Life 2.0: Holger Hoos on Responsible AI

May 3, 2023 | www.life-20.com | Length: 58:40

SUMMARY KEYWORDS
ai, systems, people, jobs, bit, good, thought, human, risk, happen, society, experts, big, natural instincts, inflection point, limitations, capabilities, research, challenge, interact

SPEAKERS
Holger Hoos, Yan Chow (Host)

Yan Chow 00:09
Welcome to Life 2.0, a podcast about the personal impact of future technologies. I'm your host, Dr. Yan Chow, a physician, a technologist, and an entrepreneur. This podcast explores upcoming innovations and how they will transform daily life for you, your kids, and their kids. Life 2.0 will interview thought leaders who can help us understand what it really means to be human in the 21st century.

Yan Chow 00:39
My guest today is a renowned German-Canadian computer scientist who is a pioneer and leader in the field of artificial intelligence or AI. His research aims to advance the ideas of Human-Centered AI, AI for Good and AI for All. He develops technology that augments, rather than replaces, human intelligence and that helps humans overcome their biases and limitations. He's working to improve the efficiency of AI by increasing performance and reducing resource needs, and to broaden access to cutting edge AI. My guest is one of the originators of automated machine learning or AutoML. His research spans machine learning, automated reasoning, and optimization, leading to innovative real-world applications in empirical algorithmics, bioinformatics and operations research. My guest co-authored the book "Stochastic Local Search: Foundations and Applications." He also has a keen interest in computer music, developing the Salieri music programming language and the Guido music notation. Academic positions and honors include the Alexander von Humboldt Professor of AI at Rhine-Westphalia Technical University in Aachen, Germany. He is also professor of machine learning at Leiden University in the Netherlands, as well as adjunct professor of Computer Science at the University of British Columbia. He was selected as a Fellow in the Association for Computing Machinery, the Association for the Advancement of Artificial Intelligence, European Association for Artificial Intelligence, and he is board chairman at the Confederation of Laboratories of Artificial Intelligence Research in Europe, otherwise known as CLAIRE, which he cofounded to promote European excellence in AI research and innovation.

Yan Chow 02:21
Who is my amazing guest? He is Professor Holger Hoos. Welcome to Life 2.0, Professor Hoos!

Holger Hoos 02:28
Hey, I'm really happy to be here! The person you're speaking of sounds a little bit intimidating, I can hardly believe that should be me. But you know, whatever, I'm really happy to be here and to have a chance to chat with you, and maybe provide a little bit of insight, enlightenment, and hopefully also entertainment, to your listeners.

Yan Chow 02:48
Before we get started, how can people get a hold of you if they want to contact you?

Holger Hoos 02:53
I'm actually reasonably easy to reach on LinkedIn or Twitter, and otherwise, on, you know, the webpage of my group at RWTH Aachen University, there's a contact email address, as well, where it's some of my dedicated staff members help me not to forget about anything that I should be answering. So that's also a good option.

Yan Chow 03:12
In the description of what you've done, all the things you've accomplished, you do mention something called Human-Centered AI, or AI for All. What do you mean by that?

Holger Hoos 03:21
So, they're two slightly different concepts. And actually, Human-Centered AI, you've explained this yourself already quite well, when you so kindly introduced me. It's the idea to study AI systems and algorithms and approaches that do not aim to replace people in their natural abilities, interests and also intelligence, but to rather augment them to make them better at things that perhaps they couldn't do so well, or even better at things they do quite well, but where there is a benefit in doing that even better, like diagnosing rare diseases, for example, right? So, the key concept is to augment rather than to replace human intelligence. And I like to contrast this with what I refer to as All Out AI, which is, you know, you do whatever you can do, whatever is technically possible, you do it because there's probably some benefit in it. And, you know, I understand people who do this because it's intellectually intriguing. As I just related, I myself was driven and still am sometimes driven by the question, you know, what is possible with this technology? But I do think by now we've reached a state of affairs, where one has to be a little bit careful, perhaps, and this being careful prompts one to ask the question, what should this really try to achieve? Should we try to achieve improving on human capabilities across the board? And for me, the answer to this is very clear. We should not, because if we try to do this, first of all, I think there's a certain modicum of danger there. But secondly, also we would replace people in many activities that they enjoy, that they're good at, that gives them meaning in which they don't want to be replaced.

Holger Hoos 05:00
So why would we want to do that? I think Human-Centered AI is a nice concept, because, you know, it gives us a direction that I think is more meaningful, and also more valuable for society and individuals. So, now quickly to the concept of AI for All, which you also asked about. I think AI for All is connected to that. The key idea here is that you basically strive to develop AI techniques and deploy AI systems in such a way that in principle, everybody can benefit rather than just a chosen few. So AI for All is sort of

in contrast to having AI developments driven by and certainly benefiting mostly the shareholders of a few global companies. That is inconsistent with the idea of AI for All. Even though everybody might be able to have access for payment, or for donating their data, for example, to their services, that's still not AI for All because it's not AI that is for the benefit of all. And of course, it's a complicated concept, in a sense that how can we ever do anything that benefits everybody equally? I think the answer is we cannot. But it's a nice principle to strive for, nonetheless, I feel.

Yan Chow 06:07
Let's say you have health insurance, and you pay a certain premium every year. And then you notice your neighbor with similar symptoms as you do, similar conditions, paying a lot less. And so, when you go to the health plan, the insurance company, saying "How did you figure out my premium?" They would say, well, we have this AI-driven algorithm, factoring all the things that you're doing, and so on and so forth. But when you get that explanation, you don't really understand how they came to that decision. And so, AI for All, in a sense, you could think of it as everyone actually has their own AI that they can use as their advocate. Do you ever see a point where that could happen, where everyone gets their own assistant, their own AI to, in a sense, put them on a level playing field?

