Google's Larry Page wants to save 100,000 lives but big data isn't a cure all

30 June 2014, by Eerke Boiten

Talking up the power of big data is a real trend at the moment and Google founder Larry Page took it to new levels this week by proclaiming that 100,000 lives could be saved next year alone if we did more to open up healthcare information.

Google, likely the biggest data owner outside the NSA, is evidently carving a place for itself in the big data vs life and death debate but Page might have been a little more modest, given that Google's massive Flu Trends programme ultimately proved unreliable. Big data isn't some magic weapon that can solve all our problems and whether Page wants to admit it or not, it won't save thousands of lives in the near future.

Big promises

Saving lives by analysing healthcare data has become a major human ambition, but to say this is a tricky task would be an enormous understatement.

In the UK, the government has just produced a consultation on introducing regulations for protecting this kind of information alongside care.data, a huge scheme aiming to make health records available to researchers and others who could work with it.

Given the ongoing care.data debacle, this is a broadly sensible document and a promising start for consultation. In particular, it identifies different levels of data. Data that could be used to identify an individual person should not be shared in the same way as other types of data.

But, like Page, the UK government is also presenting a false vision for big data. It has said review after review have found that a failure to share information between healthcare workers has led to child deaths. It's an emotive admission but rather beside the point in the big data perspective. It is indeed entirely credible that many tragic failures within the NHS might have been prevented by someone sharing the right information with the right person. Sharing is essential, but when the NHS talks about sharing, it means linking and sharing large medical databases between organisations. Surely no case review has ever claimed that the mere existence of a larger database of information would have got the right knowledge to the right person.

Medical data sharing may be a good thing in many ways, but unfortunately there is no clear case yet that automated analysis of data prevents child deaths and other tragedies. It is only big data, not magic. Preventing child deaths appears to be brought in as emotional blackmail, expected to trump the valid concerns over the NHS' big data plans.

Big disappointments

The fact is, we are not as advanced as we would
like to believe. This month, 60 years after Alan Turing died, his test for recognising "true" artificial intelligence made the news again. One in three human test subjects mistook a computer programme called Eugene Goostman for a 13-year-old Ukrainian boy. But Eugene didn't really pass the test. The programme was simply good at playing the game and relied heavily on the fact that a 13-year-old probably wouldn't know the answers to many of the questions.

The programme fell back on the same tactics used some 42 years ago by Parry, a programme that tricked people into thinking it was a paranoid schizophrenic, and the even earlier Eliza programme which had proved hard to distinguish from a real Rogerian therapist. So much for progress.

The research field of artificial intelligence – or more modestly, machine learning – has been active for 60 years and passing the Turing test is its original Holy Grail. And many of the brightest minds in computer science have worked in this area. Computing power has been increasing exponentially over that time and the web provides a massive amount of samples of human communication to learn from. The fact that we have made such slow progress despite all these developments shows just how hard it is to turn vast amounts of data into human intelligence.

Be wary of big claims

This should teach us to be wary of anyone who makes bold claims about the potential of big data. Google Flu Trends sought to derive information about the spread of illness by gathering data when people searched for terms like "flu". But we've seen time and time again that machines don't understand humans and can't mimic real human qualities.

A prime example can be found outside healthcare. It's now broadly accepted that in the course of its surveillance programmes, the NSA had obtained information that might have prevented 9-11, but failed to join the dots.

Edward Snowden's revelations made it clear that the NSA and GCHQ are collecting large "haystacks" of communications data. The intelligence services have made various claims that the analysis of this prevented serious terrorist attacks, but these claims have not stood up to detailed scrutiny. Given the amount of computing power the NSA possessed, even before the internet age, it must have been applying machine learning techniques to its bulk data for at least 30 years. Still, no evidence has been presented of any significant needles being found as a result – at least not any that is available to the public.

This all goes to show that using machine learning to process vast amounts of data, such as the information held in healthcare databases, won't save lives alone. The kind of human insight needed to put the information to proper use still can't be replicated by computers, even after decades of trying.

Doctors need to be able to ask the right questions and use their unique human qualities to make life changing decisions for their patients. Similarly, researchers still need to formulate their hypotheses and ask the medical databases targeted questions. They are not machines, and we should be grateful for that.

This story is published courtesy of The Conversation (under Creative Commons-Attribution/No derivatives).