math. And they – their approach, I heard this from some of the professors there, was viewing computer science as more of a kind of applied discipline. They said, “Hey, we don’t have a major in automotive science, so why would we have a major in computer science?”

And right around the time I was there, so this is 1992, 1993, right at the beginning of my time there, they realized, no, this isn’t a fad, and this isn’t some applied discipline, it’s quite transformative. Again, I wasn’t privy to those discussions as a freshman. All I know is they did create a new concentration, as it’s called. And being an eager beaver, I ran to Howard Lewis to declare my new concentration and he signed my paper form, if you can believe that. And he raised his head and he said, “Hey, you’re the first.”

So, I feel like my – I’ll go down in –

**KEVIN SCOTT:** Wow.

**OREN ETZIONI:** -in history as the first person to major in computer science at Harvard.

**KEVIN SCOTT:** That is so cool. That’s really, really neat. And so, what was that program like? Because I’m guessing, you know, given that it was early days, and this was true even when I was a freshman in 1990, and like even in 1990, we were still trying to figure out, like, what a computer science curriculum looked like and, you know, like the entire, you know, ferment of the field was just sort of being developed. So, what was it like being that first student?

**OREN ETZIONI:** Well, so, again, they’d been teaching computer science courses for a while, it was just under applied math. And I would say that the Harvard curriculum was very mathematical, so it was influenced by people like the great Michael Rabin, Turing Award winner, inventor of some of the key theorems in theoretical computer science. Les Valiant was there, you know, teaching combinatorics. So, it was a very theoretical curriculum, which I actually really appreciated because, you know, you want to get that mathematical grounding right.

But at the same time, there was very little AI. So, I was fortunate that, you know, two T stops away was MIT and tech square, and so very quickly I started hanging out at MIT, where Minsky was there and actually Hofstadter, who was, you know, a god for me at the time, was visiting for a year or two. And so, I felt like I was very fortunate because I got my AI at the MIT AI lab and I got my computer science grounding and broader education at Harvard. So, that was really a wonderful time.

**KEVIN SCOTT:** Yeah, and then you fell in love, I’m asking, not asserting. You fell in love with, you know, with computer science – with the field, and decided to go to graduate school and get a PhD?

**OREN ETZIONI:** So, again, what I was really in love with is the fundamental questions. For me, computer science and computers was always a tool. And, actually, I was also studying cognitive science and philosophy of language science at the time, and I was debating whether I should go to grad school in philosophy or in computer science, because these were methodologies for me to get at these fundamental questions.

And then I talked to some people and I realized the people were graduating with a PhD in philosophy and working in moving, right, it was not a great career path. Whereas I could tackle these issues in a more empirical way and, frankly, just a more satisfying from a career point of view way, as a computer scientist. So, I decided to go to grad school in computer science. But it wasn’t obvious. I could have ended up being a philosopher.

**KEVIN SCOTT:** And what was your dissertation on?

**OREN ETZIONI:** So, I went on to study with Tom Mitchell, who’s really the father of machine learning or one of them in many ways. It was on machine learning, a subfield of machine learning that’s much more symbolic than the kind of neural network deep learning stuff that we do today.

**KEVIN SCOTT:** Yeah, but he must have even been thinking about statistical methods back then. I’ve got, like, one of Tom’s textbooks that I’ve had for a very long time. And I think it’s the first place that I ever read about Bayesian inference. So, I don’t know whether he was thinking about things like the statistical approach to machine learning back when you were his student or not, but he certainly has done some interesting work there.

**OREN ETZIONI:** Absolutely. He was and he did, and his classic textbook covered it. Geoff Hinton was visiting at the time, and there was very lively and intense discussion of his ideas. I think the thing that Tom and certainly I missed was the power, the potential power of neural networks to – when you have awesome amounts of data and tremendous amounts of computational power. So, at the time, they didn’t necessarily do better than other statistical mechanisms. So, it was very clear that we wanted statistical methods, but it was a lot less clear that we wanted neural networks.

**KEVIN SCOTT:** I think that’s one of the sort of classical timing problems in computer science and engineering, where it was a good idea, but because – we effectively didn’t have the data volume and the distributed computing infrastructure that came along with the internet revolution and like these big internet companies and their distributed computing and cloud infrastructure, which wasn’t called cloud infrastructure at the time.

And so, I just sort of wonder what are we missing right now? Like, the good ideas that happened 10 years ago that are, you know, left by the wayside that are now feasible and like we don’t even have someone – I mean, the good thing about Geoff is he was stubborn, like, he was convinced that it was a good idea and he kept pushing on it, you know, for a very long time until the conditions were right for it to be successful.