Holger Hoos 06:53
It's a very interesting idea. That could be rather nice, actually. Of course, one would have to be very sure somehow that that personal agent, this personal assistant, really had one's own well-being and interest first and foremost in mind, and, you know, to certify such a thing strikes me as not so easy to do, especially if it's, you know, maybe just a few companies providing the agents. But I don't think that invalidates your idea, I think it's still a very nice idea. And if we look at how things are currently starting to work, of course, people have AI-ish assistance on their smartphones, for example, right? You could extrapolate that a little bit further. The concern I would have, just thinking a few steps ahead, the present situation with these voice assistants on our devices is that they are effectively controlled by just a few commercial entities. And even if we say that these commercial entities are currently benign, or under, you know, sufficient regulatory oversight, or have, you know, their own mechanisms in place which we trust, even if we were willing to say that... there would still be something a little bit disconcerting about basically one or two factories churning out the agents that do everything on our behalf. That doesn't sound quite like also a good competitive market situation, aside from all sorts of other concerns.

Yan Chow 08:09
I know you have a personal interest in health care. And I of course do, and I think it'd be a good exemplar of what can happen with AI.

Holger Hoos 08:17
There is another even bigger problem with insurance, in particular health insurance, in that there is no doubt in my mind that as a necessary consequence of our current increase in abilities and capabilities of the AI systems, we will also be much better able to predict the true cost that somebody will cause the medical system, the health system, right? And so now I think you can have two very different--and somehow also rational--approaches to what should happen once you're in this position where you can quite accurately predict what somebody might cost society in this way, right? You could say either that's

fine, we still want to be somewhat egalitarian, right? So, if somebody has the bad fortune to be born with some rare genetic condition that causes a lot of treatment costs, they shouldn't have to foot that bill. It's not their fault, right? Society should step up and cover it for them. So, this is sort of the equalizing principle behind insurance. Or you could say, well, finally we can rationalize this in the way that everybody fairly pays for the cost that they cause. And personally, I find the latter much less palatable than the former, but that of course is sort of a question of philosophy and ethics, and one can meaningfully feel differently about this.

Holger Hoos 08:17
I think it totally is, for many reasons, right? And I'm intrigued that you came up with this example. And you've obviously given this careful thought, because I think, especially health insurance is one of these areas where it's very easy to see what could possibly go wrong. You pointed out one of the things that could really cause quite a bit of discontent and grief, and that is, some AI-based system makes decisions about insurance premiums that really affect people's life, and ultimately, not just their livelihood, but also their access to affordable good care, and that system might not be able to explain its decisions. So, in other words, it might be very untransparent, and then we're in the situation that whoever you managed to get hold of, if at all, they would say, well, computer said this is the way so this is the way, and that's totally unsatisfactory, right? I think, in fact, this is a situation and you know, maybe this is sort of the European Canadian in me speaking now, I know that some people in the US see this a little different, and in other countries, probably too, but you know, the countries that I've had the privilege of living in, they always had what some people call a socialized healthcare system where the underlying assumption is that everybody should have access to good healthcare, and it should not be unaffordable for everyone. And even if you don't earn a lot of money, you know, you shouldn't be forced as a result of that to have substandard healthcare. At least that's the theory, right? So, if one believes in this, then of course, it's especially problematic to imagine this computer-set scenario where a lot of inequity is generated in ways that nobody can really understand.

Holger Hoos 11:05
What I would point out, I'm not an expert in this, but you know, I have a bit of an interest--I believe when the very first insurance companies arose in order to insure things like the cargo of ships hundreds of years ago, the idea was not so much that you would insure against your own personal risk so much as you would equalize the playing field. There would be acts of nature that would lead to certain cargo being lost, certain ships never making it, and there was, as far as I know, really a solidarity aspect of insurance, where we say, look, we all pay into the pool, and if one of us is unlucky, you know, that shouldn't ruin them. The rest of us should, in some sense, then step up through this mechanism, through that pool. And I personally think, you know, the calibration in a day and age where we can more and more accurately predict risk, all sorts of risk, and the cost associated with that risk, the calibration of this equalizing principle, the distribution of risk over many shoulders, on the other hand, having people only pay for the cost that they really cause, that needs to be very carefully calibrated. And I think it also needs to be rethought because so far, our predictions weren't good enough to really have that choice to the extent that we will. And so there, AI is a real game changer.

Yan Chow 12:21

It sort of affects the viability of the insurance risk pool concept. And so, when you have individual risk and individualized premiums, that is actually not a technical issue, not even an AI issue; it's a policy and philosophy issue about what if somebody keeps smoking, and they cannot pay the cost of the treatment for their lung cancer? How do you address that? Probably different societies with different resources will address it differently, but I know that there's a whole spectrum between where you just step back and say, that's your fault, you know, we're not going to pay for that--versus we'll pay for anything and everything.

Holger Hoos 12:57
Indeed, and I think there you're 100% right, that of course it becomes a policy and a philosophical issue. But at the same time, you know, the technical affordances that come from AI system, they change the space in which this policy and, you know, moral philosophical decisions can be made. And I quite honestly think those people in society who are trained to think deeply and carefully about these sorts of policy and ethical and philosophical questions--which, by the way, wouldn't be me--I think they need a chance to catch up with current developments. And that is one of the reasons why I personally was pretty happy to sign the Future of Life Institute's letter calling for a deceleration of certain types of AI research. And now of course, we have more and more prominent colleagues doing the same thing. This is really one of the reasons: because we can't be in a situation where even top-notch AI researchers barely understand or don't understand what their systems are capable of, and what their limitations are, and what could go wrong in using them. And even if they could, it would not be right to just leave it to these people. There are other people who are actually better positioned, better trained, to think about the consequences of using technology, for example, in the insurance industry, right, and that need to be on board, they need time to catch up and understand how this changes the game.

