AI 4 Science Discovery Network+

AI4SD Interview with Professor Tony Hey
02/03/2021
Online Interview

Michelle Pauli
Michelle Pauli Ltd

11/08/2022

AI4SD Interview with Professor Tony Hey
Humans-of-AI4SD:Interview-29
11/08/2022
DOI: 10.5258/SOTON/AI3SD0243
Published by University of Southampton

Network: Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery

This Network+ is EPSRC Funded under Grant No: EP/S000356/1

Principal Investigator: *Professor Jeremy Frey*
Co-Investigator: *Professor Mahesan Niranjan*
Network+ Coordinator: *Dr Samantha Kanza*

Contents

1 Interview Details

Title	AI4SD Interview with Professor Tony Hey
Interviewer	MP: Michelle Pauli - MichellePauli Ltd
Interviewee	TH: Professor Tony Hey - STFC
Interview Location	Online Interview
Dates	02/03/2021

2 Biography

Figure 1: Professor Tony Hey

Tony Hey: 'When you have a lot of data, you need all the tricks of the trade to make sense of it'

Professor Tony Hey is the Chief Data Scientist at the Science and Technology Facilities Council. Tony's original background was in physics, completing his undergraduate degree and subsequent post-docs at the University of Oxford in the UK and then CalTech and CERN in the USA. He worked at the University of Southampton in the physics department originally before transferring to the electronics and computer science department where he created a leading research group in parallel computing. He was the director of the UK's e-Science initiative (2001-2005) and then became the Vice President in Microsoft Research afterwards.

In this Humans of AI4SD interview he discusses his long career, converting scientists to Microsoft's tools, the two dramatic developments in the past 10 years, the role of commercial cloud resources, advice for early career researchers, and much more...

3 Interview

MP: What's been your path to where you are today?

TH: When I was a kid, I used to read science fiction: authors like John Wyndham and an oddball called Fred Hoyle. His books, The Nature of the Universe and The Black Cloud, were fascinating and stimulated my interest in science. I went on to read introductions to quantum mechanics and relativity too. At that time, I was in a good school in Birmingham, but it was a conveyor belt where I had to press a button marked either Oxford or Cambridge. I ended up pressing Oxford, because at Cambridge, I'd have to do things like biology, which I didn't want to do, but I was a natural at physics, which I could do at Oxford.

I really wanted to know about the Dirac equation, relativistic electrons, and positrons, which you didn't really get taught as an undergraduate in those days. That's why I did a PhD in Oxford. There was a seminal moment when I went to Caltech, and heard from two Nobel Prize winners: Murray Gell-Mann and Richard Feynman. They didn't care about whether you lived or died, they cared about whether you were researching something interesting, which was a wake-up call. My thesis at Oxford was fine, but it wasn't an interesting problem in the world.

I then went to CERN for a couple of years, before coming back to Southampton, where I progressed in physics. At that stage, physics had an understanding of gauge theories, and we had a putative theory of weak and strong electromagnetic interactions. To solve the strong interactions, you needed to use computer simulations, as the coupling constants were very large. That was how I got into doing computer simulations of field theory. While I was on sabbatical at Caltech in 1981, I heard Carver Mead give a wonderful talk about the origin of Moore's Law and how there were no engineering obstacles to making things smaller, faster and cheaper for the next 30 years; it actually went on for over 40 years. When I came back, I wanted to have faster computers, but I couldn't afford a supercomputer. What I could do, however, was put lots of small, cheap microprocessors together and do parallel computing by dividing a program into little bits. When I moved to electronics and computer science, I set up a group in parallel computing. We were one of the first groups in the world to look at how you write programs when they need to be broken up into little pieces. It was quite pioneering.

Then I got sucked into being the head of the Department of Electronics and Computer Science. The electronics scientists and the computer scientists never talked to each other, and some regarded the others with contempt. But funnily enough, I discovered I could actually bring them together.

Later, when I was the dean of engineering, I was asked to lead the UK's E-Science programme, which was a wonderful experience for five years. There we did what would be called data intensive science nowadays. That's how I came across Jeremy Frey: I worked with him on the Combichem project. Jeremy was one of these very rare things, a really original chemist. He was always very positive, always looking for the next things, and he did some wonderful experiments. We did lots of pioneering work in the UK, which involved computer scientists working with chemists.

After that, I got a job in Microsoft Research, where I was involved in technical computing. I was extremely lucky to get that, and luck plays a huge role in one's career. You've got to be in the right place at the right time, but you also have to be lucky. In my work with Microsoft, I got the opportunity to do external research with universities around the world: from the UK and US to India and China.

In one sense, it was a big change to move from working in universities to Microsoft, but in another, it wasn't so big because the research community is international. I had already spent time doing sabbaticals in the US and had a large number of international connections in Asia. It helped to already have a large network.

