AI 4 Science Discovery Network+

AI4SD Interview with Dr Heather Kulik
29/11/2021
Online Interview

Michelle Pauli
Michelle Pauli Ltd

11/08/2022

Humans-of-AI4SD:Interview-32

AI4SD Interview with Dr Heather Kulik
Humans-of-AI4SD:Interview-32
11/08/2022
DOI: 10.5258/SOTON/AI3SD0229
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 Dr Heather Kulik
Interviewer	MP: Michelle Pauli - MichellePauli Ltd
Interviewee	HK: Dr Heather Kulik - MIT
Interview Location	Online Interview
Dates	29/11/2021

2 Biography

Figure 1: Dr Heather Kulik

Heather Kulik: 'AI has dramatically changed how science is done"

Heather J Kulik is an Associate Professor in Chemical Engineering at MIT. She received her B.E. in Chemical Engineering from Cooper Union in 2004 and her PhD in Materials Science and Engineering from MIT in 2009. She completed postdocs at Lawrence Livermore (2010) and Stanford (2010-2013), prior to returning to MIT as a faculty member in 2013 and receiving tenure in 2021.

In this Humans of AI4SD interview she discusses the different ways her group is using machine learning, algorithm aversion, supporting researchers' mental health and her advice for early career researchers.

3 Interview

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

HK: I'm an associate professor in chemical engineering at MIT, where I've been for about eight years. My undergraduate degree was in chemical engineering at the Cooper Union, and I received my PhD in 2009 from MIT in the department of material science. I then went on to do a chemistry postdoc at Stanford and the Livermore National Laboratory. Currently at MIT, my group does research in the areas of computational chemistry and machine learning. We started out focused on first-principles modelling, but increasingly have incorporated machine learning into our efforts.

MP: What research are you currently working on?

HK: My group focuses primarily on accelerating the discovery of new materials and challenging spaces that are lacking in intuition: for us, that's metal-organic frameworks and transition complexes. These are materials that have promise both as functional materials, things in quantum sensing and quantum computing, as well as in the detection of small molecules, in separations and catalysis. What we're addressing are the outstanding challenges in energy and resource utilisation, looking at how we can get to answers faster using computational techniques.

Because the properties of transition metal complexes and metal-organic frameworks are governed by the laws of quantum mechanics, it's very difficult to come up with heuristics for why a material behaves the way it does. Whereas we can look at the structure of an organic molecule and know how it will behave, these rules break down with transition metal chemistry. It's also a space where the approximate methods computational chemists have developed tend to fail. This is especially the case in the area of transition metal chemistry, in which my group works, which contains mid-row transition metals with unpaired spins. The molecules which interest us belong to a class of materials called strongly correlated materials. Not only do their properties depend on the laws of quantum mechanics, but they also rely on very high-cost quantum mechanical modelling for accuracy, which is inherently incompatible with high throughput screening and accelerated discovery.

MP: How is machine learning making a difference here?

HK: My group uses machine learning in a number of ways. First, assuming that density functional theory is correct, it can help us get to answers faster. That means we're able to accelerate the discovery of new materials, taking the discovery efforts that would normally last decades down to weeks.

The second way machine learning helps involves the fact that all approximate models are slightly incorrect. Machine learning can help us focus on the regions where they can be trusted the most, where consensus is highest and uncertainty is lowest. Here, machine learning helps us to get the right kind of answer, which can predict a module on a computer that could then be realised experimentally.

The third way we're using machine learning is a newer effort for us. With some of these properties, the only way you can learn about them is if someone takes the time to do an experiment, but we're not able to model these experimental results on a computer. Instead, we use natural language processing and automated image extraction, which aggregate the

answers collected by experimentalists over thousands of papers. We then train machine learning models directly on that data, in order to make predictions about what makes a material behave the way it does. There, we've looked at questions like: at what temperature will a material break down and lose its structural integrity? We've done this with metal-organic frameworks, which are emerging, exciting materials for separations, catalysis, and quantum computing. However, their chief limitation right now is that they're not robust: they're not stable at high temperatures and harsh conditions. Being able to predict what it is that makes a material stable is an important step towards realising them as engineered materials with practical applications.

MP: What challenges have you faced with this research?

HK: One challenge is that the materials we study are quite large, really at the limit of what computational methods can be used for. So the combination of the large system size and the difficult electronic structure makes it hard to know if we're getting to the right answer when it comes to computational modelling.

There are also a number of challenges with predicting thermal stability. Experimental groups often don't report their data in the same way universally, and while a group might say that the material's quite stable, the data might show that it's not as stable as they claim — there's a positive publication bias. That means that when materials are not stable, we don't always have that data to feed our machine learning models.

My group has worked to overcome this challenge by publishing both our dataset and the machine learning models on a website called MOFSimplify. There, we encourage folks to give feedback on both our data and the predictions that our models make. They can rate it from one to five stars, depending on how they feel about it, and they can also deposit their own data. We're hoping this will be a community resource helping to collect more data, including unpublished data, and to solicit feedback on how well our models make predictions. That interface between what the models predict and what the expert thinks of it will be important to improving machine learning models.