Yan Chow 14:21
That is assuming a moral society, of course. So, when you have other nations and other entities pursuing AI for any and all advantage in whatever realm, that is also another challenge that will arise and did arise, for instance, with the onset of the internet, where we started to get viruses. So, very complex issues and very challenging and maybe some of them may not even be solvable by human thought.

Holger Hoos 14:49
I agree with you. I mean, we should definitely talk, if you would like to, a little bit about problems not solvable by human thought, because those are the problems where we might need AI to help us--not to do it for us, but to get us into a position where we can do things. And I believe there are some areas that people are generally not very aware of where this is already happening in a very positive and very impactful way.

Holger Hoos 15:11
But you know, to briefly comment on what you just suggested, you know, let me push it a step further, you were very cautious in the way you worded this, but let me be less cautious. Let me say there is an argument to be made that says, if we decelerate a bit, if we take a more cautious approach, if we focus on AI for All and Human-Centered AI, surely some bad actors out there will not. They will go for all out AI, for full, you know, for all out profit, whatever goes, goes. We cannot let this happen; therefore, we

shouldn't step on the brakes, we should also do these things, right? I personally think philosophically, morally, this is a very problematic way of reasoning, because it essentially says, just because there's somebody else who's doing something that I think is bad, I have reasons to believe it's not the right thing to do. I better do it too, right? This is the kind of arms race that I think humanity should not engage in. And I would point out that there are areas in which as global societies, we mostly have rejected that kind of thinking.

Holger Hoos 16:09
So, one example is human cloning, which I believe is still basically banned, where you could also say, sure, those who pursue it nonetheless might have pretty large advantage, ultimately, right? So, could we really afford to leave it to some bad actors? Or nuclear proliferation, right? So you know, certainly if we go all out and have more advanced research on nuclear weapons, and so on, than we already do, we can hedge against the fact that some bad actors might do that, despite the fact that they shouldn't because under nonproliferation they shouldn't even be doing anything with this at all. But nonetheless, we say no, here we draw a line, right? The same with chemical and biological weapons. So, I do think it's a little bit of a problematic argument that some people bring forward that just because country X will never be reasonable and never, you know, do the right thing, neither should we.

Holger Hoos 16:59
The domain in which I think this is great to illustrate is, of course, climate change and global warming. There this argument is also being made all the time, you know: what use is it if Germany, for instance, tries to be a good global citizen in cutting down certain types of emissions? Much, much bigger countries with bigger populations and less developed that will not do this. But nonetheless, I mean, sometimes you have to do what's right and lead by example and create coalitions of the willing, rather than just being cynical and saying, well, you know, surely somebody else will not abide by this, so why should we? And game-theoretically, I think we can, there are provable cases where if one has this sort of thinking that as long as some actors out there that might ruin it for all, I don't have any moral obligation to play fair or good or do the right thing, then everybody loses for sure. So, you know, anybody who knows game theory will never fall for this fallacy, basically. So, since I know just enough game theory, I'm a little bit more optimistic and I'm saying, well, you know, let's do what's right, and then try to inspire others to do what's right, to work with them patiently, and maybe also with a little bit of pressure, right? Because some of these actors, when it happens at the state level, they are susceptible to pressure, right?

Yan Chow 18:09
And certainly, there may be other kinds of leverage that you can use, and maybe that could be guided by AI to play the game properly. So, what do you see as the major challenges going forward in the development of AI, besides the policy and philosophical guardrails that we need to keep in mind?

Holger Hoos 18:27
So, I think one of the major challenges that's very much connected with that is that current hype and craziness and also excitement, understandable excitement, about the so-called large language models and generative AI. This is very much of a different nature in that the capabilities we have here are exactly NOT those logical reasoning ones that are hard to learn for people, and that they also have a

hard time relating to unless this is what they do in life, right? So, and since very few people are Sherlock Holmes or similar characters, very few people can relate to that. But we can relate to, you know, AI systems writing texts, summarizing articles, helping us doing background research, polishing our language, helping us in creative endeavors like, you know, making animations or poetry or music--THAT we can all relate to, basically, and that is what these new types of AI systems that are now in vogue are doing. But if you use them for anything where factual correctness or thoroughness or logical consistency matters, like, for example, in journalism or in science or in medicine, then you very quickly realize their limitations, and the limitations are, to a large part, that they cannot reason very well. Their reasoning capabilities are very, very limited, and instead, they tend to hallucinate. They tend to make stuff up that can be very entertaining, and it can be good in creative endeavors, but it can be devastating if you do it in science or in medicine or engineering or in many other areas where, you know, analytical thinking is important.

Holger Hoos 19:59
So one of the big challenges is to sort of marry this other kind of AI with this superficial kind of AI that is broad, that is conversant in natural language, that is much more accessible, that is also much more ill-defined in what it can do and what it cannot do--to marry those two together in a fruitful way. But that is a major challenge. And I would think that most of my colleagues in AI see it exactly the same way. The second big challenge that I'd like to highlight for you, as I said earlier, we study and build AI systems a lot like we approach complex phenomenon in nature, meaning that, you know, very different from an airplane, for example, or a building that we design, we do not understand things at all levels of detail very well with AI systems. Right now, there are a lot of emergent properties, there's a lot of surprises for people who develop them. And that is not per se bad. But it should give us pause, right, if even the experts don't understand the capabilities and limitations and characteristics of the systems that they're building, and we have plenty of evidence that this is the case for all these systems that people are now talking about it, that they're using.

