3 Takeaways Podcast Transcript
Lynn Thoman
(https://www.3takeaways.com/)

Ep. 137: A Noted MIT Dean on the Age of AI and the Rapidly Evolving Relationship Between AI and Humans

INTRO male voice: Welcome to the 3 Takeaways Podcast, which features short, memorable conversations with the world's best thinkers, business leaders, writers, politicians, scientists, and other newsmakers. Each episode ends with the three key takeaways that person has learned over their lives and their careers. And now your host and board member of schools at Harvard, Princeton, and Columbia, Lynn Thoman.

Lynn Thoman: Hi, everyone. It's Lynn Thoman. Welcome to another 3 Takeaways episode. Today, I'm excited to be with Dan Huttenlocher. He's the Dean of MIT's Schwarzman College of Computing. He's also an inventor who holds over 20 patents. And he's the co-author of The Age of AI with Henry Kissinger and former Google CEO Eric Schmidt. For anyone who's interested, Eric Schmidt was a guest on 3 Takeaways in episode number 7. I'm excited to find out how Dan sees the future of AI [artificial intelligence], ranging from upsides we wouldn't have thought possible to perhaps unimaginable downsides. Welcome, Dan, and thanks so much for our conversation today.

Dan Huttenlocher: Thanks so much for having me. I'm delighted to be here with you all.

LT: I am thrilled as well. Dan, are we integrating non-human intelligence into the basic fabric of human activity? And if so, can you give some examples that people might be surprised by?

DH: We absolutely are. I think, as one can see from the news these days, there are all kinds of places where particularly machine learning and artificial intelligence is being used in a manner that really, at least, appears to be human level intelligence. We can have interesting discussions and debates about what human level intelligence really means, but certainly performing tasks we would have thought required human intelligence. And these examples go from everyday things to very specialized ones. So in the medical domain, increasingly AI methods are being used in the radiological domain for early detection of breast cancer, for early detection of lung cancer. And these are advances that really are standing to improve that early detection and therefore really save people's lives. And so really compelling cases to very specialized things in the sciences.

LT: How is today's AI where machines learn on their own, different from previous types of machine learning? Didn't we use to teach machines by, what I'm going to call human wisdom? So for example, in chess, didn't we feed in all of the grand master games so that the computers could learn from those examples as opposed to now where machines are learning on their own just by playing against themselves?

DH: Yes, this is a perfect and fantastic question. These systems, where we constructed them, I would say, to codify human expertise. As a computer scientist, I wouldn't actually call that machine learning. I would call it AI, but not machine learning, because the machines weren't really yet learning. We took a bunch of human expertise. Chess was sort of the really almost kind of prototypical example of that. In the 1980s, there was the IBM Deep Blue, which was the first computer chess program to beat a grand master. And all of that was done by, not only taking lots of

games that had been played by humans and feeding them into a computer, but it also involved a lot of experts in chess and how to actually structure the way this was being created.

DH: In that era, we thought of these things as kind of, often the phrase was used, expert systems, that the experts would help the system be able to exhibit certain human levels of expertise. And that proved to work in some limited cases, but not to apply more broadly to other domains. Chess was one of the limited cases where it actually worked quite well, because we could show them grand master level chess performance. But now with the machine learning algorithms today, there's two big changes from that.

DH: The first is, machine learning algorithms still learn a lot from human behavior, but they often learn from very broad records of human behavior. And then there was a big advance in machine learning methods where you just played the AI chess program against itself. So instead of showing it any human games, you just gave it the rules of chess and you had it play a lot of games against itself. Now, you had to give it some indication, other than just winning and losing, of what were better and worse moves. And then it would sort of improve on that as it played. So there's a little bit of human input about the nature of chess at the very beginning, but remarkably little.

DH: But the thing is, once the machine's playing against itself, it can play billions and billions of games, many more than have been played in the history of certainly recorded human chess, the games that get published. And that ability to learn from that self-play and the scale of that self-play has really pushed not just chess, but many, many other games to the level where the machines outperform humans almost always.

LT: How is the new AI, such as ChatGPT, different, for example, from a Google search?

DH: This is a great question, and it's one where I'd love to start by disabusing people [laughter] from what's been covered in the press a lot, which is somehow the association of ChatGPT and search, exactly the one that you mentioned. It's out there all over the place. ChatGPT is an example of what we call a large language model. It models the structure of human language very broadly, and those are instances of something that we call generative models or generative AI, so you may hear that phrase. And the thing about generative AI is that it does what it sounds like. It generates things, and it generates things that could look like they were produced by a human. It generates human-like output. And these large language models are phenomenal at writing things that sound like a human wrote them, but they don't have the sort of deeper level of understanding and what I would call grounding, tying it back to actual facts. But in search, you care quite a bit about the grounding in facts. In fact, if you look at a lot of the popular press coverage and so forth, it's people being sort of concerned or disturbed about the fact that ChatGPT is "hallucinating" and making things up. But that's because it understands language and it's trying to synthesize something that sounds human-like and is a sort of eloquent way of expressing something.