What was difficult, however, was that academics love using Google and wouldn't be seen dead using Microsoft tools. They also use Linux and practice open source. I couldn't be seen to be funding people to use Google, but it wasn't realistic to think that people would throw away their nice Macs for a clunky PC. It was a complicated balance. Our solution was to do open source tools around a Microsoft product. For example, we made an open source plugin for chemical equations in Microsoft Word. Then we did something similar for Excel. For scientists working with relatively small amounts of data, their standard tool for organising it was an Excel spreadsheet. Our plugin helped you actually annotate and get metadata about the data in your spreadsheet. We also created a free tool called Worldwide Telescope, which you could use to make wonderful displays of big astronomical surveys. You could take a tour through the solar system, our local galaxy or, in fact, the entire universe.

One of the people we funded as a Microsoft research fellow was Professor Fei-Fei Li. She led a project called ImageNet, with which you could test your computer vision system to see, for example, if your computer could distinguish between a poodle and a labrador. The project involved a huge volume of images, which people were paid to label, then you could see if a computer vision system could be as accurate as a human. The answer was no, it had huge error rates. But then in 2012, there was the discovery of deep neural networks, which reduced the error rate by 10%. Now, with the error rate on that particular benchmark, computer systems are much better than humans at detecting these things. That was a big breakthrough, which people in industry are now using for things like autonomous vehicles and speech processing: they're all applications of deep neural networks.

I then returned to the UK, because I wanted to make sure that UK scientists were not disadvantaged with respect to their US counterparts. I managed to persuade the Turing Institute that they needed to have some computing if they wanted to get involved with deep learning. Some people hear that and think, "We've been doing AI for the last 20 years, it's just more of the same," but that's not the case. Deep learning was a big change and there's the possibility of breakthroughs like AlphaFold.

When I arrived back at the Rutherford laboratory, nobody was interested in machine learning. It started off with just me running a machine learning group, but we've grown now after I got funding for a few people. We use NVIDIA GPUs, which can do lots of calculations at once, making them very useful for training networks in a short space of time. I started calling this kind of research AI for Science and there are now AI for Science programmes at the Turing Institute. The AI4SD programme is also an AI for Science network. They're good because they raise people's awareness that when you have a lot of data, you need all the tricks of the trade to make sense of it.

I have a friend who is a pathologist, and they used to use Computer Vision to do image analysis of body organs. But now they're thrown those away and they just use deep learning. It's a dramatic change. That's really why I wanted to make sure the UK wasn't being left behind. We've made some progress, but there's plenty more to do.

MP: Where would you say we are right now, a decade on from the Fourth Paradigm?

TH: The Fourth Paradigm was Jim Gray's phrase and it's a good one. That's why we wrote the book on it, to get scientists to reflect on where they were. In the past 10 years we've seen the end of Moore's Law: chips don't just get faster, cheaper, and smaller, they can get smaller and you can get more memory, but you won't get faster computing because the chips eventually melt! That plus deep learning are the two things that have changed the most dramatically in the past 10 years.

MP: Where are we with open science at the moment?

TH: Open science is one of my passions, ensuring that data is fair, findable, accessible, reusable, and interoperable. It's the interoperable bit that's contentious. People want to get all the semantic data out of the paper, and they believe you can only do that using exotic technologies.

Schema.org is a good example of interoperability: when I was at Microsoft, it was the only thing Microsoft and Google could agree on. It allows you to provide semantic information on, for example, a website, so that the system can figure out when, for instance, I search "Casablanca," that I'm looking for the movie rather than the place in North Africa. This tool isn't as exotic as the semantic web, but it is more familiar to people.

So we're getting better with open access. David Litman, who did the National Library of Medicine, had the idea of linking papers to data and digital object identifiers, so you can search from the paper to the data. That's where Jim Gray got his vision from to do that for all the datasets that you could, and then add new data to yours, and so on. You could increase the efficiency of the whole process.

MP: What needs to happen at the national level to support first-class research?

TH: One of the things that concerns me is that in the National Science Foundation, for example, they're concerned with giving academics the sort of huge computing resources that are reachable in industry. They are trying to set up a cloud, if you like, through commercial clouds, which grants access to state-of-the-art, large amounts of computing. Maybe we should do something like that in the UK. Although many of my colleagues like building systems themselves, I think there's an argument for using, when appropriate, commercial cloud resources. Companies can get hacked very easily but Microsoft and Amazon are paranoid about keeping their clouds safe.

MP: What advice would you give to early career researchers?

TH: They have to do something they care about; they need some passion. If I could go back in time, I'm not sure I would do particle physics again, but maybe that's because I've seen the exciting times take place.

I was listening to The Life Scientific this morning, and there was a cosmologist from Cambridge who was looking at the production of carbon dioxide from farming. She suddenly switched to using her statistical techniques for gravitational lensing for analysing farming data, and she has found her passion. The question is, what will you tell your kids when they ask you "What were you doing when climate change was happening?" Do you want to say "I was wondering about the stars and cosmology"?

If I were an early career researcher right now, I would hope to be doing something I was passionate about. I'd like the UK to remain relevant in the future. With AI in general, we need to ask how we can use it to solve some of the problems in disease or climate change, these are the grand challenges I'm trying to formulate. I'd like us to have some ambition in the UK to solve the problems that are really hard and interesting. All I can do nowadays is cheer people on!