Holger Hoos 21:11
Then I think there is a need for a very targeted effort to explore precisely this: the strengths, weaknesses, limitations and characteristics of these systems. So, we need other directions in AI that deal with that, and I also personally think we need something that I would call monitoring AI systems. And what I mean by that is where monitoring is part of the noun, actually, where these are AI systems whose job it is to watch over other AI systems and make sure that these other AI systems don't go off the rails, so to speak. And as more and more sort of, you know, not particularly qualified people--in terms of AI knowledge, right, and AI expertise--use AI systems, I think these sorts of guardrails, these technological guardrails, will become more important. Regulation alone will not do this. And education alone will also not do that. I think we need AI tools that help us build better AI systems and watch the AI systems that we're building and deploying and manage those systems. So that is also a tremendous challenge. And I see this as an entirely positive challenge. But I do think that these are directions where there are tremendous opportunities to make a difference in the world, to make things better and safer, and, you know, also societally more compatible.

Yan Chow 22:17

So, Holger, from your perspective, is this recent hype around generative AI--does it reflect a fundamental advancement in AI? Or is it more evolutionary? In healthcare, for instance, we were kind of stuck, kind of stagnant, until COVID hit us. And then all of a sudden, there was a big boost to the advancement of healthcare, the reengineering of processes. How do you get organizations that have traditionally been very averse to change, very scared of risk, to actually move in a reasonable way to adopt technology, to adopt AI? Is this a fundamental change in AI that people, you know, as you mentioned, are not fully familiar with? Or is this something evolutionary?

Holger Hoos 23:08
I think it is a sort of a phase transition, an inflection point. And I think it is one that has come in a way that has even experts surprised. Now, of course, you know, there is a bit of a development that some people have been saying for a while might lead to this inflection point. But now that I feel it's basically here, I think many of us are quite surprised that it came so suddenly, and just to not be misunderstood, I am not talking about the inflection point where machine intelligence equals and then surpasses human intelligence, I personally think we are very far away from that--and that is a very good thing, because I honestly think we are not ready for it. And it's also very unclear to me whether going there would even be desirable.

Holger Hoos 23:56
But the inflection point that I'm talking about is one where AI systems have all of a sudden become, in one sense, good enough, and in another sense, broadly accessible enough, that all of a sudden, people really grasp the potential of this technology, on the one hand side. When I started studying this, it used to be a field for some dreamers and specialists and nerds. And now it's something that you can go and talk to your grandmother about AI, right? And she would have heard about this, this is amazing! I would have never thought this would happen when I started studying AI. So, it's now become a mainstream thing. And why is that? Well, it is because it's become powerful and useful enough that it doesn't just touch people's imagination, but it's become real enough that people start to really care. And that takes two forms. Some people get totally enthusiastic and carried away and say, you know, now this is the solution to everything, you know, let's go all out on it. And others get very scared and worried. And in between? There seems to be very little in between those two extremes.

Holger Hoos 25:02
One major factor, by the way, in this inflection point, this phase transition, is this recent development that now we can interact with AI systems like ChatGPT in natural language. There's something very fundamentally different about being able to relate to a system in relatively fluent natural language, compared to, you know, programming it with some sort of formal programming language or some sort of highly technical approach. This really makes a difference, also the way people feel and think about these systems. And of course, you know, we've seen it in science fiction movies for decades, right, all the way back to the 1960s, we had this vision of strong AI. And one key characteristic of these strong AI systems, as you know, "2001: A Space Odyssey," I mean, that's a film that literally is from the 60s, one of the key characteristics of these AI systems was always that you could have a real conversation with them. And now we feel we're getting to the point where we can really have that conversation. And of course, people then jumped to the conclusion, yes, now we're here, full-on human level AI is either here or just around the corner, and we get deceived. And that is something about the inflection point

that I find problematic, we now get very easily carried away--even the experts, even people who really should intellectually know better--get carried away. We ascribe capabilities and characteristics to these systems that they really don't have. And we do that because we can interact with them in this very natural, open-ended way.

Yan Chow 26:33
I think that the human tendency is to ascribe humanness to things that appear to have human qualities. And I think it's interesting that philosophically that entity that we speak to doesn't even need to be fully human, we still tend to relate to human-like qualities almost intuitively. Recently, I read a really interesting article about if we get to the point where it's good enough--not perfect, but good enough--that we can do things like create an avatar of somebody we know, and then we can have the avatar respond to us, you know, with the perceived emotion, with speech and natural language, that we might bring up the possibility of resurrecting our dead relatives. Let's say, our grandmother passed away before my kids, the grandkids, got to know them, we can resurrect them and have them speak the way that they would have spoken if they had been alive. I think that's both exciting and scary.

Holger Hoos 27:28
So, I couldn't agree more with you. Exactly like you say, it's both exciting and scary. I'm not sure what to think about this. And of course, you know, to give some credit to the entertainment industry, Hollywood studios and other big producers have prepared us for this for quite a while, right? I'm not sure about you, but personally, I think Star Trek is quite awesome. Not all of Star Trek, but quite a bit of it. So, there you have a bit of a personal thing for me. And there are Star Trek episodes from 10, 20 years ago that really do a good job in exploring what that kind of thing would mean. So, some people have been doing a pretty good job in thinking ahead. And now we're getting closer to realizing these kinds of capabilities. Now, have we thought enough about them? Before we sort of make it widely accessible? I'm not so sure. I genuinely sit a little bit on the fence on this. But I tend to think that one of the good things about the current, you know, excitement about AI is that it does trigger a certain public discourse, it inspires people to think about this, maybe more seriously than they would otherwise. And that I think is good. And there should be more of this thinking about it.