DH: These are not being verified with underlying facts in any way in the current models. So there's still a ways to go there. And I think, somehow maybe it's Microsoft getting very involved and Bing Chat alongside the search engine, and so forth. But this technology still has a way to go. And in and of itself, it's not really a search tool as we're used to search tools.

LT: AI solutions, though, can be amazing. Can you give some examples of AI solutions that are different than and better than human ones? If I remember correctly, in your book, The Age of AI,

you talk about two examples that to me were stunning. The AlphaDogfight one and the Google cooling systems.

DH: AlphaDogfight is, pilots are trained in simulators. They spend lots of simulator hours in addition to in a real airplane. And the Defense Department got interested in using machine learning techniques to train an AI to fly these simulators in dogfights. And the simulators are quite realistic these days. And they tie to the underlying physics of the real aircraft and so forth. So the ability to fly the simulator well maps extremely naturally onto the ability to fly an actual airplane well. Which should make you feel good every time you get in a commercial airliner, because those pilots are also doing a lot of very accurate simulation flying as well as their commercial flying. And they were able to take relatively current machine learning techniques in a fairly straightforward way, train them in dogfights and get performance that was at the level of the very best human pilots, which was quite striking.

DH: I think a number of these examples are striking in that regard. The Google cloud computing data centers findings were also just looking at how to optimize when computers are on and off and at what levels they're running and what the cooling settings are is extremely important given the amount of power going and cooling going into data centers these days. And again, humans try to tune these things. They have lots of dashboards and lots of data, but there are a bunch of things you're trying to optimize at once.

DH: And so it was a place where there's a lot of data and some relatively straightforward criteria, about sort of maintaining temperatures that don't harm the machines and lowering the kilowatt hours that you're using. And so it was a very well suited problem to machine learning. And again, they were able to improve by, I've forgotten the numbers now, but it was some significant percentage, like 20% or something over what human experts were doing in terms of reducing the power usage without negatively impacting the computing at all.

LT: I was fascinated by the AlphaDogfight example that the AI was victorious over the best fighter pilots, but also that they thought of and they executed maneuvers that humans couldn't execute. I mean, the humans would come up behind the enemy plane, whereas the AI would come straight toward it. So I thought that was fascinating that the AI would come up with solutions that humans hadn't thought of and couldn't actually execute.

DH: Yes, and this comes again a lot out of this self-play that we talked about also for the game of chess. One of the really striking things about AlphaZero, which was the first of the self-play chess programs that really made substantive advances, is it discovered various strategies in chess at, approximately 2000-year-old game that many great minds have looked at, that were completely outside of the set of strategies that people had developed. The game of chess has changed as a result of that. Humans have now studied those strategies and are using them.

LT: Why does AI sometimes get things spectacularly wrong? And can you give some examples?

DH: I think there are a lot of places where AI does the unexpected. So I wouldn't classify that as much as spectacularly wrong, because some of the unexpected is actually extremely valuable, because it's things that humans haven't thought about. When we're dealing with machine learning algorithms, one way to think about them is many of us have known a very creative person who has lots of ideas, and some of them are really worth following up on, and some of them should go into

the dustbin very quickly. [laughter] And you can think of machine learning as that kind of a colleague. And it takes human judgment to understand which of these "crazy ideas" are crazy in a good way. They're highly creative. They're going to push understanding and discovery in new ways, and which should just not see the light of day.

DH: And I think that that's a characteristic of these machine learning algorithms that we're not used to dealing with yet. We often tend to immediately start to think of things as human. We often think of our pets as human. You know, this thing's actually synthesizing language and talking to us, this is one of the things that we've been trying to communicate in the work that I've been doing with [Henry] Kissinger and [Eric] Schmidt and things that we've been trying to do at MIT as well, and really helping us understand how do we think about how to interact with AI? What is its role vis-a-vis humans and human discovery and human understanding?

LT: Can you talk about the upsides of AI, such as drug discovery?