And I just sort of wonder, like, you know, what conditions are going to change in the future that are going to make you know some of the things that are infeasible now actually possible? It’s one of the reasons why I’m like really excited about all of this large-scale compute infrastructure that we’re building right now, because like it’s just – it may not get us to AGI, which is a thing I do want to talk to you about in a minute, but it certainly gets us something, and I’m really interested to see what that something is.

**OREN ETZIONI:** Me too. I think it’s a really interesting time to be a computer scientist, to be a computer professional. I do want to say, off the top of my head, here are three things that the current technology doesn’t yet touch. The first one is the current technology – maybe this is a good phrase – is kind of profligate in its use of compute and data. Yeah, I need millions of examples at least for pre-training and then thousands for tuning. Yeah, I need this massive amount of computation, millions of dollars of computation to build my model.

Whereas of course human intelligence, which is the standard, sits in this little box, right, that’s on top of my neck and is powered by the occasional salad and a cup of coffee, right? We know, right, you know, kids, they’ll see one example and they’ll go to the races. So, I think we can build far more frugal machines in terms of data and compute, that’s one.

And then the second thing, and this goes right back to the discussions we were having at CMU in the early ‘90s is, “What is the cognitive architecture?” In other words, okay, you can take a narrow question like, “Is this email spam or not,” or “Did I just say “B” or “P?” Speech - phoneme recognition. And you can train models that’ll do – they have super-human performance at that.

But the key thing in artificial general intelligence – in AGI – is the “G.” So, how do we build what was called, then, a unified cognitive architecture? How do we build something that can really move fluidly from one task to another, when you form a goal, automatically go and say, “Okay, here’s a subgoal, here’s something I need to do or learn in order to achieve my goal.” There’s just so much more to general intelligence than these savant-like tasks that AI is performing today.

The third topic in AI that I think we ought to be paying more attention to is the notion of a unified cognitive architecture. So, this is something we studied at CMU back in the day. And it’s the notion of not just being a savant, not just taking one narrow problem, but going from one problem to the next and being able to fluidly manage living, where right now we’re talking. Soon, I will be crossing the street, then I’ll be reading something.

Putting all those pieces together and doing it in a reasonable way is something that’s way beyond the capabilities of AI today.

**KEVIN SCOTT**: Yeah, and we’ve got a little bit of that starting –

**OREN ETZIONI:** I know, I –

**KEVIN SCOTT**: In transfer learning, like, but just beginning.

**OREN ETZIONI**: Right, but the thing about the transfer learning is that it’s still from one narrow task to another. Maybe it’s from one genre of text to another genre of text. We don’t really have transfer learning from, okay, I’m reading a book, to now I can take what I read in the book and apply it to my basketball game, right? We’re very far from anything like that.

**KEVIN SCOTT:** Yeah, the other thing that I’m also really curious about, we’ve – you know, we’ve chatted with some people on the podcast who are doing research on insect biomechanics, for instance. You know, which – and they’ve done very, very detailed studies of, like, what the neural architecture is for the control systems that manage insect flight and navigation, for instance.

And they are very, very highly specialized neural circuits. It’s not a, you know, sort of a general like deep network that you know – like maybe a deep network could learn that behavior, but like there’s certainly – there seems to be an efficiency opportunity.

And like I could not more strongly agree with that point. I forget what the numbers are, but if you look at like the world champion Go player, Lee Sedol, like, he probably got to world-champion levels of expertise on a small number like, low tens of thousands of hours’ worth of game play, which is like more than you or I would ever be willing to commit, but far, far less than the millions of simulated hours that a thing like AlphaGo Zero invested to get to the point where it could play competitively with him.

And that is – you know, that amounts to millions of dollars and millions of watts of power and it’s a very, very interesting chasm that I think we have an opportunity to cover at some point – hopefully in the not-too-distant future.

**OREN ETZIONI:** Absolutely. You know, at the Allen Institute, we’ve recently written a paper that’s going to appear in Communications of the ACM on something we call “green AI.” And the idea of it is both to think about the carbon footprint, right, and the “Can we build these more efficient systems?” But there’s another thread here that I want to highlight, which is making sure that the research that we’re doing is sufficiently inclusive.

So, back in the day, it used to be that you or I or a talented undergraduate in India or some other country with a laptop could do something really cool and write a paper about it and get noticed. If we reach the point where you have to have so much infrastructure and so much compute to do an experiment that leads to a published paper, that’s a real problem for the field, right? We don’t get to harness all the brilliant ideas and creativity of a broader population. And so we suggested some pragmatic ideas of how to fix that not by, you know, forbidding or cutting off this very exciting, high-end research, but by saying, “Okay, let’s also look at efficiency, at results of, okay, how can we run these types of models on a much more limited device?”

