Coming to grips with voice technology - For my next trick

https://www.economist.com/node/21712484

Jan 5th 2017

Talking machines are the new must-haves

IN “WALL-E”, an animated children’s film set in the future, all humankind lives on a spaceship after the Earth’s environment has been trashed. The humans are whisked around in intelligent hovering chairs; machines take care of their every need, so they are all morbidly obese. Even the ship’s captain is not really in charge; the actual pilot is an intelligent and malevolent talking robot, Auto, and like so many talking machines in science fiction, he eventually makes a grab for power.

Speech is quintessentially human, so it is hard to imagine machines that can truly speak conversationally as humans do without also imagining them to be superintelligent. And if they are superintelligent, with none of humans’ flaws, it is hard to imagine them not wanting to take over, not only for their good but for that of humanity. Even in a fairly benevolent future like “WALL-E’s”, where the machines are doing all the work, it is easy to see that the lack of anything challenging to do would be harmful to people.

Fortunately, the tasks that talking machines can take off humans’ to-do lists are the sort that many would happily give up. Machines are increasingly able to handle difficult but well-defined jobs. Soon all that their users will have to do is pipe up and ask them, using a naturally phrased voice command. Once upon a time, just one tinkerer in a given family knew how to work the computer or the video recorder. Then graphical interfaces (icons and a mouse) and touchscreens made such technology accessible to everyone. Frank Chen of Andreessen Horowitz, a venture-capital firm, sees natural-language interfaces between humans and machines as just another step in making information and services available to all. Silicon Valley, he says, is enjoying a golden age of AI technologies. Just as in the early 1990s companies were piling online and building websites without quite knowing why, now everyone is going for natural language. Yet, he adds, “we’re in 1994 for voice.”

1995 will soon come. This does not mean that people will communicate with their computers exclusively by talking to them. Websites did not make the telephone obsolete, and mobile devices did not make desktop computers obsolete. In the same way, people will continue to have a choice between voice and text when interacting with their machines.

Not all will choose voice. For example, in Japan yammering into a phone is not done in public, whether the interlocutor is a human or a digital assistant, so usage of Siri is low during business hours but high in the evening and at the weekend. For others, voice-enabled technology is an obvious boon. It allows dyslexic people to write without typing, and the very elderly may find it easier to talk than to type on a tiny keyboard. The very young, some of whom today learn to type before they can write, may soon learn to talk to machines before they can type.

Those with injuries or disabilities that make it hard for them to write will also benefit. Microsoft is justifiably proud of a new device that will allow people with amyotrophic lateral sclerosis (ALS), which immobilises nearly all of the body but leaves the mind working, to speak by using their eyes to pick letters on a screen. The critical part is predictive text, which improves as it gets used to a particular individual. An experienced user will be able to “speak” at around 15 words per minute.

People may even turn to machines for company. Microsoft’s Xiaoice, a chatbot launched in China, learns to come up with the responses that will keep a conversation going longest. Nobody would think it was human, but it does make users open up in surprising ways. Jibo, a new “social robot”, is intended to tell children stories, help far-flung relatives stay in touch and the like.

Another group that may benefit from technology is smaller language communities. Networked computers can encourage a winner-take-all effect: if there is a lot of good software and content in English and Chinese, smaller languages become less valuable online. If they are really tiny, their very survival may be at stake. But Ross Perlin of the Endangered Languages Alliance notes that new software allows researchers to document small languages more quickly than ever. With enough data comes the possibility of developing resources—from speech recognition to interfaces with software—for smaller and smaller languages. The Silicon Valley giants already localise their services in dozens of languages; neural networks and other software allow new versions to be generated faster and more efficiently than ever.

There are two big downsides to the rise in natural-language technologies: the implications for privacy, and the disruption it will bring to many jobs.

