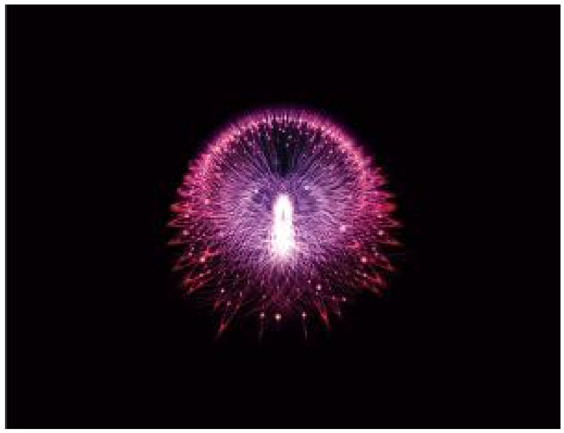
Not like us: Artificial minds we can't understand

08 August 2013 by **Douglas Heaven**
Magazine issue 2929. **Subscribe and save**
For similar stories, visit the **Robots** and **The Human Brain** Topic Guides

We have created a completely new form of intelligence, though no human can fathom how it thinks and reasons



It can see things we miss and knows us better than we know ourselves *(Image: Robert Hodgin)*

1 more image

Rick Rashid was understandably nervous. As he stepped onto the stage to address 2000 researchers and students in Tianjin, China, he was risking ridicule. He didn't speak Chinese, and his translator's poor skills in the past promised embarrassment.

"We hope that in a few years we'll be able to break down the language barriers between people," the senior vice-president of Microsoft Research told the audience. There was a tense 2-second pause before the translator's voice came through the speakers. Rashid continued: "Personally, I believe this is going to lead to a better world." Pause, repeat in Chinese.

He smiled. The crowd were applauding every line. Some people even cried.

This seemingly overenthusiastic reaction was understandable: Rashid's translator had come far. Every sentence was understood and delivered flawlessly. And the most impressive part? The translator was not human.

Performing such a task was once far beyond the abilities of the most sophisticated artificial intelligence, and not for want of effort. For years, AI was dominated by grand plans to replicate the performance of the human mind. We dreamed of machines that could understand us, recognise us and help us make decisions. In the last few years we have achieved those goals. But not in the way the pioneers imagined.

So have we worked out how to replicate human thinking? Far from it. Instead, the founding vision has taken a radically different form. AI is all around you, and its success is down to big data and statistics: making complex calculations using huge quantities of information. We have built minds, but they are not like ours. Their reasoning is unfathomable to humans – and the implications of this development are now attracting concern. As we come to rely more and more on this new form of intelligence, we may need to change our own thinking to accommodate it.

More than half a century ago, researchers laid out a series of goals that would bring us closer to machines with human-like intelligence. "We had a shopping list of things we wanted to do from the 50s," says Nello Cristianini at the University of Bristol, UK, who has written about the history and evolution of AI research.

Many items on the list can be traced back to the Mechanisation of Thought Processes conference in 1958 in Teddington, UK, which brought together not just computer scientists but physicists, biologists and psychologists too, all excited by the prospect of building a thinking machine in our image. The supposed hallmarks of intelligence they agreed on included understanding speech, language translation, image recognition and replicating human decision-making abilities.

But time passed and the shopping list got no shorter. Many researchers tried to emulate human thinking with programmed rules, rooted in logical axioms. Create enough rules, they figured, and success would follow. It proved too hard. Decades later, with little to show, AI funding dried up.

So what changed? "We haven't found the solution to intelligence," says Cristianini. "We kind of gave up." But that was the breakthrough. "As soon as we gave up the attempt to produce mental, psychological qualities we started finding success," he says.

Specifically, they jettisoned preprogrammed rules and embraced machine learning. With this technique, computers teach themselves to build patterns from data. With sufficiently large volumes of information you can get a machine to learn to do things that appear intelligent, be it understanding voices, translating language or recognising faces. "When you pile up enough bricks and stand back, you see a house," says Chris Bishop at Microsoft Research in Cambridge, UK.

Here, roughly, is how it works. Many of the most successful machine-learning systems are built on Bayesian statistics, a mathematical framework that lets us measure likelihood. It puts a number to the plausibility of an outcome given the context and previously observed correlations in similar contexts.