**KEVIN SCOTT:** But I’m so glad to hear that you all have written that paper and are pushing on that because I do think it’s one of the fundamental issues that we’ve got at this particular moment in time with AI research, these models are not only extremely expensive to train, they are extremely expensive to serve. So, you’ve got this cost thing that makes it difficult to make them widely available.

I mean, like, we could – I’ll give you an example. I won’t talk about GPT-3, but I’ll talk about this model that we built called Turing NLG, which is a 17-billion parameter transformer model.

**OREN ETZIONI:** Only 17 billion.

**KEVIN SCOTT:** Yeah, only 17 billion. So, like, it was extremely – I mean, we trained this on a very large, very sophisticated cluster of GPUs and it consumed a lot of resources and like we wrote a bunch of very specialized software to manage the distributed training task.

And the model is very, very powerful. And so, one of the things that we’re struggling with right now is - I would love to get that model into the hands of as many people as humanly possible. One impediment to getting it into the hands of as many people as possible is cost. And like I think, you know, to your point, we can bring the costs down by, you know, making a whole bunch of these investments. Like, the infrastructure could be better. You can distill the model, there’s all sorts of like really interesting things that you can do to like still preserve the model’s power and make it much cheaper to serve.

But the other, you know, interesting thing with these models is it’s a general language model. It will enable people to put it in use cases that we would find objectionable. And like by “we” I don’t mean Microsoft, I mean “we” society. And so you know, it’s – and I don’t know how you train the model to allow it to do all of the powerful things that it can do and exclude the objectionable things that we’re not going to want it to do.

And so that is another thing that makes access a little bit tricky. So, like, how do you get that into the hands of responsible people so that they can discover all of the good uses that the tech companies that have the resources to build these models will never be able to imagine on their own without, you know, opening Pandora’s box and creating more misery in the world.

**OREN ETZIONI:** Very important questions. The good news is, I do think that we’re making progress there. So, some of it is the old adage, garbage in, garbage out. So, you have to be careful what you feed. This model is kind of like an innocent child, right, who will read anything. So, you have to be careful what you feed it.

That’s typically not enough, because these things consume, right, you know, billions and billions of sentences and documents. I’m a great believer in auditing techniques. So, we also need to make models like this eternally auditable so other bodies can help discover if there’s, you know, problems hidden there or if it can be tuned in a negative direction. So, I do think that you and Microsoft are very smart to think carefully about these issues, but I do think that help is on the way.

**KEVIN SCOTT:** Yeah, well, and that I think is one of the foundational things I would like to be able to figure out sooner rather than later is just a way to allow the help to happen in an efficient and transparent and open manner, because at least that much needs to be happening.

It would have very ironic to have a situation where the very, very necessary public and open work that needs to happen on responsibility and safety and ethics and all of these other things can’t happen because the people who are doing that work outside of corporations don’t have access to the models.

**OREN ETZIONI:** Well, so then if I may ask you a question, right, with GPT-3 and Microsoft’s recent exclusive ability to license it, are you planning to make it available to academia, to places like AI2?

**KEVIN SCOTT:** Yeah, we’re trying to figure out exactly how to do that right now. Like, in a safe way and – I mean, the thing with GPT-3 is, like, it’s, you know, it’s ten times bigger even than the Turing NLG model that we built.

And so there, we probably will have to serve it on our infrastructure just because the task of figuring out how to – like, if I gave you a bag of weights or we gave you a bag of weights and then you had to go figure out how to serve it for doing your research, like, that would be its own research project that would be daunting. And so, like, we are working through those issues right now.

**OREN ETZIONI:** Well, I’m glad to hear it. And, of course, you’re absolutely right. With something that big or even the Turing NLG model, right, what people really want is an API, not a file that you download with lots of numbers in it.

**KEVIN SCOTT:** Yeah, and so the nice thing about an API is it lets you, through like the access control to the API, like, have some sort of, you know, guarantees around responsible use. And then it just – it really does become about cost.

[32:29]

And, you know, cost is – we want to be able to get the cost down to the point where you can do reasonable amounts of inference and exploration with the model in a way that doesn’t break my personal budget at Microsoft. (Laughter.)