Holger Hoos 28:36
Regarding the other thing you said, you know, I fully agree that, you know, we tend to ascribe these human qualities, even to systems that are really quite primitive and of course, the great example that some of us who are old enough, actually, I'm technically not old enough, but it was still relatively recent when I was born, and then that means people still talked about it in the 90s, is a simple AI system from the 1960s, from around 1965, before I was born, a system called Eliza and the Eliza effect. This was a system that was actually created, sort of a primitive chatbot. Very, very simplistic, super simplistic. It was created to investigate to which degree people would have natural conversations with a simulated psychologist. And what really happened was that people were amazingly willing to take this thing for real, to engage with it as they would engage with a real psychologist, to a scary degree, even after they were told that it's just a bunch of rules under the hood, and pretty simple rules. I mean, this was a thing, you know, at UBC, where I've worked until seven years ago or so, you know, we used Eliza in my first-year computer science course that I taught as an example of how you can create an illusion of, you know, intelligent behavior. So firstly, computer science students program bits of Eliza. It's so simple,

and nonetheless, people tend to project all sorts of very human-like qualities onto a simple system like this. So, you're totally right. And we've known this for--what is it now--60 years, that this happens. But we still haven't learned to sort of be careful about this. And I think it's something very deep evolutionary in us, that makes us take something that gets not even really close, but just close-ish to natural human behavior, to take that for the real thing. Maybe there is some evolutionary advantage of doing that. But now we have to be careful, because there could also be a big disadvantage of getting carried away with that sort of thing.

Yan Chow 30:29
So, what would you see as the ultimate in terms of your vision for the future? What is the best use of AI vis-à-vis human beings? What kind of roles do we take? Do we have roles? That's our number one question. The second question is, what kind of role? Do we have jobs? You know, what is our value?

Holger Hoos 30:47
That is a really interesting question. I mean, first of all, you know, I sometimes meet people who say, wouldn't it be paradise, if nobody would have to have jobs any longer, and we all were just being cared for by machines, we could just, you know, spend all of our time doing things that we really want to do. I no longer buy this, because I've just read enough serious articles that investigate the sources of human happiness, and what you find is two things invariably--all of these studies find that. One thing that is tremendously important for human happiness is connections with other humans. And so, I feel any technology that tries to minimize or take away or replace these connections is on a problematic path with respect to making us happier, or more satisfied as humans. So, that has immediate consequences. So, for example, some people get very excited about the idea of dealing with this demographic problem that many of our societies have: that there are basically too many old people that require care and companionship compared to young people who could provide this. And then some people get carried away and they say, well, this is obviously where we need very smart robots, robots that can emotionally interact with people. I personally think that's a terrible idea. These are exactly the kind of jobs where even if all the other jobs were gone, these would be the last jobs we would want to give to a machine, right? I think these would be the jobs where once WE are in that situation, we would be desperate to have a real person, even one that we don't like so much, maybe right, rather than a machine--even if the machine does a pretty good job in simulating the person. So, that's the first thing.

Holger Hoos 32:00
The second thing I would say is, I think there is pretty good agreement that what people take a surprising amount of happiness and sense of purpose from is actually their work. So, work makes people incredibly unhappy if it doesn't go well. It also can make people incredibly satisfied and happy if it goes even just well enough. To me, this idea of thinking of a life where there is no work, and you just basically, you know, spend your hours in idleness is not attractive. And I don't think I'm particularly weird in this way. I think there are many people who derive a lot of meaning from the work they do. Now, you mentioned the pandemic earlier, I think one of the things that the pandemic has taught us, other than the importance of human connection, it has taught us that as well, I believe, very strongly so. It has taught us I think that our work, our notion of work, has definitely been too inflexible. And that much is to be gained in terms of satisfaction, efficiency, but also economic benefits in certain areas by making the work environment more flexible. And I'm a strong believer in this, actually.

Holger Hoos 33:27
I'm also a strong believer in what you said earlier: that sometimes it takes a big disruption in order to shake up the system and to enable real progress. You said this very nicely about medicine, where you are the expert, and I am not. But certainly, that was my perception as well. Education is another area where in a relatively conservative country such as Germany, amazingly enough, distance education all of a sudden was adopted at a breakneck speed, and was actually, I mean, I wouldn't say it was an amazing success. But it was successful enough that it really changed people's thinking about distance and blended education. And I think that's a good thing. It's a very good thing. So, what I would say is that with AI, we should make sure that the work that gives us meaning, the work where it's important that people do it, because it's work in which people relate to other people, that that work somehow gets protected. And I wouldn't say reserved, but you know, earmarked for humans, I think that would be nice and important.

Holger Hoos 34:25
I also think that it's very, very important for us to be transparent about work and jobs and tasks done by machines versus people, I very strongly feel that everybody should have a right to know--whether they've been interacting, in whatever form, by written text or spoken word or direct interaction once we're good enough to build robotic bodies that are somewhat, you know, realistic, and we're very far away from that--but I think in all these cases, we have a right to know whether what we interact with is human or human-made, or machine or machine-made, or a mixture of both. And I think we will increasingly see that it's a mixture of both, and, you know, I think that's okay.

Holger Hoos 35:07
Let me come back quickly to your example of health insurance, which I think is an intriguing one. When I worked primarily at Leiden University before moving my primary employment to Aachen, just across the border, I had the pleasure to work together with a Dutch health insurance company that needed some advice and some academic collaboration in some of the innovative work they were doing in that space, right, and I've talked earlier about how skeptical I am of, you know, using AI in insurance and so on. But these guys, they actually convinced me that what they had in mind was something very different from what I was concerned about. And so, I said, sure, as long as we talk about that, why not, right? And their idea was simply this, you know, 90% of claims, health insurance claims, are super routine. And so, their idea was, couldn't we have a system where you just take a photograph of that medical bill, and a system looks at this and only makes two decisions. It doesn't say it's illegible or not, it just says, this is so standard, we can pay it out immediately this second, or it can say this is something a human has to look at. And I thought that's a great system. There is all this drudgery and routine work of approving these hundreds of 1,000s of standard claims that it takes no expertise, no human insight, no judgment to process. It's boring, annoying, repetitive work, and it takes time, which means that people have to wait for their reimbursements. Why not have the machine system in a position where it says yes, if this all looks like standard, we're going to be quick. But if not somebody who really is qualified will look at this in more detail. And that then frees up the people who can do the sophisticated work that requires judgment, for example, to focus on the cases that need judgment. That's a great example for the kind of AI that I actually believe we should be investing in, both economically as well as in terms of

our talent and energy, and that will make life better. And it's also a great example of AI for All because this kind of an AI system truly is of very, very broad benefit.

