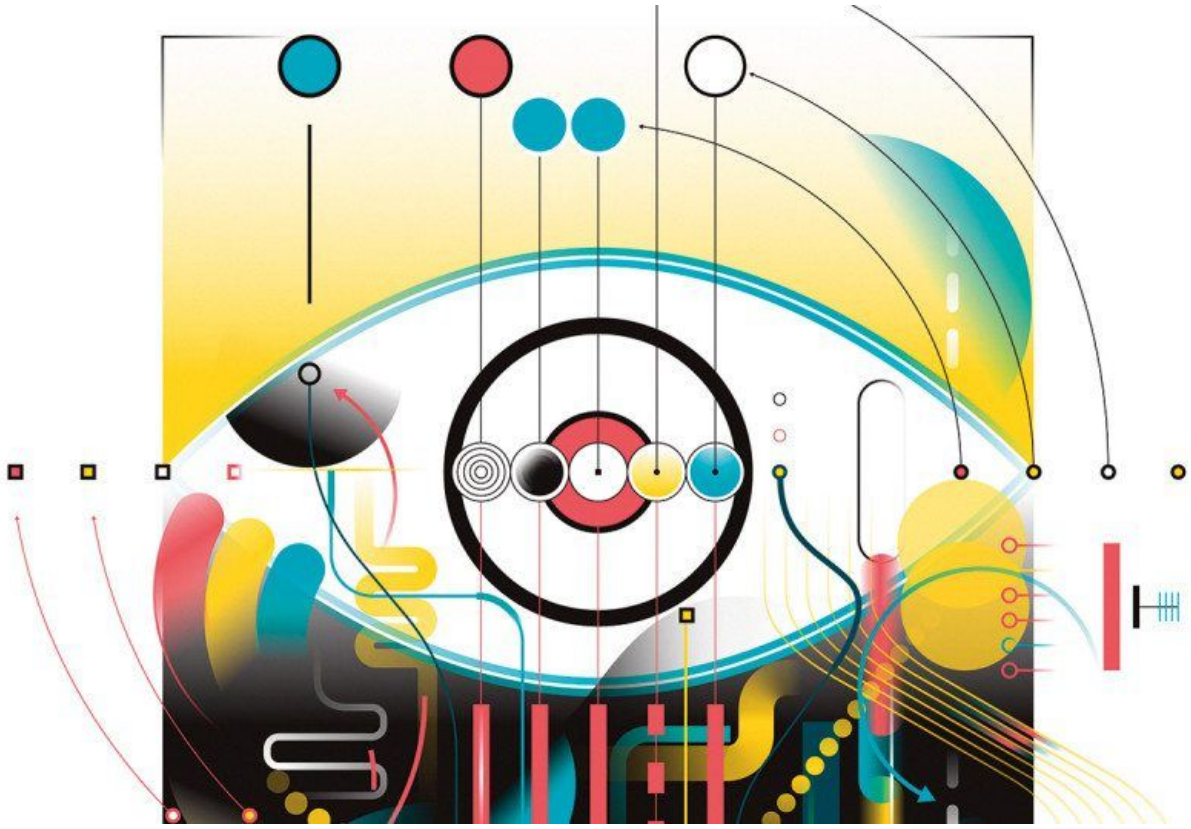
FEATURE 26 October 2016

The road to artificial intelligence: A case of data over theory

Computers that could simulate human intelligence were once a futuristic dream. Now they are all around us – but not in the way their pioneers expected



Leandro Castelao

By **Nello Cristianini**

IN the summer of 1956, a remarkable collection of scientists and engineers gathered at Dartmouth College in Hanover, New Hampshire. Among them were computer scientist Marvin Minsky, information theorist Claude Shannon and two future Nobel prizewinners, Herbert Simon and John Nash. Their task: to spend the summer months inventing a new field of science called "artificial intelligence" (AI).

They did not lack in ambition, writing in their funding application: "every aspect of

learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it." Their wish list was "to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves". They thought that "a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer."

