EXEMPLARS SESSION

The EXEMPLARS session highlighted successful models of training and programme development in artificial intelligence and data science. The speakers were Ani Adhikari from the University of California, Berkeley (UC Berkeley), Rajesh Sundaresan from the Indian Institute of Science (IISc), Vipin Chaudhary from the Case Western Reserve University, Andrew Thangaraj from the Indian Institute of Technology Madras (IITM), Rohini Srivathsa from Microsoft India, and Krishnan Narayanan and N Dayasindhu from itihaasa.

Ani Adhikari described two ways of thinking about data science and data science education, and data science itself as a way of thinking. Usually, the students who do data science are very well advanced in mathematics, computer science, probability theory, and statistics. At UC Berkeley, a different perspective on data science was adopted.

She spoke of UC Berkeley's approach: (i) start with no prerequisites; (ii) teach data science as a way of thinking, as a way of understanding the world, as a way of knowing how to work with the information on hand, how to ask questions, and use information in ways that enables one to answer the questions, able to make inferences and generalise to aspects of the world that one has not observed/cannot observe. Equipped with this thinking, students can decide whether or not they want to use it in their fields. Based on this, UC Berkeley would like to craft a programme enabling students to take higher level courses in data science, use data science in diverse fields, and integrate societal context and ethics throughout.

Ani Adhikari also explained UC Berkeley's data science engineering and educational programme. She outlined the milestones crossed from a pilot offering of Data 8, Foundations of Data Science (in 2015), the formation of a Division of Data Sciences (in 2017), offering the first data science major with a cohort of about 100 student graduates (in 2019), and the launch of a data science minor and support for the creation of a hub to advance data science (in 2020). A few months ago, 700 students received their degrees in data science. In 2022–2023, a college of computing, data

science, and society will be established to partner with Berkeley's other renowned colleges.

The number of students who have started seeing themselves as potential data scientists have increased. For example, students of public policy and fine arts now see how data science impacts their fields and how their fields can impact data science.

Rajesh Sundaresan spoke about IISc's focus on developing the artificial intelligence (AI) workforce and addressing issues of scale. The Institute could train the trainers – train the drivers for change who will then be able to bring about the next revolution.

He talked about engineering education in the twentieth century that had dealt largely with classical core topics such as aerospace, chemical, civil, electrical power, mechanical, and metallurgy. These core areas dealt with materials and energy generation, storage, distribution, and utilisation. The reason for this was the need to convert energy from one form to another, and transport this energy, materials, and people. It was necessary to build a workforce that would build these systems. The physics of these problems needed a certain mathematics training, leading to the core engineering mathematics curriculum comprising topics such as matrices, multivariable calculus, complex analysis, transforms, ordinary differential equations, partial differential equations, numerical methods, basic linear algebra, and basic probability.

There has also been a rise in the generation, storage, distribution, and utilisation of information, mostly driven by computer science and information technology entities, control and communications engineers, and operations research people. Are the traditional training programmes in these areas sufficient?

An example of transformative applications is PageRank, which is based on a random web-surfer model that combines Markov chains and novel distributed algorithms to handle scale, speed, and personalisation. Here, the traditional training was not sufficient; the mix of randomness, scale, distributed processing and data storage had to be addressed. Similarly for machine translation, linguistics and statistics had to be combined to get translation at scale; corpora had to be gathered and curated.

The core topics that students need to know to be able to make substantial contributions to such domains are rapidly mixing Markov chains for the Markov Chain Monte Carlo (MCMC) method, computational linear algebra, high-dimensional statistics, network analytics, computational topology, particle filters for data assimilation, random matrices, algebraic geometry applied to optimisation, cryptography and privacy, game theory, and mechanism design. These topics are often not covered in basic courses, and only lightly touched upon in advanced courses.

The traditional programmes do not address these modern requirements of high levels of computational, mathematical, and data analytic skills. The training is scattered and not cohesively organised. There is a lack of guidance, direction, and opportunities when students work on these systems at the undergraduate level.

