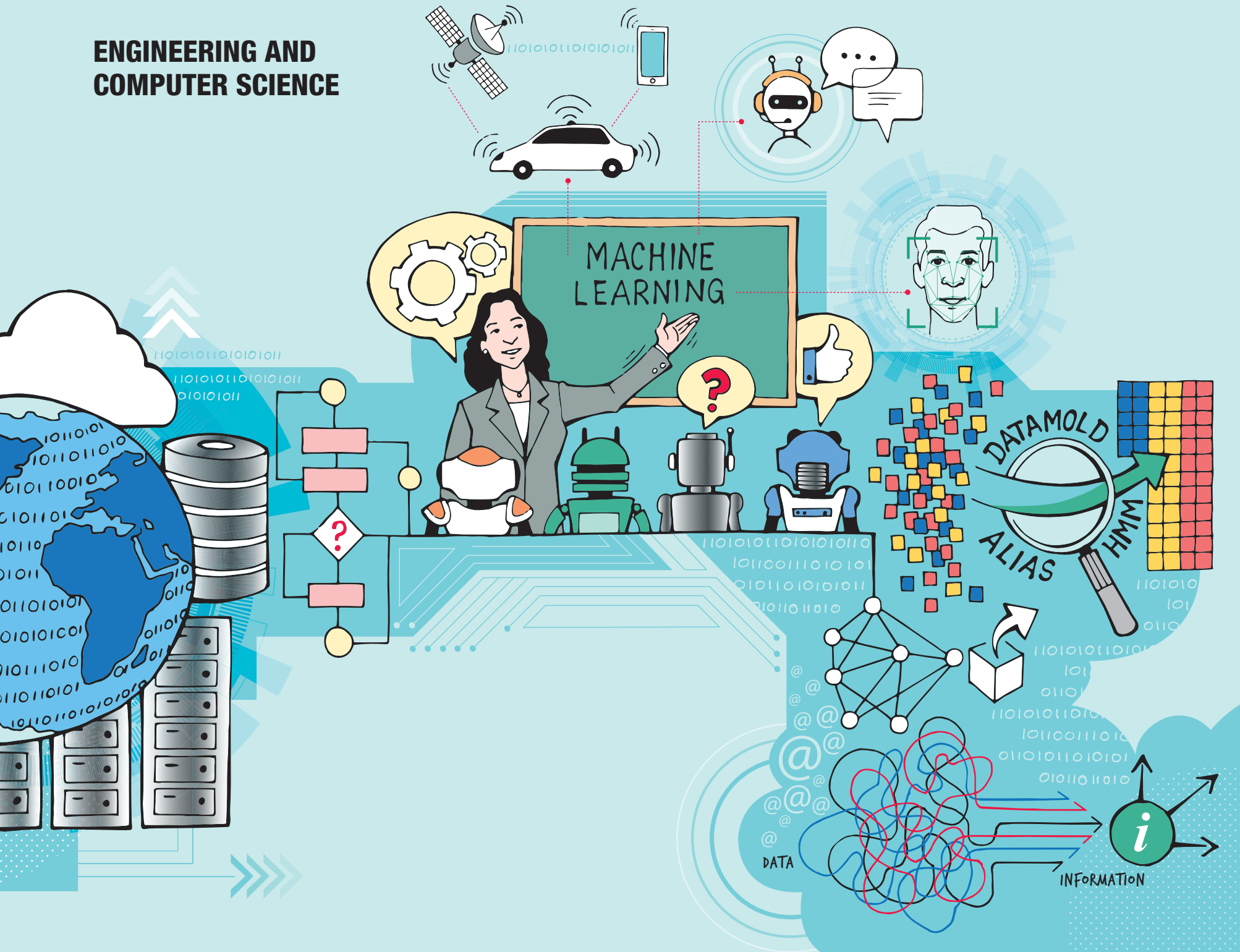


# ENGINEERING AND COMPUTER SCIENCE



## INFORMATION IN THE AGE OF DATA

We live in the age of data. According to one report in 2018, 2.5 quintillion bytes of data were created each day. This figure grows every day.

With all this data being generated how do we make sense of it all? How do we go about extracting relevant information from these vast oceans of data? This is where machine learning comes in. Machine Learning is the science of teaching computers with examples where explicit programming is difficult. Machine Learning is behind the recent success stories we hear about machines driving cars, chatting like humans, and recognizing your face in a photo.

Prof. Sunita Sarawagi's research involves machine learning where she teaches machines through examples to automate tasks that humans find too tedious. This includes cleaning up badly written addresses into a proper structure and detecting duplicate addresses in large lists such as those kept by the income tax department. Prof. Sarawagi and her collaborators developed a software package called DATAMOLD which efficiently improves address structuring. Sarawagi employed the theory of Hidden Markov Models to create this solution for unstructured data. HMMs are a class of probabilistic graphical model that allow us to predict a sequence of unknown (hidden) variables from a set of observed variables. She developed a software package called ALIAS that can efficiently dig out mutants of addresses in a dataset of millions of addresses.

Prof. Sarawagi developed algorithms for extracting factual answers to questions from noisy tables on the world wide web. She showed how a machine can self-teach itself by harnessing the diverse redundant ways in which information is distributed over the Web. She along with collaborators developed a new type of model called the Semi-Markov Conditional Random Field that was key to developing such an extractor.

Prof. Sarawagi's research has huge implications in a world where vast quantities of data need to be converted into useful information that can then be used for everything from new business ideas to real social change.