DH: I'll talk about one thing just because I'm so excited about it. It's a later set of advances that are happening, which is a more junior colleague of mine here at MIT who straddles chemistry, chemical engineering, and computer science. And he has been looking at the automated discovery of what are called small molecules in chemical engineering and chemistry. And many drugs are small molecules. So the natural way that both he and others had been looking at these machine learning problems is that you understand properties of the molecules. And that's where that work had all been going. But here's the problem with it. It was finding molecules that had the desired properties, but they were much more complicated to actually synthesize in practice, than something that a human chemist or chemical engineer would have come up with. They didn't take into account the challenges of how to actually synthesize these things.

DH: They would just get any arbitrary molecular structure by just putting it together like Lego blocks. What his insight was, "Let's train this AI also on things that are easy to synthesize." And how do you do that? Well, he used basically a chemical catalog of simpler compounds that you use in the lab to do synthesis. And by adding that to the training process, he enabled the AI to not only find these things that have properties where people haven't been able to find reasonable compounds yet, but ones that were vastly simpler to synthesize. It's this kind of work, I think, where machine learning can really drive things.

LT: Achievements which were once presumed to be human, such as writing a song or discovering a medical treatment, can now be generated by AI. Can you talk about that? What do you see as the potential of AI?

DH: So the thing I'm most excited about, about AI broadly, but just this generative AI that's a creativity side, writing and creating images, is really as a tremendous sort of complimentary thinking partner to humans. Very much in the way we were talking about with chess, or with AlphaZero with chess, or with AlphaDogFight. Because these systems learn in a different way than humans learn, they can get to very different places than people. And if that's in a mode as a collaborator or a partner, a sort of thought partner, if there's some interrogatory mode of interaction, they can be amazing at enabling us to do more. How many times have you sat down to write something, I'll speak for myself anyway, and you just can't really get started. So you just toss a few prompts at ChatGPT, and it writes some stuff, and then you can get in this mode of saying, well, that's a pile of garbage, and now this is kind of intriguing. So that's really what I'm most excited

about these, I'm really excited about the partnership, the collaborative potential of these machine learning systems.

LT: AI can use trillions of parameters. As you know, Facebook's newsfeed recommendation algorithm alone is based on over 10 trillion parameters. And the next version of ChatGPT will, at least it's reported, be over 500 times more powerful than the current version. How can we monitor and oversee AI?

DH: I'm a firm believer in understanding the context in which it's being used. And if it's being used as an autonomous decision process, recommendation process, et cetera, without direct human involvement, if it's not being used in this mode as sort of a collaborator partner of a human, then I think it needs to be held to certain standards of behavior, just like we hold humans to certain standards of behavior.

LT: Humans have personalities and desires that are not keyed to other individuals. Do you think people will prefer computer relationships to the messiness of human relationships?

DH: This is a possibility. I mean, these are issues that we really need to look at and understand. And again, I think there are ways in which they can be very positive. Something that could really be willing to spend all day explaining things to a five-year-old could be very good for that five-year-old's development, assuming we know what it's explaining and what its models are and so forth.

LT: What are the three takeaways you'd like to leave the audience with today?

DH: My first takeaway is that we're now living in a world where AI exhibits human-level intelligence, but that human-level intelligence does not mean human. Chatbots like ChatGPT can write as well as an educated person, and there are many other things that are maybe less visible to the general public in scientific discovery and solving business problems, but AI is totally unlike people. It doesn't have motivations. It doesn't have emotions. It doesn't have intuition. It doesn't have curiosity. It doesn't have empathy, characteristics that are central to being human and to thinking like a human.

DH: Takeaway number two is that the differences between human thought and AI can be very powerful for augmenting and complementing both human intelligence and human creativity. AI can serve as a collaborator, a partner, an interlocutor that has a different form of intelligence than any human that you can have those kinds of relationships with. And we don't yet understand the range of applicability of AI to these kinds of roles. You can already see the use of AI assistance in computer programming, and what's sometimes called pair programming, where instead of sitting with another human, you sit with the AI and make suggestions. And you can certainly envision the use of AI in education.

DH: The third takeaway, maybe the most important, is that it's unnatural for people to interact with something intelligent without attributing human characteristics to it. But it is critical that we do not interact with AI as if it were human. We have to be particularly on guard in situations where the human characteristics like intuition, emotion, curiosity, empathy could be particularly important to achieving good outcomes.

LT: Dan, this has been fascinating. Thank you.

DH: Well, thank you so much for having me. Great to be with you.
[music]

OUTRO male voice: If you enjoyed today's episode and would like to receive the show notes or get new fresh weekly episodes, be sure to sign up for our newsletter at https://www.3takeaways.com/ or follow us on Instagram, Twitter, LinkedIn and Facebook. Note that 3Takeaways.com is with the number 3, 3 is not spelled out. See you soon at 3Takeaways.com (https://www.3takeaways.com/)

This transcript was auto-generated. Please forgive any errors.