**OREN ETZIONI:** Sure, which I’m sure is quite sizable. But here’s a simple idea that we advocated in the Green AI paper. The reporting standards, when we report accuracy of models, you know, how often they get things right, are very clear and so on. But the reporting standards on how much did it cost you to produce that performance, aren’t. People often don’t report that, or if they do, they don’t report some of the dead ends that they went into.

If they reported that more rigorously, then as you said, the distillation efforts – basically, the efforts to build a cheaper model, maybe a far-cheaper one, would be a lot easier. Because I could say, okay, here’s your model, it costs, let’s say, $100,000 to train it. Here’s my model. It only performs 70% at the level of your model, but you know what? It costs 1/10th to produce.

Or here’s my model, maybe it’s only 40% of your model, but I can run it on a phone. So, the – but I can only publish that result if I have a baseline, right? And that baseline has to include the cost, because then I can make that.

So, a really simple step on the part of everybody – kind of the rich players in the ecosystem of simply being rigorous on reporting their costs would enable everybody else to start whole new – really sub areas of the field.

**KEVIN SCOTT:** Yeah, I think that’s an interesting idea and, certainly, something we will think about. I mean, one of the interesting things that I’ve seen in – so, my Microsoft Research reports to me at Microsoft. And, you know, one of the things that we’ve struggled with this same thing even internally, right? So, like, you may imagine that, you know, Microsoft is you know like has a very big capital and R&D budget, which means that you know, once we have one of these models, then everyone can use it, which isn’t even true internally.

So, you know, the amount of resources required to train a very big model and then to, like, go serve it in an application where you may have a billion users is –it’s sort of daunting on the training side, there’s just limited amount of compute.

[34:15]

And so, like, you have to decide who is going to be training what. And on the serving side, we really do very carefully measure how much it costs to, you know, in terms of compute and power and depreciation on the capital for a single API call to a thing, because we have to make sure that when we’re putting it into a product or a service that you know, you’re profitable still.

And, so, you know, our teams inside of the company even have these issues about, like, you know, what engineering work do I have to do to make this thing useful for me? Because we haven’t even sorted the problems of access out inside of the company. And I know for a fact, you know, that everyone is struggling with this right now because these models are so big.

**OREN ETZIONI:** That makes a lot of sense, yeah.

**KEVIN SCOTT:** Yeah, so, you started your career – or early parts of your career, you were doing, you know, a bunch of super-interesting stuff in information retrieval and you – you know, you were starting companies, you – like I think may have built the first comparison shopping site on the Web.

How did you – it sounds like you’ve always been interested in artificial intelligence, and you know, my question is: Like, have you – you know, were you always able to see, you know, this through line through all of this stuff that you were doing, like, everything always connects back to AI? Or was it, like, all right, these are just super interesting problems of the moment that I have the skill you know to address and the interest in.

**OREN ETZIONI:** Anybody who tells you that over a long career and here I’m betraying my age, I’ve been doing this for 30-plus years, there’s a connecting through line and a grand plan is either far smarter than I am, or far more strategic, or is just kind of selling you a bill of goods, you know, the Brooklyn Bridge, as they say. There are themes that I’ve always been fascinated by.

For example, all the companies that have started have always been about empowering the consumer with information – particularly about price, like when to buy your airline ticket, right, that was so you get the best price. That was Farecast, or the first company I co-founded with Professor Dan Weld on online comparison shopping, “Where do you find the best price?” particularly this was in the world before Amazon.

Amazon was just getting started. So, and it wasn’t just, by the way, about price, it was also about selection, right? Nowadays, we kind of assume we can always find it at a decent price at Amazon, say, but A, sometimes you can get a better price elsewhere, and B, at the time, often it wasn’t clear where to buy a particular good.

So that was a big theme for me. And the toolkit, we all have our toolkit. Certainly, AI and information retrieval related to that, machine learning, that was my go-to toolkit, so it was natural to go there.

But I kind of view a lot of these as side adventures. I feel like my two long-term passions are – one is the fundamental intellectual question of AI that we talked about earlier, “How do we build an intelligent machine?” And then the second one is, “How do we use AI as a technology to make the world a better place?”

And, you know, better search engine can do that, better shopping can do that. Right now, at the Allen Institute of AI – or AI2, as I call it – a project that I’m particularly proud of is semantic scholar. How do we make scientific research more efficient and more productive using AI? And actually a callout to Microsoft Research and the work on the Microsoft Academic Graph. Right? We’ve made extensive use of those resources to build Semantic Scholar.

**KEVIN SCOTT:** That is awesome. So, I want to dig a little bit into the particulars of these big models that folks are building right now, particularly around language, although soon they’re going to be multimodal and you know sort of applied to a bunch of different domains.