To address this, IISc has formulated programmes that involve training at various levels. There is a new BTech in Math and Computing, where students will be trained to build these technologies of the future. There are many basic and thematic certificate programmes such as in computational and data sciences, deep learning, digital manufacturing and smart factories, and digital health/imaging. IISc offers Master's programmes (MTech) in Artificial Intelligence, Robotics and Autonomous Systems, and Quantum Technologies. To retrain professionals while they are working, an MTech (online) in Artificial Intelligence, Data Sciences and Business Analytics is also offered.

Rohini Srivathsa spoke about Microsoft India's engagement with the AI ecosystem over the last three-and-a-half to four years on multiple dimensions of AI skilling. In 2018, NITI Aayog released the National Strategy on Artificial Intelligence (NSAI), and in 2019, Microsoft India began its efforts.

Rohini Srivathsa pointed out that by 2024, India is going to be the largest developer ecosystem in the world. How would huge developer communities, who have all the possible cognitive services on the Cloud today, going to use it responsibly? In 2019, Microsoft held a Week of AI, wherein the target was the developer ecosystem. Multiple sessions of large-scale basic hands-on

training on machine learning algorithms were held. Over 45,000 developers have been skilled so far.

Microsoft also partnered with NASSCOM and GitHub to conduct a Microsoft AI Classroom Series for students. Partnering with the Ministry of Human Resource Development, an AI curriculum was created; this curriculum is part of the Central Board of Secondary Education (CBSE). 'Train the trainer' sessions were also conducted, wherein there were different modules for each subject. For example, the AI module comprised search algorithms, probabilistic reasoning, Markov decision process, and reinforcement learning while the data sciences module included introduction to programming tools for data science, mathematical foundations, machine learning, and case studies.

Ten AI Digital Labs were started in partnership with ten higher education institutions such as the Birla Institute of Technology & Science Pilani and Periyar University, to provide infrastructure, curriculum, training, access to Cloud, AI services, developer tools and support. An AI Business School was also started, which had four modules on strategy, making an AI-ready culture (AI-driven decision making in various functions of the organisation), AI 101, and ethics and responsibility in AI.

In a country of India's scale, connecting skilling to employability is important, said Rohini Srivathsa. Through Future Ready Talent, a virtual internship platform, Microsoft has partnered with the Government and various organisations to provide training to students in technology skills and enable them to work towards solving real-world problems using Microsoft Azure and GitHub tools. The credits earned during the internship programme would be recognised in the students' college degrees. More than 160,000 students had registered from more than 1200 colleges spread across tier 1, tier 2, and tier 3 cities, and 100% of the 2021–2022 students are employed.

Vipin Chaudhary talked about AI for (i) computer science and data science students, focussing on undergraduate and graduate education; (ii) non-data science and non-computer science students.

At the Case Western Reserve University, there are standard undergraduate AI/machine learning (ML) courses, and specialised AI courses at the graduate level for computer science and data science students.

For non-data science and non-computer science students, there is an AI minor across the university. How do you deliver an AI minor without having students take the foundational courses. Here, there are two tracks: a technology track and a cognitive science track. In the technology track, around 15 courses are offered that span different departments such as neurology of behaviour, neurobiology, computer vision, advanced algorithms, manufacturing and automated systems, applied controls, and robotics. A minor would typically have around five to six courses. A student would choose three courses from this track and two required AI courses. In the cognitive science track, there are courses such as introduction to neurobiology, linguistic analysis, mathematical logic and model theory, physiological psychology, cognitive psychology, and cognition and information processing.

Vipin pointed out that every university has a data science programme, and most of them have data science institutes that are located in different parts of the institute and driven by research agendas. Earlier, a domain that needed data science would start teaching maths, statistics, and computer science. Each unit would teach the same courses related to data science, leading to a wastage of resources. To address this, a new department was formed at Case, which was a combination of data science and computer science. The undergraduate course in data science has a requirement of 11 core courses such as introduction to data structures, mathematics, algorithms, and introduction to probability; two breadth courses such as cognition and computation, econometrics, and bioinformatics for systems biology; at least three foundation courses; two technical electives. Students have to choose three to four courses from other domains ranging from economics, finance, neuroscience to materials science. Students can also create their own domains.