MP: What has surprised you in this research?

HK: One of the things we focus on is structure property relations: we try to understand which atoms matter where in a material. In order to do this, we have to do feature selection, which involves identifying which parts of the material matter by using statistical techniques. Recently, we published a result that I had never expected. Feature maps are ideally built on the highest-fidelity data, because raw predictions coming out of a low-fidelity model are not going to lead to a good result. What we found, however, was that among different models, the feature maps are consistent. This was a surprise for me because I had always thought that feature maps were likely sensitive to where we got the data from. Had we done a careful experiment or a careful calculation that was more accurate, we thought the feature maps would change, but we found that these feature maps are pretty invariant and universal.

In terms of the experimental MOF (metal-organic framework) work, the things that surprised us are perhaps less surprising in retrospect. Most people think of MOF stability as governed by the identity of the metals and the nodes of certain elements. Generally, we found that the elements people thought led to stable MOFs did not, and that it's more about linker chemistry. Linkers are the connectors between the nodes that make up a MOF periodic material. So that was somewhat surprising, how small, subtle changes in the linker chemistry

could change predicted stability so much.

MP: How has AI machine learning changed how science is done?

HK: AI has obviously already changed the way we do research, because it's changed the focus of how we develop tools, and what we spend our time on. In that sense, AI has dramatically changed how science is done. The uncharted area is the question of the role of expert knowledge: what can an expert do that an AI or machine learning model can't do? How can the relation interaction between machine learning models and the expert be optimised? You want there to be the opportunity for feedback in which people feel like their expert knowledge is valued and can influence how model predictions are made. At the same time, however, people tend to overestimate the accuracy of their expert knowledge. If someone tells you that a machine learning model is accurate 90% of the time, you might focus on the 10% of the time that it's wrong. But you might not be aware that you, as an expert, are only accurate maybe 85% of the time.

One underserved area involves understanding the best way to leverage expert knowledge and not sidelining the expert. We also need to understand in what ways machine learning models can free up experts to focus on different tasks, and how machine learning can automate those more tedious things to allow researchers to focus on hypothesis testing.

In that respect, the main thing my group has started doing is looking at machine learning models that predict the decisions expert computational chemists would normally make to automate workflows. This means that all the tasks that we normally carry out to set up a series of calculations to discover new materials can instead be carried out autonomously by the machine learning models themselves.

MP: Is it hard to get to the point where the expert is able to trust the automated machine learning model?

HK: Yes. This area of research actually has a name: algorithm aversion. We're trying to understand how experts interact with machine learning models versus how a novice, who perhaps has less resistance, does. This area of research is still in development, and it has implications across everywhere that we interact with AI.

The thought is that there needs to be an opportunity for the machine learning model to be overridden, but for this to be a complementary interaction. A lot of the machine learning models we've developed replace things that are the tedious steps: predicting if the density functional theory calculation will succeed, for example. We're very happy to automate away these types of things. What we haven't had to address yet is the question of making ourselves obsolete, but it's something that I think about a lot. When I present my research, I often get questions like, "Are you trying to replace us?" and "What is my role as an expert researcher?"

I think the most important thing to study is how do we maximally leverage a machine learning model so experts are still developing knowledge. That involves not simply outsourcing everything to machine learning models, like, for example, how no one can find their way around anymore without Google Maps. We need to ask how much do we need that hands-on expertise to develop both intuition and expert knowledge? What is that expert knowledge and how can it be leveraged with machine learning?

MP: Your research group places a large amount of emphasis on supporting researchers' mental health. Why is this important to emphasise?

HK: There's a good trend in the US at the moment to acknowledge that people are whole humans with a need for a balance between their professional and personal lives. MIT is a pretty fast-paced place, and it attracts those types of students. It's important to remind them to take a break and have time away from their work, to ask them "Have you taken a vacation this summer? Have you taken a break?" I try my best not to ask students for things on nights and weekends. I remind my students that I get a full night's sleep and I encourage them to as well. It's important to recognise that our roles as mentors involve reminding people that it's ok to take time away, and not to push themselves too hard.

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

HK: The questions I get increasingly from younger folks are: "What is it like to be a woman in academia? Are things getting better? Is there more inclusivity?" This might be counterproductive to say, but it's helpful to ignore all of that for as long as you can. Don't count yourself out of science just because you don't see as many people who look like you at each level. The fact that you've made it this far probably means that you have a lot of skills and perseverance. That would be my advice to women and other underserved groups, in particular.

My advice to early career researchers more generally is read as much as you can, collect as much knowledge as you can, and think about skills development. Recently I read a book called So Good They Can't Ignore You by Cal Newport. It talks about how developing skills as capital is one way to ensure job security. I know that, depending on where you are in your career, obtaining an academic position can feel like a lottery. But this book makes the good point that no matter what skills you develop, you'll never regret having developed great skills that could help you in industry or in academia. My advice is to think about how to make your contribution to the field singular and to focus on developing unique skills.