

HOW MACHINES LEARNED TO SPEAK HUMAN LANGUAGE

AND WHAT DOES THAT MEAN FOR THE WAY PEOPLE USE COMPUTERS?

THE ECONOMIST

JAN 11TH 2017

THIS past Christmas, millions of people will have opened boxes containing gadgets with a rapidly improving ability to use human language. Amazon's Echo device, featuring a digital assistant called Alexa, is now present in over 5m homes. The Echo is a cylindrical desktop computer with no interface apart from voice. Ask Alexa for the weather, to play music, to order a taxi, to tell you about your commute or to tell a corny joke, and she will comply. The voice-driven digital assistants from America's computer giants (Google Assistant, Microsoft's Cortana and Apple's Siri) have also vastly improved. How did computers tackle the problems of human language?

Once, the idea was to teach machines rules—for example, in translation, a set of grammar rules for breaking down the meaning of the source language, and another set for reproducing the meaning in the target language. But after a burst of optimism in the 1950s, such systems could not be made to work on complex new sentences; the rules-based approach would not scale up. Funding for human-language technologies went into hibernation for decades, until a renaissance in the 1980s.

In effect, language technologies teach themselves, via a form of pattern-matching. For speech recognition, computers are fed sound files on the one hand, and human-written transcriptions on the other. The system learns to predict which sounds should result in what transcriptions. In translation, the training data are source-language texts and human-made translations. The system learns to match the patterns between them. One thing that improves both speech recognition and translation is a “language model”—a bank of knowledge about what (for example) English sentences tend to look like. This narrows the systems' guesswork considerably. Three things have made this approach take a big leap forward recently: First, computers are far more powerful. Second, they can learn from huge and growing stores of data, whether publicly available on the internet or privately gathered by firms. Third, so-called “deep learning”, which uses digital neural networks with several layers of digital “neurons” and connections between them, have become very good at learning from example.

All this means that computers are now impressively competent at handling spoken requests that require a narrowly defined reply. “What's the temperature going to be in London tomorrow?” is simple (To be fair, you don't need to be a computer to know it is going to rain in London tomorrow). Users can even ask in more natural ways, such as, “Should I carry an umbrella to London tomorrow?” (Digital assistants learn continually from the different ways people ask questions.) But ask a wide-open question (“Is there anything fun and inexpensive to do in London tomorrow?”) and you will usually just get a list of search-engine results. As machine learning improves, and as users let their gadgets learn more about them specifically, such answers will become more useful. This has implications that trouble privacy advocates, but if the past few years of mobile-phone use are any indication, consumers will be sufficiently delighted by the new features to make the trade-off.