Increasingly, devices are always listening. Digital assistants like Alexa, Cortana, Siri and Google Assistant are programmed to wait for a prompt, such as “Hey, Siri” or “OK, Google”, to activate them. But allowing always-on microphones into people’s pockets and homes amounts to a further erosion of traditional expectations of privacy. The same might be said for all the ways in which language software improves by training on a single user’s voice, vocabulary, written documents and habits.

All the big companies’ location-based services—even the accelerometers in phones that detect small movements—are making ever-improving guesses about users’ wants and needs. The moment when a digital assistant surprises a user with “The chemist is nearby—do you want to buy more haemorrhoid cream, Steve?” could be when many may choose to reassess the trade-off between amazing new services and old-fashioned privacy. The tech companies can help by giving users more choice; the latest iPhone will not be activated when it is laid face down on a table. But hackers will inevitably find ways to get at some of these data.

Hey, Siri, find me a job

The other big concern is for jobs. To the extent that they are routine, they face being automated away. A good example is customer support. When people contact a company for help, the initial encounter is usually highly scripted. A company employee will verify a customer’s identity and follow a decision-tree. Language technology is now mature enough to take on many of these tasks.

For a long transition period humans will still be needed, but the work they do will become less routine. Nuance, which sells lots of automated online and phone-based help systems, is bullish on voice biometrics (customers identifying themselves by saying “my voice is my password”). Using around 200 parameters for identifying a speaker, it is probably more secure than a fingerprint, says Brett Beranek, a senior manager at the company. It will also eliminate the tedium, for both customers and support workers, of going through multi-step identification procedures with PINs, passwords and security questions. When Barclays, a British bank, offered it to frequent users of customer-support services, 84% signed up within five months.

Digital assistants on personal smartphones can get away with mistakes, but for some business applications the tolerance for error is close to zero, notes Nikita Ivanov. His company, Datalingvo, a Silicon Valley startup, answers questions phrased in natural language about a company’s business data. If a user wants to know which online ads resulted in the most sales in California last month, the software automatically translates his typed question into a database query. But behind the scenes a human working for Datalingvo vets the queryto make sure it is correct. This is because the stakes are high: the technology is bound to make mistakes in its early days, and users could make decisions based on bad data.

This process can work the other way round, too: rather than natural-language input producing data, data can produce language. Arria, a company based in London, makes software into which a spreadsheet full of data can be dragged and dropped, to be turned automatically into a written description of the contents, complete with trends. Matt Gould, the company’s chief strategy officer, likes to think that this will free chief financial officers from having to write up the same old routine analyses for the board, giving them time to develop more creative approaches.

Carl Benedikt Frey, an economist at Oxford University, has researched the likely effect of artificial intelligence on the labour market and concluded that the jobs most likely to remain immune include those requiring creativity and skill at complex social interactions. But not every human has those traits. Call centres may need fewer people as more routine work is handled by automated systems, but the trickier inquiries will still go to humans.

Much of this seems familiar. When Google search first became available, it turned up documents in seconds that would have taken a human operator hours, days or years to find. This removed much of the drudgery from being a researcher, librarian or journalist. More recently, young lawyers and paralegals have taken to using e-discovery. These innovations have not destroyed the professions concerned but merely reshaped them.

Machines that relieve drudgery and allow people to do more interesting jobs are a fine thing. In net terms they may even create extra jobs. But any big adjustment is most painful for those least able to adapt. Upheavals brought about by social changes—like the emancipation of women or the globalisation of labour markets—are already hard for some people to bear. When those changes are wrought by machines, they become even harder, and all the more so when those machines seem to behave more and more like humans. People already treat inanimate objects as if they were alive: who has never shouted at a computer in frustration? The more that machines talk, and the more that they seem to understand people, the more their users will be tempted to attribute human traits to them.

That raises questions about what it means to be human. Language is widely seen as humankind’s most distinguishing trait. AI researchers insist that their machines do not think like people, but if they can listen and talk like humans, what does that make them? As humans teach ever more capable machines to use language, the once-obvious line between them will blur.