For students who are not computer science and data science majors, an applied data science programme was developed. The goal was deep domain knowledge and broad applied data science knowledge. Six courses are offered, with a graduate data science certificate and an undergraduate minor. Around 60–70 students of the 1600 students admitted to Case every year go through the minor, and the data science major is one of the fastest growing majors at the university (the rate is higher than computer science).

Andrew Thangaraj spoke about issues of scale in India. The number of students passing out of Class 8 in India is 215 lakhs, which is close to the total number of people of that age. However, there is a dramatic drop by the time they come to Class 10, and an even bigger drop when they come to Class 12.

The number of people who gain undergraduate admissions is 84.65 lakhs, which is 27% of the gross enrolment ratio (GER; the fraction of people in that age group who should be in college compared to how many are there in college). Out of this, more than 60 lakh students enter the Bachelor of Science (BSc), Bachelor of Arts (BA), and Bachelor of Commerce (BCom) streams.

Andrew pointed out that the big problem in India is college admissions. There is extreme competition for top institutions; the costs of studying at quality private institutions is high; there is low employability and low GER. What is needed are large public universities offering bachelor's degrees that are employment oriented, of high quality and affordable by the majority.

In the above context, what is IITM doing? Andrew spoke about two major digital interventions. Under Swayam, nearly 1000 courses are offered every semester from more than 100 organisations; more than 25 lakh people enrol and learn for free; more than 3.5 lakh people write in-person exams; for more than 2 lakh people, the credits are transferred into their degree programmes. IITM started the Bachelor of Science (BS) in Data Science and Applications in January 2021. This programme is offered in hybrid mode – online instructions with in-person exams every month. There are no eligibility criteria for admission other than a pass in the Class 12 exam. However, there is a qualification process wherein the students study the basics that are essential for

doing data science over a period of four weeks, after which they appear for an exam. If they pass the exam, they enter the programme. The programme comprises multiple levels. First, there is a foundation level in which students take courses in mathematics, statistics, and computing. It is possible to exit at this stage with a certificate. The next level is the diploma level, where it is possible to get two different diplomas – one in application development and the other in data science. Students have to do four projects at the diploma level. If students pass both the diplomas, they can get into the degree level. At the degree level, it is possible to get a three-year BSc or a four-year BS.

Many students who have taken up BS in Data Science and Applications at IITM are also doing another degree online. This has been made possible by the University Grants Commission (UGC) online degree regulation that students can be on campus in one programme and do another programme online. This is a solution to the domain knowledge versus data science problem. Andrew opined that the much-needed large public university in India can have academic contributions largely from elite institutions and can be in a hybrid digital mode. There is also an ongoing discussion about a digital university.

Krishnan Narayanan spoke about a survey that was conducted, involving the higher education institutions in India, to understand the AI and data science landscape. A hundred and thirteen responses were received from a diverse set of institutions.

One surprising finding was that in India, there is a very high prevalence of Bachelor's programmes in AI. Of the 62% of respondents who were already offering AI courses, 80% were offering a Bachelor's degree. Standalone AI programmes seemed to be very popular. There is also a high prevalence of units completely dedicated to AI. Online programmes are being offered both at the undergraduate (UG) and graduate (G) levels. Some of the courses that are not being offered at the UG and G levels by 25% of the departments are philosophy of AI, AI and brain sciences, multi-agent systems, human–computer interaction, speech processing, robotics and automation, and ethics. The respondents' answers on the challenges that they faced when offering AI and data science programmes were in computing infrastructure and data availability, and collaborations (industry and international).

N Dayasindhu spoke about some action plans that they had decided upon such as conducting the current visioning workshop; studying AI and DS programmes in India and the US, and identifying exemplars (course curriculum, datasets and tools, resources); organising follow-up workshops/meetings on AI and DS workforce development; accelerating efforts to scale up computing infrastructure and access to large datasets; identifying potential mechanisms to address some of the challenges such as faculty development, access to specialised content, research pipeline, and collaborations.