But one of the interesting things that you all did. I think this was sometime the middle of last year was this system called Arista that is your test-taking system. And so you’ve been working on this for a really long time, and my understanding is that you were able to leverage some of these newer, self-supervised language models like Google’s BERT model in particular, to finally get the system to the point where it can reach parity with students at taking science tests. Talk a little bit about that.

**OREN ETZIONI:** Sure, so this really starts with the late Paul Allen’s vision. And he asked the question, “Gosh, if our technology is so great, be it computer science or AI, why can’t it pick up a book – a textbook – read the textbook and then answer the questions at the back of the book?”

And he had a project even before Allen AI, which started in 2014, he had an earlier project that attempted to do that in various ways and were not successful.

So, when we launched Allen AI, AI2, in 2014, we said, “Why don’t we focus on grade school tests?” Let’s work our way up to reading a college-level biology textbook.

And the way we do that is we’re going to have the program take a fourth-grade science test, and eighth-grade test, take the SATs. And the beautiful thing about that is we can measure the system’s performance and we’ll take it on unseen tests, right, the same way, you know, the Regents New York State publishes a new test every year.

So, we’ll take an unseen test, the machine has never seen the test before. And like a kid, we’ll measure the performance. We’ll also compare it to human performance. We’ll have a benchmark where we say, “Okay, how is our progress going?”

And I think that helped Paul Allen, reassure him that he could get a sense, right, without being there every minute, get a sense of our progress and that we were really making progress rather than building towers in the air. So that became our benchmark is these science tests. And we struggled mightily with these tests in part because it turns out the tests are written in English and they require a lot of background knowledge. They require understanding of various phrases. They have phrases in there like “the onset of winter.” What the heck is the onset of winter? So, a lot of tricky issues that we don’t have time to get into.

We were making steady progress, but it was hard. It was very difficult. We set ambitious goals every year and we struggled mightily to meet them. What happened, as you said, this new class of models came, and they actually originated at AI2. We had a model called ELMO, won the best paper award in 2018. It was quickly followed by Google’s model called BERT, which is a nod to ELMO, won the best paper award in 2019.

And people have gone from there, and Turing NLG, which you mentioned, right, is yet another instantiation. Others have Roberta, et cetera. There are all these models are actually – I should mention because they’re actually pretty simple. All they do, right, is take a word and say, “We’re going to figure out what this word means based on its context.” Not its context in one page or one sentence, but in all the places it appears in billions of sentences, we’re going to compute statistics on its context and these statistics are going to enable us to predict if I see a word or a sentence or a phrase, what’s going to come next – what might even come before – just based on what typically happens.

It turns out that that basic idea, with a lot of technical bells and whistles, is incredibly powerful – more powerful than I think anyone would have anticipated. So, we started using that in our work and we said, “Okay, it’ll help. This tide will boost all boats.” But we never anticipated how much it will help.

So, we found that very quickly, we were able to get 90% – at least on the multiple-choice parts of the test – that involved text. And so, all of a sudden, this model led us to pass and even ace a fourth-grade test, eighth-grade test, even 12th-grade science tests. And that was a surprise. And it leads me to make a prediction. I think that we’re going to see – we’re already seeing, but we’re going to see even more in the next five years, tremendous applications of natural language all over the world.

There’s machine translation, which we’ve already seen. There’s – in the medical and the healthcare system, everywhere where there’s text, which is kind of everywhere, right, because we have text in our emails, we have text in physician records, we have text in scientific papers, we have text in insurance claims, you know, you name an arena of life, I’ll tell you the text that’s there.

Our ability to understand that has really had an inflection point. And that inflection point is going to result in both improved science, but also new startups, new capabilities out of companies like Microsoft and Google. It’s really an exciting time to work on what’s called NLP.

**KEVIN SCOTT:** Yeah, it is. I mean, we had a similar phenomenon, I guess, with convolutional neural networks, which like really were a step function improvement in image recognition and like some of these visual domain tasks. But you know, if anything, because so much of our world is about written human language, like, the impact that I’m seeing from these models like ELMO, I mean, it’s just unbelievable.

Like, I completely agree with you about the far-ranging impact and, like, probably we’re going to see an acceleration over the next five years of, you know, companies and applications and all sorts of interesting things.

I want to pick your brain, though, as we reach these inflection points on domains where AI starts to get really good it always comes with implications for what does it mean for you know individual humans and greater society.