It took rather longer than a summer, but 60 years and many disappointments later, the field of AI seems to have finally found its way. In 2016, we can ask a computer questions, sit back while semi-autonomous cars negotiate traffic, and use smartphones to translate speech or printed text across most languages. We trust computers to check passports, screen our correspondence and fix our spelling. Even more remarkably, we have become so used to these tools working that we complain when they fail.

As we rapidly get used to this convenience, it is easy to forget that AI hasn't always been this way.

At the Dartmouth conference, and at various meetings that followed it, the defining goals for the field were already clear: machine translation, computer vision, text understanding, speech recognition, control of robots and machine learning. For the following three decades, significant resources were ploughed into research, but none of the goals were achieved. It was not until the late 1990s that many of the advances predicted in 1956 started to happen. But before this wave of success, the field had to learn an important and humbling lesson.

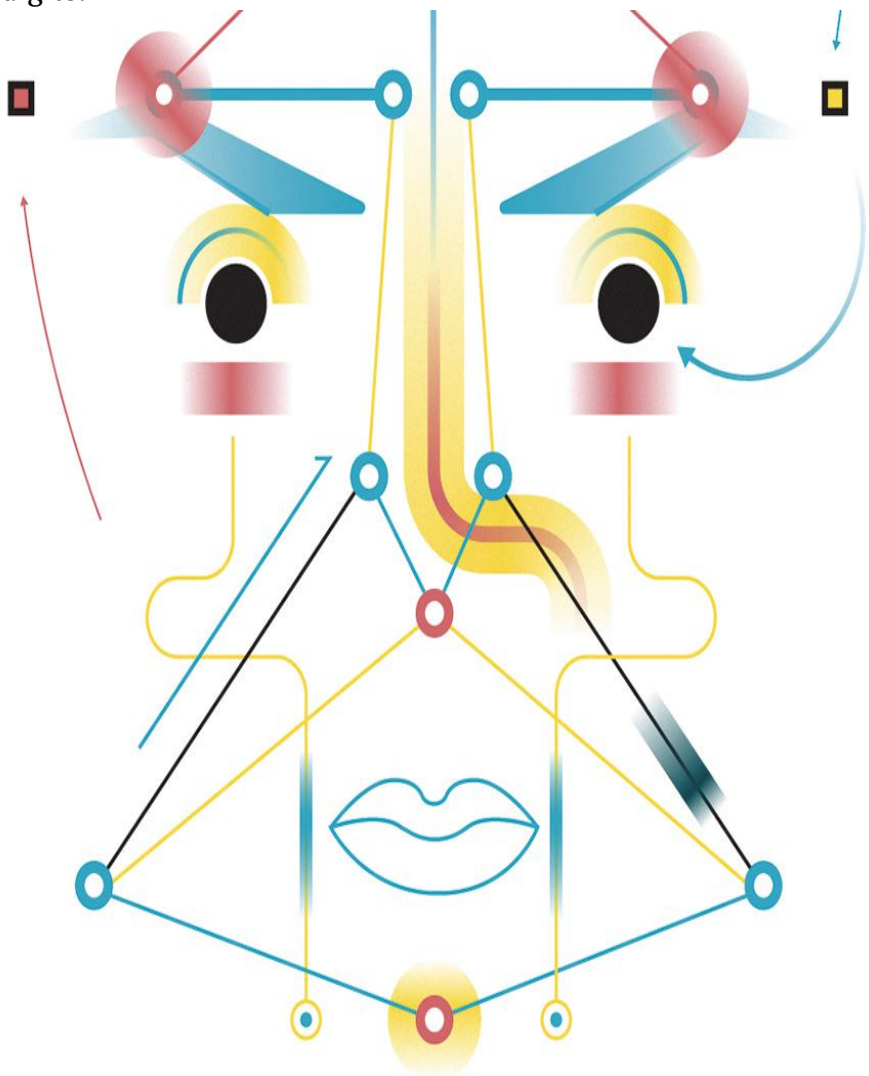
While its goals have remained essentially the same, the methods of creating AI have changed dramatically. The instinct of those early engineers was to program machines from the top down. They expected to generate intelligent behaviour by first creating a mathematical model of how we might process speech, text or images, and then by implementing that model in the form of a computer program, perhaps one that would reason logically about those tasks. They were proven wrong.