Yan Chow 37:13
We are actually involved in activities like that, where we automate 70, 80, 90% sometimes, of processes that really could be rules-based, very automatable, and it frees up people. But it does bring up, what do people then do? What is the highest value work that they can do that requires decision-making, or creativity, or problem solving and things like that?

Holger Hoos 37:33
Yeah, I think this is not the first time we're kind of facing this situation, that technological progress makes certain kinds of activities basically no longer viable, right? This happened to a large extent for manual labor in the industrial revolution, as you know, and people were very concerned about it back then, and for good reasons. I think there is a challenge and an art in making these sorts of pretty drastic transitions in a societally compatible and acceptable way. Let me just say, I don't envy the policymakers whose job it is to effect and to guide this transition. I don't envy them that responsibility. I think it's really hard work. It's very easy to bash politicians. But sometimes I think we should step back and think about the difficulties that they are facing too. And this is one of them, right? So, I do feel that yes, there is a difficult transition ahead of us with respect to certain kinds of jobs and activities, and I would fully expect, and I also wouldn't see a problem with that, at the end of this people will work less, fewer hours. I think that's okay, actually. You know, I mean, my recollection is that from history lessons in high school, and so on that in Germany, where I went to high school, in the 1800s, people worked routinely 50-60-hour weeks. Now, some of us do this as well, not because we're contractually obliged to, right, or we would starve if we wouldn't, but because you know, we have other incentives, sometimes we just get carried away at the university, for example, but very few people need to work these hours in Western countries, right? Many other parts of the world, people still do. But I think an argument can be made that technological progress tends to correlate with, you know, larger and larger segments of the population doing work that is less dangerous, less detrimental to health, and also requires fewer hours. And I see this as a further significant step in this direction. So, I would imagine--and I don't think this is problematic--that in 20 years, people in highly industrialized countries might work on average only 25 hours, maybe even less than that. But what they do in these hours will count for more, because it's going to produce more value. I think that's fine. But it's not a transition that you want to happen, to have happened too rapidly, because the jobs that could get lost if it's not through natural attrition, you know, people just retiring and not being replenished in these jobs, that would be very hard on them.

Yan Chow 37:38
One of the stories that's often repeated--don't know if it's true or not--when the internet first came, people, again, were very worried about job loss, and eventually, but it took time, five times as many jobs were created after the internet. So, the challenge is, what are the positions that are going to be created by AI? One of the ones that I've read recently about is actually the skill of prompt writing, you know, being able to converse properly with a machine to get what you need out of it. And I'm sure there'll be many, many other things like that.

Yan Chow 40:33

But in a sense, those jobs have to be at a level that humans can handle. Yeah, I see that in the future, there will be AI that humans cannot understand, you know, that we like the benefits, but we don't really know or we can't explain--or do you think it is explainable at some point, the whole art, the whole science of making AI explainable?

Holger Hoos 40:54
So first of all, Yan, I have to chuckle a little bit, because you're saying in the future, there might be a point. I'm saying that point is here, right, we DON'T understand the AI systems that we have! That's the first thing. The second thing I am going to say sort of semi-seriously is you're talking about this new job of prompt writer. You know, in human interactions, we call these managers, right? They're the prompt writers that sort of help other people maybe do their job if they do it well. I mean, a little bit more seriously, we all know that, even when we don't talk about technological systems, but just people, the skill of relating to people and interacting with them in just the right way is absolutely crucial in organizing anything that requires more than one person working on a job. And the more people you have, the more difficult these management and social engineering problems become. So now we're talking about basically, you know, managing AI systems, for lack of a better word. I do think you're right, in that this will become a major skill, a very valuable skill, especially in this transition phase, where AI systems are still relatively brittle, and it takes a lot of insight and experience to sense and to detect when they go off the rails even subtly.

Holger Hoos 40:54
So I think now, I would maybe liken it to--I mean, I don't have personal experience with this--but I imagine that people who train wild animals, they need to develop a lot of skill in predicting the unpredictable, to know the unknowable, to deal with something, in this case, you know, maybe a tiger or a lion or something like that, that clearly is pretty intelligent, but also in a way that's very different from us. And they need to somehow make this safe and effective and entertaining, as it were. I mean, let's not talk about the philosophical issues of training animals. But I think there's a challenge there that people can perhaps relate to. And so, I think interacting with AI systems nowadays is sort of like this, only more extreme in that the AI system, of course, is less sophisticated in many ways. And the sophistication it has, in terms of its capabilities, it's a lot more sort of idiot savant, right? So, there's a lot more sort of very specialized, but that of course, is going to change. That is one of the things I would expect to change over the next 5, 10, certainly 15 years, and our way of interacting with these systems will become more natural and less guarded. And it will take less specialized expertise to interact with these systems meaningfully, to make them do the right things and to make them not do the wrong things. Because if we do our job, we in this case are people like me, right, then we will develop AI systems that are less brittle, that are more aware of their own limitations, and that will, for instance, increasingly say: look, this is something I can't help you with, because it's outside of what I've been trained to do.