Like, one of the things I’ve been really, really asserting to my kids’ teachers are that, you know, we have these systems that already are pretty good at taking standardized tests, and like, we – you know, we probably don’t want to be training our kids as test-takers, you know, unless there’s some much better understood cognitive benefit of the test-taking activity or you know like it really is a verification that they have you know ingested and learned, like, deeper knowledge than they need just to do the test-taking activity, because the machines are going to be able to take the tests with super-human performance very soon, I would guess.

You know, which means that – you know, I’m sure you’re glad that you went to Harvard and you had, like, a pretty diverse training, like, you were interested in cognitive science and mathematics and got a liberal education. And I think liberal education becomes more and more important as we get expert machines in these narrow domains.

Like, what are some of the other things that you all are thinking about – because this is part of the AI2’s mission, right, is to think about AI’s impact on the world, like, how are you thinking about these things?

**OREN ETZIONI:** Well, I think the topic of education in the modern world is near and dear to my heart. You know, I have three kids and a ten-year-old. And I worry both about the fact that they’re not really getting nearly as much computer literacy as I would like, and literacy with statistics and the ability to analyze data, right, which we have more and more of.

That impacts even their ability to be a good citizen, right? So many of the issues – let’s take climate change, right, that we face, you know, come to statistics or the issues of what is going to be the role of computers and algorithms in society.

So, I really think that we do need to have the basics, right? You don’t want people who, you know, can’t read, because the computer will read to me or can’t do arithmetic, right? So, you want to have the basics, but you want to go way beyond that. And you want to develop the skills of working together with the machine. The machine will do its part and the kid will learn to use it in important ways.

At AI2, we think about a number of issues not so much having to do with kids and education, but certainly having to do with bias. How do we prevent machines from amplifying the bias that’s in their input data, right? Because these models that we talk about, right, they take typically data from the past, they crunch it in the present, and then they make predictions or even decisions in the future.

So, if our past has racism and sexism and other ‘-isms’ that are very unfortunate, or more than unfortunate, they can be horrific, the last thing we want to do is carry them forward to the future.

And, again, that’s a very hot area of research, both at AI2 and more broadly we had a paper a few years ago won the best paper award called *Men Also Like Shopping*, that looked at the bias that’s actually in images, right?

You type in shopping, you’ll see more images of women than men. How does that affect our computer vision systems? Et cetera, et cetera. So, there’s a lot of work there.

I would say that the focus at AI2 has been on the beneficial use cases. So, there’s some work to do to fight against the negative ones, but why do we even go into this in the first place. We go into why do we build this advanced technology? Why do we do this basic research? We do it because we see opportunities for technology to make the world a better place.

And, of course, now when we’re all – the entire international society is questing for a vaccine to COVID-19, I think that’s a very important illustration of that, right? We are reliant on technology, and by the way, AI is heavily used there, but we’re looking for technology to solve some of humanity’s thorniest problems. And we’re working to build what I would call “beneficial AI” systems.

**KEVIN SCOTT:** So, what do you think we collectively can do? And, like, you can interpret “we” however you want – we the technology industry, we academia, we the governments of the world’s nations can be doing to leverage the power that we’re building with AI and to get people prepared?

I mean, you mentioned numeracy, for instance, which is mathematical equivalent of literacy, which I think is at least as important as literacy in the modern world or, like, 21st century.

But, like, what are the other things, like, policy-wise, education-wise, investment-wise that we should be doing to, like, receive the benefit that AI is going to be able to create over the next handful of years?

**OREN ETZIONI:** Well, in terms of policy, I think we do actually have to be very careful not to use the kind of blunt and slow and easily distorted instrument of regulation to harm the field. So, I would be very hesitant, for example, to regulate basic research. And I would, instead, look at specific applications and ask, “Okay, if we’re putting AI into vehicles, how do we make sure that it’s safe for people? Or if we put AI into toys, how do we make sure that’s appropriate for our kids, for example? The AI doesn’t elicit confidential information from our kids or manipulate them in various ways.”

So, I’m a big believer in regulate the applications of AI, not the field on its own. I think some of the overarching regulatory ideas, for example, in the EU, there’s the right to an explanation. And it sounds good, right? AI is opaque, it’s confusing, these are called black box models. Surely, if an AI system gives us a conclusion, we have a right to an explanation, that sounds very appealing.

Actually, I think it’s a lot trickier than that because there are really two kinds of explanations of AI models. One is explanations that are simple and understandable but turn out not to be accurate. They’re not high-fidelity explanations, because the system is complex. And a great example of that is if you go to Netflix and it recommends a movie to you, they’ve realized that people want to know, why did you recommend this movie? And say, “Well, we recommended this movie because you liked that movie, right? We recommended Goodfellas because you liked The Godfather.”

