IMPLEMENTING LABORATORY AUTOMATION SYSTEMS THAT USE ARTIFICIAL INTELLIGENCE TECHNIQUES TO AUTOMATICALLY EXECUTE CYCLES OF SCIENTIFIC EXPERIMENTATION

LARISA SOLDATOVA, Reader in Data Science and Director of Online MSc Data Science, Goldsmiths, University of London

Dr Larisa Soldatova is Reader in Data Science in Goldsmiths, University of London and Director of the Online MSc Data Science Programme. Larisa is an internationally recognised expert in knowledge representation, laboratory automation and data analytics. She is involved in a number of international projects on the development of semantic standards, e.g. the ontology for Data Mining, ML schema, laboratory protocols EXACT. Selected awards: BCS Machine Intelligence prize (2006), nomination for the World Technology Award (Software) (2006), RCUK Fellowship (2007-2012), Meta-QSAR grant (EPSRC) (2012-2014), AdaLab grant (EPSRC) (2014-2018), Big Mechanism Grant (DARPA) (2014-2018), ACTION on cancer grant (EPSRC) (2018-2022).

What are the latest developments in implementing lab informatics in 2019?

I have been working in this area for 15 years, and I have been involved in projects focusing on automating laboratories. The aim is to have IT systems that can do everything without human intervention, from thinking about what questions to research, which involves generating hypotheses and designing experiments, to doing real experiments: analysing results, updating knowledge, and based on this producing new knowledge, again generating more hypotheses and so on. These systems have existed in academia for 15 years. Pharmaceutical companies and other industries may have more complex lab systems, but not if put together with software. The most advanced systems are integrated, and they have machine learning components, many AI tools, and real automated labs.

From an academic point of view, we always struggle with limited funds. It is not even comparable to industry. The only way we can produce such advanced systems is to be very intelligent in order to save this money. An example of what we are trying to do is using active machine learning, but again, in academia, we just don't have computer power and resources comparable to pharma companies. It forces us to look at other solutions. For many problems, this might actually be the most suitable solution. In drug discovery, usually pharmaceutical companies are screening a lot of compounds using high throughput methods. If there are then some indications of activities, they pick them up and study them more closely to see if it can be a potential drug or not. We cannot buy millions of compounds. It is why we use different methods, for example, active machine learning. It's very clever; from studying even 20 chemical structures it can start predicting which chemical to test that will give you the most information about what experiments will be the most beneficial to run. So that is why it's called active, as it can tell you which chemicals to test. We have very interesting results with hundreds of chemicals. Our largest library, Hopkins Library, is 1500 chemicals. These are challenges of academia; they are challenges but also opportunities to try out something different. We would like to work with industry, of course, and we do have some collaborations, but it can be challenging. We have had good connections with Pfizer, for example, but then they closed the site in UK. You need to start building new collaborations from the beginning. There is also another trend of people who worked in academia starting their own successful companies. Personally, I would like to see more of such interactions. This is why I'm grateful for such events like this one; it is an opportunity to improve collaboration. I have been participating in several such events and found them interesting, and I also see from attendances that interest is growing. Many of our postdocs and PhD students who come to train in academia move into companies,

and they're doing fantastic work now in this AI area. In addition, some academics later move into companies. That is why I think that, in many cases, industry is often more advanced than academia.

What are the main areas on which you are currently working?

Our main interest is in automated scientific discovery; how much you can do to create new knowledge, and how much of it you can automate. Our answer is that it is possible but it's hard to make it scalable and practical. There were a lot of talks at the event about FAIR data and data integration; it's a huge problem, as it's hard to put things together, but there are other issues. It's not only FAIR data, there should be FAIR software and there should also be connections between software data and wet lab experiments. In our recent project this was the most pressing challenge. There are huge advantages in AI and in laboratory automation of this high throughput screening, but in terms of putting it together so that there is no slow stage of interpretation in between there are still challenges. As such I believe the next steps in research will be from more advanced integration of more heterogeneous entities. There is a huge distance between the wet labs and computational labs. Even in connecting simulation with real experiments it is still not how it should be. I believe that the next big step will be interpreting knowledge. Right now, this deployment is driven by availability of huge data sets and what you can do with them. The next step is when you have knowledge: the meaning of this data on the cloud, how different data sets are connected and interrelated, what we know about it, to what extent can we trust it, how reliable is it? Also, there are gaps, there are duplications, contradictions. So, it's coming closer to human reasoning. We do that every day without realising we deal with the information around us and how it can be fragmented, incorrect, contradictory, yet we're still making some decisions. I believe that the next step will be going one level up from Data to Knowledge, with all this machine learning decision-making using all the information.

What are your expectations for the future of this field? Is it bright?

Yes. Not many people would call me an optimist, but yes! This area will only grow in importance and in recognition, and there are so many exciting areas to look at. I can give you an example of how this importance and recognition is increasing. I am Russian, so I need visas I need to go through immigration. When I go to conference, even several years ago, I go through immigration and they ask: "what is this conference about?", and you have to explain to an immigration officer what it is you are doing and they look at you very sceptically. Last year, you just had to mention it's an AI conference and they say "oh it's about the robots? Yes, please, come in!" People finally know what it is, so the future looks very optimistic. But seriously, of course, there are a lot of concerns that AI will kill us all. The reality is always somewhere in between; there is a lot of over excitement and a lot of scepticism, but the reality is that it is an extremely promising technology. Of course, it can be dangerous, just like any technology - electricity is dangerous if you don't use it correctly. It's up to us how we use it. However, I agree that there should be conversations about it, and there should be relevant policies implemented. The people who are working in these areas should be involved; at the moment, they are probably not involved enough. This means that the people who are developing policies don't understand enough. But again, this is because the technology has progressed so quickly. People haven't realised what's happening and suddenly it's here; and its influencing everyday lives quite profoundly. You can talk to your smartphone using voice recognition and face recognition. It is all around us. It has been here several years and trying to predict where we will be in the next three years is exciting, but of course there are also problems.

You recently spoke at our PharmaTec Series. What were the three main takeaways of your presentation?

In my talk, I talked about the result of our European project on the lab where we try to develop this fully automated cycle of improvement of systems biology models. It's an extremely complicated process and it is clear that without automation we will not be able to scale up because biology is just so complex. We made it work for that particular task, identifying the problems and the possible solutions. So, I like to deliver a message of looking at it at the larger scale, how it can be applied to other problems, what needs to be done for this. If you want to move forward with scientific discovery, to make it faster, cheaper, and more efficient, this is what we need to do. We need integration, we need standards, we need open software, and we need to do it with as many partners, with industry and academia working together.