Let's say we want an AI to answer questions about a simple topic: what cats like to eat, for instance. The rule-based approach is to build, from scratch, a database about cats and their dietary habits, with logical steps (see diagram). With machine learning, you instead feed in data indiscriminately – internet searches, social media, recipe books and more. After doing things like counting the frequency of certain words and how concepts relate to one another, the system builds a statistical model that gauges the likelihood of cats enjoying certain foods.

Of course, the algorithms underpinning machine learning have been around for years. What's new is that we now have enough data for the techniques to gain traction.

Take language translation. In the late 20th century, IBM used machine learning to teach a computer to translate between English and French by feeding it bilingual documents produced by the Canadian parliament. Like a Rosetta stone, the documents contained several million examples of sentences translated into both languages.

IBM's system spotted correlations between words and phrases in the two languages and reused them for fresh translation. But the results were still full of errors. They needed more data. "Then Google comes along and basically feeds in the entire internet," says Viktor Mayer-Schönberger of the Oxford Internet Institute at the University of Oxford.

Like IBM, Google's efforts in translation started by training algorithms to cross-reference documents written in many languages. But the realisation dawned that the translator's results would improve significantly if it learnt how people speaking Russian, French or Korean actually conversed.

Google turned to the vast web it has indexed, which is fast approaching the fantastical library imagined by Jorge Luis Borges in his 1941 short story *The Library of Babel*; it contained books with every combination of words it is possible to have. Google's translator – attempting English to French, for instance – could then compare its initial attempt with every sentence written on the internet in French. Mayer-Schönberger gives the example of choosing whether to translate the English "light" with the French "lumière", referring to illumination, or "léger", for weight. Google has taught itself what the French themselves choose.

Google's translator – along with the Microsoft one used by Rashid – knows nothing about language at all, other than the relative frequency of a vast number of word sequences. Word by word, these AIs simply calculate the likelihood of what comes next. For them it is just a matter of probabilities.

These basics are more or less intuitive. The complexity arises from the vast numbers of correlations made within enormous amounts of data. Google's self-driving car, for example, gathers almost a gigabyte of data each second to make predictions about its surroundings. And Amazon is so good at getting people to buy more because it recommends items based on billions of correlations from millions of other purchases.

Too big to fail

The translation of Rashid's speech – guessing what he had said, what the translation should be and how his voice would say that in Chinese – shows just how powerful statistical AI can be. "These systems don't perform miracles," says Bishop. "But we're constantly surprised how far we can get just by looking at the statistics of very large sets of data."

These intelligent algorithms are beginning to influence every realm of life. A month after Rashid's speech, for example, the Netherlands Forensic Institute in The Hague employed a machine-learning system called Bonaparte to help find a murder suspect who had evaded capture for 13 years. Bonaparte can analyse and compare large volumes of DNA samples; something that would be far too time-consuming to do by hand. The insurance and credit industries are also embracing machine learning, employing algorithms to build risk profiles of individuals. Medicine, too, uses statistical AI to sift through genetic data sets too large for humans to analyse. IBM's Watson even performs diagnoses.

"Big data analysis can see things that we miss," says Mayer-Schönberger. "It knows us better than we can know ourselves. But it also requires a very different way of thinking."

In the early days of AI, "explainability" was prized. When a machine made a choice, a human could trace why. Yet the reasoning made by a data-driven artificial mind today is a massively complex statistical analysis of an immense number of data points. It means we have traded "why" for simply "what".

Even if a skilled technician could follow the maths, it might not be meaningful. It wouldn't reveal why it made a decision, because it wasn't arrived at by a set of rules that a human can interpret, says Bishop, who thinks this is an acceptable trade-off for systems that work. Early artificial minds may have been transparent, but they failed. "You've got an explanation, but it's an explanation of a wrong prediction," he says. Some have criticised this shift (see "Chomsky vs Google"), but Bishop and others argue that it is time to give up on expecting human explanations.

"Explainability is a social agreement," says Cristianini. "We decided in the past it mattered. We've decided now it doesn't matter."