Well, if you look under the hood, right, the model that they use is actually a lot more complicated than that. So, they gave me a really simple explanation that’s just not true. So, that’s one kind.

The other kind is I can give you a true explanation, but it’ll be completely incomprehensible. So, now if the EU says, you know, you have a right to an explanation, what you’re going to end up with is one of these two horns of the dilemma – something that’s incomprehensible, or something that is inaccurate.

So, I think that it’s really important that we are careful not to go with kind of popular notions like right to explain, but instead, think through what happens in particular contexts.

**KEVIN SCOTT:** Yeah, I think that is an extraordinarily good point. These models are already at the complexity where they’re as complex as some natural phenomenon. We’re not able to explain many natural phenomena because, you know, when we get down to the point of like these are the electrostatic interactions of atoms that comprise this system. You have to look at the phenomenology of the system. It’s why statistics is going to be such a really important skill for everyone. It’s why understanding the scientific method and having an experimental mindset I think is important.

I think this is such a good point about not deceiving ourselves that an incomprehensibly complex answer to a question of like “Why did this thing do what it did?” even if it’s couched in terms of language that we might otherwise understand, that’s not real understanding.

**OREN ETZIONI:** Exactly. And I’m not suggesting that the solution is, hey, just trust us, you know, we’re – we’re all (inaudible, crosstalk)

**KEVIN SCOTT:** Yeah, yeah, yeah, for sure –

**OREN ETZIONI:** – going to work. But, again, going back to the auditing idea, rather than an explanation, if we want – you know, one of the most jarring ones are uses of AI in the criminal justice system, right?

**KEVIN SCOTT:** Yes.

**OREN ETZIONI:** To help make parole decisions and things like that. Well, we should audit these systems, test them for bias, right? The press should be doing that, the ACLU should be doing that, regulatory agencies should be doing that. But the solution is not to get some strange explanation for the machine. The solution is to be able to audit its behavior statistically and test it, hey, are you exhibiting some kind of demographic bias?

**KEVIN SCOTT:** Yeah, I mean, one of the things we do at Microsoft is we have these two bodies inside of the company, this thing called the Office for Responsible AI that sits in our legal team. And we have this thing called Aether, that’s the AI and ethics committee inside of the company.

What we do with both of these bodies is we try to have both the lawyers and the scientists thinking about how you inspect both the artifacts that you’re building in your AI research, but their uses. And we have a very clearly defined notion of a sensitive use. And depending on how sensitive a use a particular model is being deployed in, we have different standards of auditing and scrutiny that go along with it.

And, recommendations, like, for a criminal justice application, for instance, you may say that a model can only advise, like, we do not condone it making a final decision. You know, just so that there’s always human review in the loop.

**OREN ETZIONI:** I think that’s smart. And I also think that this relates to another key principle when we think about both regulatory frameworks and ethical issues. Whose responsibility is it? The responsibility and liability has to ultimately rest with a person. You can’t say, “Hey, you know, look, my car ran you over, it’s an AI car, I don’t know what it did, it’s not my fault, right?” You as the driver or maybe it’s the manufacturer if there’s some malfunction, but people have to be responsible for the behavior of the machines.

The same way that, look, I’ve got – the car’s already a complex machine with 150 CPUs and so on, I can’t say, “Oh, well, the car ran you over, I had very little to do with it.” The same is true when I have an AI system. I have to be the one who’s responsible for an ethical decision. So, very much agree with you there.

**KEVIN SCOTT:** Yeah, so we are just about out of time, but before we go, I wanted to ask you, what do you do for fun outside of work?

**OREN ETZIONI:** Well, I would say that spending time with people, you know, my family. Like everybody in the Northwest, you know, getting outdoors and hiking is fun. I also – I love team sports. My knees no longer tolerate it, but for many years, I played soccer and basketball.

So, you know, do a variety of those things. But most recently, particularly under COVID, I have to admit that I’ve developed a vice. And that is playing Bug House online. So, Bug House is team chess, right, where you have two players facing off two other players. And you have – used to be five minutes, now online three minutes is popular.

So, you have a three-minute game. And in that game, you have to defeat your opponent, working together with your partners. That is, if your partner is the white player and you’re the black player, she might hand you a black piece, right, because she takes her opponent’s black piece, she hands it to you. This is all mediated by the computer. And you place it on your board instead of making a move. So, it’s a crazy, crazy game, which is what the phrase Bug House alludes to. And it has communication, it has adrenaline, it has stopwatch, you know, split-second timing, and it’s a wonderful distraction from worrying about when we’re going to have a vaccine.

