

Impact of Big Data

Dr. Sunnie Chung leads the Big Data Lab at CSU and its future is looking very bright



THE BIG DATA LAB at CSU is no place for old textbooks. It's a place where the targets for research and teaching are always moving, at the speed of the internet.

"In our history of technology, either in science or engineering, there is no such technology evolving this fast," says Dr. Sunnie S. Chung, the lab's director.

"Big data" refers to the collection of data sets so large and complex that it's difficult to process with traditional systems and methods. Massive data are constantly generated by systems or a variety of Internet of Things (IoT) devices capturing transactions, interactions and observations of a general population or environments. All of us both generate and consume data in our daily activities.

The volume of data is increasing explosively. There are 44 zettabytes of data in the digital universe as of 2020, Dr. Chung says, and the rate of expansion is accelerating. Get used to the word zettabyte — that's a trillion gigabytes. Imagine the challenge of designing and constantly redesigning a curriculum for a field that is changing that rapidly.

"In computer science, especially in big data-related areas, every six months you see the changes — like new methods developed in research becoming new industry standards or some methods becoming obsolete," Dr. Chung says.

And, now universities are playing catch-up. "Big data and its related areas have evolved so fast, mostly from industry research at the beginning," she says. "There's not enough workforce educated and trained in these new areas. Industry is waiting with these amazing jobs and opportunities."

So, establishing curricula to generate data scientists for big data has become one of the most important priorities to every university, according to Dr. Chung.

In social media, Dr. Chung has been studying systems for real-time opinion analysis. In IoT, her interest is in cybersecurity and the privacy of cloud computing systems for mobile health applications and network intrusion detection systems. Finally, in the health science area, Dr. Chung has been developing an intelligent system that identifies urban food desert defect areas and their impact to the development of diseases in Cuyahoga County.

Cyber security and privacy in the IoT area is a good example of the research challenges in the big data era. IoT devices communicate with cloud servers to send the data in encrypted forms. "There's no way to process them in a cloud without decrypting them," Dr. Chung says.

"The problem is the cloud is public, so all these sensitive data are in the public and they need to be decrypted for computations by the cloud servers. That raises serious security and privacy issues."

The solution is "very special encryption methods," she says.

Dr. Chung is doing research on encryption methods called "semantic preserving encryptions." That puts the data in the specially encrypted forms that the cloud servers can process while the data is still encrypted. The challenge is to make the special encryption strong enough to protect the big data without being breakable by hackers and yet fast enough to complete the computations over the encrypted data for the big data applications.

"If the computation over the encrypted big data is too slow, then it's useless," Dr. Chung says. "It's a pretty challenging area, and it has been a focus of research in collaborations between academia and the major big data industry such as Google, Amazon, Microsoft, IBM and Oracle, which builds the cloud computing systems."

One of the most exciting initiatives in computer science is the establishment of the Internet of Things Collaborative (IOTC) between CSU and Case Western Reserve University. The IOTC is funded by the Cleveland Foundation and led by Dr. Nigamanth Sridhar, professor in the EECS department and an NSF program director. Dr. Chung contributed significantly to the establishment of this multi-million dollar (more than \$7 million over six years) collaborative, and she has received seed funding to perform research in IoT.

CSU has been an important part of the local big data community.

One of Dr. Chung's favorite activities has been the organization of five workshops on big data at CSU in the last five years. These workshops have provided an excellent platform to the regional big data research and development community, both in industry and in academia, as well as to students. She says the workshops have been attended by about 170-230 people each year, and CSU students have had an excellent networking opportunity to be connected to the big data industry.

"Attendees came from local big data industry and universities, and presenters came from big data industrial research groups and academia," she says. "A lot of them were invited from Silicon Valley. This has been the perfect platform for CSU students to be exposed to the big data industry."

Representatives from Google, Cloudera and IBM Watson have taken part in the workshops. Northeast Ohio industries such as the Cleveland Clinic, Progressive, Rockwell, Parker Hannifin, IBM Health Watson and



WE'RE SEEING REALLY STRONG, POSITIVE RESULTS IN TERMS OF STUDENT SUCCESS.

DR. SUNNIE CHUNG



NASA that rely on big data have been there. So have other area universities like Case Western Reserve University and Kent State University.

Dr. Chung says she's seeing "really strong, positive results in terms of student success." That includes hires by Google, Amazon, Facebook and Yelp, but also by major Northeast Ohio employers like FirstEnergy, Hyland Software, Parker Hannifin, Rockwell, Sherwin-Williams, Progressive and the Cleveland Clinic. "Every industry needs big data technology and big data-related knowledge and skill sets," she says.

Dr. Chung has developed courses to help students prepare for those opportunities. The new curriculum has also had a positive effect on the research activities in the computer science program, with a significant increase in the number of thesis and dissertation projects, as well as in the number of graduate and undergraduate research projects in the last few years.

Dr. Chung's undergraduate students have had success in continuing for their doctoral studies with full scholarships/fellowships at iconic institutions, such as Northwestern University and the University of Southern California. Her first Master's thesis student, Danielle Aring, is now working for FirstEnergy as the first data scientist that CSU graduated. Another one of her Master's thesis students, Andrew Yu, is now teaching at Penn State Behrend. Nick White, whom Dr. Chung mentored in his independent study and senior design project and who won first place in the senior design poster competition in 2016, is now working for Facebook and attending Stanford University for his Master's. Finally, Suhua Wei, whom Dr. Chung advised in her research project, is now working at Amazon.

Dr. Chung's future plans are to enhance research on big data, AI and cyber security and establish global big data and data science programs at both the undergraduate and graduate levels. This will lead to the recruitment and retention of even more students and will help establish the computer science programs at CSU as among the best big data and data science programs.