Holger Hoos 43:39
And I'm not sure about you, Yan, but you know, what is one of the most sophisticated things I do in my job? I'm a professor, you know, so I would say one of the most sophisticated things I do is, I guide young researchers to the degree where they leave the university with a PhD degree. What is a PhD? A PhD is a certificate that says: this person now is an independent researcher; they are now my peer. So,

my job at the university to a large degree, as far as education is concerned, is to produce people who I can confidently release with my blessing to do the kinds of things that they couldn't do when they started, or they shouldn't have done when they started. And so, my feeling is that ultimately, the question we have to ask such people, and I do in PhD exams, one of the most important questions is the following. And of course, you don't ask it literally like this, you ask it in a somewhat, you know, sophisticated way, but you ask these people something that clearly goes beyond their level of knowledge, and you know it and they know it too. And you want to see how they react. You want to see how they react when they get out of their depth.

Holger Hoos 44:48
And I think you know, in medicine, you probably are looking for this as well, you know. I imagine maybe people in the operating theatre, people will have the guts and the maturity to say, I'm now in over my head. I need help, I need a second opinion, right? My GP during the pandemic really impressed me. I had a little issue. It was a little issue, totally harmless. But he said at some point, you know what, this is something I would like to have my colleague, who is in the same practice, have a look at it, I hope you won't mind. I hope you won't think I'm incompetent or anything like that. I said quite on the contrary, you know, you just impressed me with that statement more than with anything before because I think it's important that experts know their limitations. This is really a crucial part of what it means to be an expert in anything, you have to know your limitations. And that's why I think before we can accept AI systems as experts, and I'm not talking about expert systems, I'm talking about AI systems doing an expert's job, they need to be aware of their own limitations, they need to be able to articulate that, and they need to very clearly signal under the right circumstances: I'm in it over my head, I need help. I can't do this, I won't do this, it would be irresponsible for me to do this thing. And we see far too little of that. Instead, what we see is so-called AI experts, human AI experts, that demonstrate a shocking lack of what I would call intellectual humility and sort of skepticism when dealing with their own and their colleagues' creations by precisely not asking the question: is that still, you know, within its limits, and to understand what this thing is doing, but quite on the contrary, by, you know, just running away with it-- and that makes me a little concerned. I think we, the AI community, and everybody using AI systems too, we need to focus on the limitations and on the weaknesses, not because it's so scary, but because that's what we would do in people too. We would ask, is this person qualified to design the bridge? Is this person qualified to do the brain surgery? Is this person qualified to push that button that will launch this mega rocket that if it blows up on the path might do a ton of damage? We ask these questions; we should ask these questions of AI systems just the same.

Yan Chow 47:03
People don't perceive risk. Going back to the example of the car, if they hadn't had a lot of experience with driving, they may not perceive the risk. And so, with AI right now, everything is positive, and everything is a lot of fun, and people don't perceive the risk until something happens. And COVID was something that showed us how inflexible healthcare was--until something happened to test that. So, I think one of the interesting ideas, and I know that generative AI might be able to do this is to simulate what might happen. So, is there a place for simulation as a way to create different scenarios of risk that we really would have to look at, more concrete than just theoretical?

Holger Hoos 47:42

It's an interesting thought. I mean, you know, AI-based simulation, I think, is a very important area for many reasons. Actually, one of the reasons why I'm very excited about it and I think why it's absolutely crucial to develop this further is climate change challenge because there, we have exactly one system. Whatever we do to interact with it is being very carefully reflected, so much better to have a digital twin of that system, where we can run scenarios. What a great way to maybe persuade at least enlightened policymakers of what's really at stake here and showing them credible simulations of what could happen. Now, as for AI itself, I think that's very difficult because ultimately, you're asking for a system that we have today to maybe simulate a system that we may or may not have in five or 10 years. That's tricky, right? So, I think you're asking for a simulator that simulates something that's probably far more complex than itself and while this is not in principle undoable, I think it's very difficult to get this accurate.

Holger Hoos 48:38
My feeling is, for what it's worth, we talked about driving. So for the better or for the worse, a lot of 16 year olds are driving cars, right, big heavy cars too. And I think it's astounding how well they do this on average. I mean, of course, we know that there's higher risk in that age group, especially the middle part of the population. But we know that they still can be amazingly safe drivers, considering the complexity of that, as for example, measured at the difficulty of getting autonomous driving to work. So, what I would say is, how do these kids learn how to not go into the blind corner at 100 miles an hour, right? They don't learn this by simulation, at least not in Germany, or the Netherlands and Canada, not when I last looked. They don't go to the driving simulator and fail 30 times until they finally figure it out, right? But quite on the contrary. I think maybe this is the first time in our conversation that I disagree a little with you. I think people actually tend to have reasonably good intuitions about what is risky and what is not, and these intuitions need to be maybe nurtured and developed in the right way, but I think the seeds are actually there.

Holger Hoos 49:47
But maybe it takes a bit of guidance to see this, and so with driving and autonomous driving, one of the lessons that will always stick with me when I took driving lessons many, many years ago was the question that the instructor, he said the following: look at this picture here; it was a picture of a typical residential area. And he said, now you're driving along here at, let's say, 30 kilometers an hour, because you're really responsible. And behind this car out comes a basketball, what do you do? And everybody in the room said, "I step on the brakes," right? Because you know that after that ball with a high probability comes running a kid, right? So, this is common sense reasoning, you don't even need to teach people that. But this is something that's so different from current machine learning. Current machine learning needs lots of data, it needs lots of bad experiences. This is a situation that is a great combination of sample size one and common sense. And of course, you know, a lot of machine learning research goes in this direction too. But we're really not there.

