We have so much genomic data—we need AI to help us grapple with it

Genomics is set to become the biggest source of data on the planet, overtaking the current leading heavyweights – astronomy, YouTube and Twitter.

…

There is a need for computing to move ahead of the upcoming challenges in genomics – not just to ensure we can store and process the vast quantities of data. There is also going to be increasingly diverse data to grapple with, as efforts to sequence the genomes of all life on earth are underway. … More widely, 'real time' genome analysis of bacteria and viruses can help to track outbreaks of infectious diseases …

Though the concept of machine learning has been around since the 1960s, it's only in the last 10 years that it's really been applied in genomics. Three factors have converged allowing its potential to be realised – the algorithms are sophisticated enough, the data sets needed to train the algorithms now exist, and the computing power to train those algorithms exists.

Dr Nicole Wheeler, a researcher at the Centre for Genomic Pathogen Surveillance at the Sanger Institute has used supervised machine learning to train an algorithm to spot genome sequences in Salmonella bacteria that are associated with a deadly bloodstream infection, as opposed to mild food poisoning. …

Read full, original post: Getting smart about artificial intelligence