**KEVIN SCOTT:** That sounds like a lot of fun.

This has been an amazing conversation. I’m so happy to know that there are people like you – and institutions like AI2 that are not just advancing the state of the art, but like, thinking very, very carefully about how these technologies can have a positive benefit for society. So, like, thank you so much for the work you do and for being here with us for an hour today.

**OREN ETZIONI:** Well, thank you, Kevin. I really appreciated the opportunity to talk to you and to talk to your audience. It’s a real pleasure.

**KEVIN SCOTT:** Awesome.

[MUSIC]

**CHRISTINA WARREN:** Well, that was Kevin’s conversation with Oren Etzioni. Wow, you went through so many different paths and talked about so many different interesting things.

For you, Kevin, I’m curious, as someone who works on the industry side of artificial intelligence, what excites you or I guess is most interesting to you about some of the stuff that Oren is pursuing from more of the academic perspective?

**KEVIN SCOTT:** Well, look, I think AI2 is doing some of the most interesting work in the field right now, and you sort of heard in our conversation, this giant leap forward that we have had in natural language processing and natural language understanding over the past handful of years.

Started at AI2 with their work on this technology called ELMO, which then resulted in BURT and Roberta and Turing NLG and GPT and a whole bunch of these technologies that really are reshaping how natural language and AI is working right now.

One of the things that Oren was chatting about that’s really impressive both from a technical perspective as well as just beneficial use for society is the work that they’re doing on Semantic Scholar, which is applying some of these really advanced AI technologies to the task of trying to understand and make more accessible the huge amount of scientific literature that the researchers of the world are producing right now.

And things like that become even more important when you have a moment like we’re in now with COVID, where getting that research digested and to the right people as quickly as humanly possible so that you really are able to get those right people to understand the salient points can just mean the difference between literal life and death as we’re scrambling to develop therapies and vaccines for SARS coronavirus-2.

So, it’s just wonderful that you have this combination of such incredibly clever people there who are also working on these problems that can create such benefit in the world.

**CHRISTINA WARREN:** Yeah, no, I totally agree. And when he was talking about some of that work, I was thinking just about all the – the use cases and – and you, obviously, go to some of the most important ones, which would be, you know, life-and-death decisions. But just in general, the ability to really help get people the right information by you know translating and kind of digesting and getting the essence of those documents out. It’s really powerful.

And that’s just talking about, you know, in one language. I think about what could happen when we talk about, you know, like, machine translation and that sort of work, too. I’m excited about what GPT-3 and other things that I – that you know about that I read about and try to kind of understand promise and it’s really exciting to think about.

**KEVIN SCOTT:** Yeah, I think in general, it’s an exciting time to be working on AI right now. And just a good reminder from this conversation today that it is equally important to be thinking about how you point that work in a direction that is safe and responsible and benefits the public good.

**CHRISTINA WARREN:** What Oren is doing, he is still a professor. He works with students and he’s training the next generation of people who – whether they’re going to be researchers or people working in industry are going to be solving these problems and inspiring those next generation of thinkers and innovators is awesome, and I’m glad we have people like him doing that kind of work.

**KEVIN SCOTT:** Yeah, I mean, as I wrote about in my book and I’ve talked about over and over and over again, we should be thinking about AI as a tool, and a tool that can enhance and augment the things that human beings are trying to do. And I think Oren raised a really good point in our conversation today around the important role that education has to play in preparing our future citizens and people even in the workforce today in being able to pick those tools up and make the best possible use of them.

And it’s a little bit different than, you know, the education system that we have right now, which is fundamentally engineered around the needs of an industrial economy. In the future, we may need to train people to better operate inside of an AI economy.

**CHRISTINA WARREN:** Yeah, no, that’s really interesting things to think about. And I also really liked when he was talking about how, you know, he thinks of AI as “augmented” intelligence, and I think that to your point, especially if we’re thinking about having to train people to work inside an AI economy, that’s where that kind of comes into place is being able to augment or in some cases supplement other types of learnings and other things that people are doing so that we can adjust and I guess be agile.

**KEVIN SCOTT:** Yep, absolutely.

**CHRISTINA WARREN:** Okay, well, that’s a wrap. Thank you so much to Oren for joining us today. Also, buy Kevin’s book or read it, because it’s fantastic. Get it from your local library, your favorite local bookstore, wherever. It’s awesome.

And to our listeners, thank you for joining us. Thank you for being in this conversation. Send us a message anytime at behindthetech@Microsoft.com. And tell us what’s on your mind. And please, stay safe out there.

**KEVIN SCOTT:** See you next time.