Podcast Name: *ACM ByteCast*
Episode: Yoshua Bengio

Welcome to the *ACM ByteCast* podcast, a series from the Association for Computing Machinery! The podcast features conversations with researchers, practitioners, and innovators at the intersection of computing research and practice about their experiences, lessons learned, and visions for the future of computing. In this episode, host Rashmi Mohan, the Director of Engineering at Splunk and a member of ACM Practitioners Board, interviews guest Yoshua Bengio. Yoshua is a Full Professor at the Department of Computer Science and Operations Research at the University of Montreal. He is also the Founder and Scientific Director of MILA, Montreal Institute for Learning Algorithms at the Quebec AI Institute. He was part of the trio that won the highest award in computer science and the ACM Turing Award in 2018. He is also a published author and the most cited scientist in Computer Science.

To begin, Yoshua introduces himself and what drew him to this profession. He was always excited about computers in his teenage years and learned how to program them by himself. In graduate school, he was the only one truly interested in this subject and started reading papers that talked about synergy between the brain and machines. He is interested in the gap between deep learning methods and human intelligence, and wants to better understand what can go wrong with AI to mitigate those risks. In graduate school in his AI class, they learned about logic and symbol manipulation using tools and searching to make computers make their goals. They wanted to build intelligent machines that were inspired by how the brain works and this could be broken down into small bits of computation. His supervisor was not interested in this field, and he started going to conferences to network and find others who shared his passion and a similar vision.

Next, Yoshua discusses how intelligence is a huge bag of tricks with pieces of knowledge that are combined to create reason. But how do we get that knowledge or intelligence? Knowledge is acquired by learning procedures and found they could make an intelligent entity able to do tasks. Knowledge is the relationship between symbols, and symbols don’t have any other meaning unless they are related to other symbols. Each word is associated with a word vector, or representations of the word, and we process all of that information to create meaning. Yoshua worked with neuroscientists, and a large fraction of innovations are inspired by how the brain works and conditions. In machine learning, they track how well a learning is doing and the measurements to show success. Yoshua was more interested in the bigger principles that would help them move toward smaller machines and be transformative. He specifically studied attention mechanisms, which were a way to select the specific pieces of information in the layer and to focus on pieces to create the next computation. For example, if you were translating a paper from one language to another, you would keep a pointer to mark the place that you’ve already translated. This completely changed the performance of the systems and translations and drastically changed the quality of the translation.

In addition, Yoshua talks about how AI machines can now pass as humans and why that is significant. AI can now anticipate its performance metric and choose the right words with higher probability. There is still not a good understanding of the relationship between the quantitative mechanics and more qualitative ability that the machine is understanding, and mastering language models. His concern in AI is artificial general intelligence, or AGI, which is a human level intelligence, which means the machine is as competent as a human in most tasks that we can evaluate. He believes we could reach AGI levels in the next few years, but he is worried that AI systems will be smarter than us and escape our control. There needs to be safety protections in AI to make it behave well and to also protect it from being hacked. AI could also influence us through language through psychological manipulation. We need to understand what can go wrong and guarantee that AI will behave well according to our norms and values. There are also biases and discrimination in AI because of misalignment, and we don't know how to build the systems that align well enough with our values.

To prevent this problem, the first thing to do scientifically is find out what can go wrong and finding ways to mitigate those risks. This can include misalignment, misuse potential, hacking of AI, and AI becoming smarter than humans. The second is political in creating protocols and international treaties that are followed globally so that no single person, company, or government can abuse that power. His advice for researching is to bring about global awareness of the risks and then invest in the right research. In closing, he is most excited about AI safety and building algorithms that could give us guarantees about safety with probabilistic guarantees.

Key takeaways:
1:59 - Yoshua Bengio introduces himself and what drew him to this profession.
5:10 - What did you expect to encounter and what surprised you?
6:43 - Were there others involved in this AI interest?
8:25 - Definition of deep learning and its importance in AI.
13:27 - What kind of interdisciplinary collaborations did you have?
15:00 - What milestones were you tracking in your research?
19:22 - AI machines can pass as humans now and why that is significant.
22:06 - What caused this acceleration in AI?
24:20 - Concerns in AI.
32:20 - Biases and discrimination.
34:18 - What should we do?
37:53 - Advice for researchers on raising awareness.
39:51 - What are you most excited about in this field?

Links:
Learn more about Yoshua Bengio.
Learn more about Rashmi Mohan.

Tags:
Technology, IT, computer science, technology development, technology solutions, technology problems, research, generative AI, researching AI, society, Yoshua Bengio, AI concerns, biases, discrimination, misalignment, misuse, AI safety, deep learning