Peter Flach at the University of Bristol, UK, tries to teach his computer science students this fundamentally different way of thinking. Programming is about absolutes, but machine learning is about degrees of uncertainty. He thinks we should be more sceptical. When Amazon's AI recommends a book, he says, is that because of machine learning or because the company has books it can't shift? And while Amazon might tell you, for example, that similar people bought the book choices it presents, what does it actually mean by "people like you" and "books like this"?

"It may be at some level there's always going to be machinery that we have to trust even if we don't fully understand it," says Flach.

The danger is that we give up asking questions. Could we get so used to choices being made for us that we stop noticing? The stakes are higher now that intelligent machines are beginning to make inscrutable decisions about mortgage applications, medical diagnoses and even whether you are guilty of a crime.

In medicine, for instance, what if a machine learning system decides that you will start drinking heavily in a few years' time? Would doctors be justified in withholding a transplant? It'd be hard to argue your case if no one knew how the conclusion was arrived at. And some may trust the AI more than others. "People are too willing to accept something that an algorithm has found out," says Flach. "The computer says 'No'. That's the issue."

There could be an intelligent system somewhere making its mind up right now about what kind of person you are – and will be. Consider what happened to Latanya Sweeney at Harvard University. One day, she was surprised to find that her Google search results were accompanied by adverts asking "Have you ever been arrested?" The ads did not appear for white colleagues. This prompted a study showing that the machine learning behind Google's search was inadvertently racist. Deep within the chaos of correlations, names more commonly given to black people were linked to ads about arrest records.

"There are profound ethical dilemmas," says Mayer-Schönberger. Many people have expressed concerns about privacy in an age of big data. "To be honest, I'm less worried about privacy and more worried about the abuse of probabilistic prediction," he says.

To navigate this world, we will need to change our ideas about what artificial intelligence means. The iconic intelligent systems we have built do not play chess, nor plot humanity's downfall. "They are not like HAL 9000," says Cristianini. They have gone from chaperoning our time online and nudging us towards a purchase to promising to predict our behaviour before we know it ourselves. We cannot avoid them. So the trick will be to accept that we cannot know why these choices were made, and to recognise the choices for what they are: recommendations, mathematical possibilities. There is no oracle behind them.

When people dreamed of making AI in our image, they may have looked forward to meeting these thinking machines as equals. The AI we've ended up with is alien – a form of intelligence we've never encountered before.

This article appeared in print under the headline "Higher state of mind"

Chomsky vs Google

Do we need to understand the artificial intelligence we create? The question sparked an unlikely tussle between two intellectual heavyweights from quite different realms.

During MIT's 150th birthday party, Noam Chomsky, the father of modern linguistics, was asked to comment on the success of statistical methods for producing artificial intelligence. It turned out Chomsky is not a fan.

Chomsky's work on language has influenced many who study human intelligence. At the heart of his theories is the idea that our brains essentially have hard-wired rules. That might help explain why he disapproves of the modern approach to AI, which has thrown out rules and replaced them with statistical correlations. In essence, it means we don't know why these AIs are intelligent, they just are (see main story).

For Chomsky, proponents of statistical techniques are like scientists studying bee dances who produce an accurate simulation of bee movements without asking why the bees do it. Chomsky's point is that statistical techniques provide predictions but not understanding. "That's a notion of success that is very novel. I don't know of anything like it in the history of science," he said.

Peter Norvig, the head of research at Google, shot back at Chomsky in an essay on his website. He bristles at Chomsky's comment that the statistical approach has had "limited success". On the contrary, it is now the dominant paradigm, Norvig writes, and generates several trillion dollars of revenue per year. In the academic equivalent of a diss, he compared Chomsky's views to mysticism and to comments made by Fox News presenter

Bill O'Reilly, who once fallaciously criticised science for failing to explain why Earth's tides exist.

Yet Norvig's main disagreement was more fundamental. In short, he argues that scientists such as Chomsky who seek to build ever-simpler and more elegant models to explain the world are outdated. "Nature's black box cannot necessarily be described by a simple model," he says. Norvig's point is that Chomsky's approach provides an illusion of understanding, but isn't rooted in reality.

What started as a debate about AI, then, appears to be more about the nature of knowledge itself.

***Douglas Heaven** is a writer based in London*

From issue 2929 of New Scientist magazine, page 32-35.