Holger Hoos 50:45
Now what I would say is, people are actually pretty good at this kind of stuff, and they should also be pretty good at dealing with the risks and challenges of AI. We just have to open their eyes to it, we have to reinforce their natural instincts of caution, without making them panicky and/or negative, because I will say this, I'm happy to go on record and have done so many times with concerns about AI and so

on. But I also think we can totally not afford to forego the opportunity to use AI to solve some of our great challenges, at least to some degree. This includes climate change, it includes inequity. I think AI holds one key--not THE key, but one key--to making progress on all of these, if we do it right. And therefore, I think the real challenge is not should we do it or should we not do it, but how should we do it so that it comes out right, empowering people, recognizing their natural instincts, and helping them course correct, when their natural instincts mislead them, for instance, into trusting a system they shouldn't trust, just because superficially the system behaves human-like, right? There, we should caution them, we should help people to become natural skeptics, but we should also reinforce their natural instincts where, you know, their instincts are right with respect to AI. You know, we can do this in terms of training people to become better managers, to be better in interacting with other people, why shouldn't we be able to help them become better with, you know, dealing with artificially intelligent systems and their limitations and weaknesses? Of course, we can!

Yan Chow 52:13
Very good points. The time has really flown. What are some last thoughts that you may have before we close? Also, reiterate how people can get a hold of you.

Holger Hoos 52:24
Sure. So first of all, it's been great spending this time with you. Really insightful questions, I have to say--questions that also make me think, which is great. Secondly, what are my final thoughts on this topic? So first of all, there's a meta final thought, right, that there can't be final thoughts on AI because we're still early in the journey, right? So, it's a very exciting development that we're witnessing here. It's not scary, but it's one that demands of us caution and prudence. I think we should become far better than we naturally are at detecting hype and also sort of the hallmarks of weakness of machine-generated content, because if we don't, this could really open the door to mass manipulation--to, you know, a few bad actors having a disproportionate impact on public opinion, all very problematic prospects, I don't think it's inevitable that these things come to pass and come to bite us. But I think we need to watch out, and we need to actively do things in order to make people aware. So, this is one of the reasons why I very much enjoyed talking with you today. Because my hope is that some of the things I said perhaps stimulate people to think, and maybe even stimulate them to think not only enthusiastically, which is always great, but also a little bit skeptically. And I think a healthy dose of skepticism, when it comes to AI is a good thing. It's a very good thing.

Holger Hoos 53:51
The other point I'd like to make is that it's not enough to be worried and scared about AI, we also have to embrace the positive sides, the prospects of AI really helping us solve problems that we couldn't solve in an unaided way, right, and to deal with situations, with systems, also with each other in ways that are responsible and respectful and effective more so than we could without that kind of additional help. And I do think there is a lot of value in AI systems that help us realize and overcome our weaknesses, and that amplify our strengths rather than trying to replace our strengths and share our weaknesses. I think this is really important. In order to make sure that AI research and development can be focused at least to good part on this,

Holger Hoos 54:39

I think it's very important that we now come to the same decision as we came more than 20 years ago when people got the ability to decipher the human genome. And then it was said, this is too important for humanity, for society as a whole, to leave it to industry alone. I'm not anti-industry at all. I think industry is doing an amazing job and plays a very important role in AI as well, right? So, it's all good. But AI is too important to leave it to industry alone to develop and to use responsibly, right? So what I think we really need is a big public investment that makes sure that top notch AI research and development can also--not exclusively, but also--be done in the public domain, paid by the public, responsible to the public, and focused on the needs and worries and concerns of the public. We cannot expect companies to do this, it's against their DNA, it's against their nature, right? For a good reason, they should be worried, amongst other things, of course, but they should be very worried about profitability.

Holger Hoos 55:09
We need top notch AI development in a space where short term profit is not the primary driving factor. It's really crucial. And so therefore, I really, really hope that political decision makers will wake up to that need and make the investments that are necessary to make that possible, so that the brightest minds in the area of AI and closely related areas no longer have a choice to make: do I want to do top-notch cutting-edge research, which means that I have to work for profit-driven industry, whether I want to or not? Or do I care so much about the benefit that can come from my work beyond just profitability, that I'm willing to work in environments that don't allow me to do cutting-edge research? That is not a choice a bright young person should have to make. And so if I had one wish, I would say in five years, the best and the brightest, they should be able to make the choice between working in a great industry environment, for profit, for economic benefits, and all the good things, or to work in a similarly well-equipped and high-end public facility where they can do it for the betterment of society and for the public good. And some will choose one way, and some will choose the other. And that's the way it should be. But I would want our best and brightest to have that choice. Currently they don't, and the way it's going, they won't--unless there will be a massive public intervention.

Yan Chow 57:08
This technology is so important, there is a place for a national or international initiative that actually is already going on, but maybe in countries that are not thinking the same way. Tell us again, how to get a hold of you.

Holger Hoos 57:20
So, the easiest way to interact with me, I'm afraid, is through the social media that I use, which is LinkedIn and Twitter, so I do try my best to react to people reaching out there to me. And another good way is through my very capable staff who are helping me to stay on top of email by not sending email to me, but actually to the addresses that are published on our website here at RWTH Aachen University.

Yan Chow 57:45
Great. Holger, thank you so much for sharing your perspective, a very informed perspective, on AI. I really enjoyed our fascinating conversation and, like you, I look forward to where AI is going in the next five or 10 years--and more than that, how we as societies and countries, and as a human race, decide

how we will relate to and how we will therefore develop AI. I hope when we look back 10 years from now, we'll feel even more optimistic.

Holger Hoos 58:12
As a naturally born optimist, that's certainly my hope and actually my expectation as well! It's been truly a pleasure doing this with you, Yan. Thanks for giving me the chance, and I do think your listeners will enjoy our conversation as much as I have.

Yan Chow 58:27
I'm sure they will. Thank you so much, Professor Hoos, really appreciate your time and your perspective.

Holger Hoos 58:32
Thank you.