They also expected that any breakthrough in AI would provide us with further understanding about our own intelligence. Wrong again.

Over the years, it became increasingly clear that those systems weren't suited to dealing with the messiness of the real world. By the early 1990s, with little to show for decades of work, most engineers started abandoning the dream of a general-purpose top-down reasoning machine. They started looking at humbler projects, focusing on specific tasks that were more likely to be solved.

Some early success came in systems to recommend products. While it can be difficult to know why a customer might want to buy an item, it can be easy to know which item they might like on the basis of previous transactions by themselves or similar customers. If you liked the first and second Harry Potter films, you might like the third. A full understanding of the problem was not required for a solution: you could detect useful correlations just by combing through a lot of data.

Could similar bottom-up shortcuts emulate other forms of intelligent behaviour? After all, there were many other problems in AI where no theory existed, but there was plenty of data to analyse. This pragmatic attitude produced success in speech recognition, machine translation and simple computer vision tasks such as recognising handwritten digits.



Data beats theory

By the mid-2000s, with success stories piling up, the field had learned a powerful lesson: data can be stronger than theoretical models. A new generation of intelligent machines had emerged, powered by a small set of statistical learning algorithms and large amounts of data.

Researchers also ditched the assumption that AI would provide us with further understanding of our own intelligence. Try to learn from algorithms how humans perform those tasks, and you are wasting your time: the intelligence is more in the data than in the algorithm.

The field had undergone a paradigm shift and had entered the age of data-driven AI. Its new core technology was machine learning, and its language was no longer that of logic, but statistics.

How, then, can a machine learn? It is worth clarifying here what we normally mean by learning in AI: a machine learns when it changes its behaviour (hopefully for the better) based on experience. It sounds almost magical, but in reality the process is quite

mechanical.

Consider how the spam filter in your mailbox decides to quarantine some emails on the basis of their content. Every time you drag an email into the spam folder, you enable it to estimate the probability that messages from a given recipient or containing a given word are unwanted. Combining this information for all the words in a message allows it to make an educated guess about new emails. No deep understanding is required – just counting the frequencies of words.

But when these ideas are applied on a very large scale, something surprising seems to happen: machines start doing things that would be difficult to program directly, like being able to complete sentences, predict our next click, or recommend a product. Taken to its extreme conclusion, this approach has delivered language translation, handwriting recognition, face recognition and more. Contrary to the assumptions of 60 years ago, we don't need to precisely describe a feature of intelligence for a machine to simulate it.

While each of these mechanisms is simple enough that we might call it a statistical hack, when we deploy many of them simultaneously in complex software, and feed them with millions of examples, the result might look like highly adaptive behaviour that feels intelligent to us. Yet, remarkably, the agent has no internal representation of why it does what it does.

This experimental finding is sometimes called "the unreasonable effectiveness of data". It has been a very humbling and important lesson for AI researchers: that simple statistical tricks, combined with vast amounts of data, have delivered the kind of behaviour that had eluded its best theoreticians for decades.

Thanks to machine learning and the availability of vast data sets, AI has finally been able to produce usable vision, speech, translation and question-answering systems. Integrated into larger systems, those can power products and services ranging from Siri and Amazon to the Google car.

Researchers' attention is now focused what fuels the engine of our intelligent machines: data. Where can they find data, and how can they make the most of this resource?

One important step has been to recognise that valuable data can be found freely "in the wild", generated as a byproduct of various activities – some as mundane as sharing a tweet or adding a smiley under a blog post.

Engineers and entrepreneurs have also invented a variety of ways to elicit and collect additional data, such as asking users to accept a cookie, tag friends in images, rate a product or play a location-based game centred on finding monsters in the street. Data became "the new oil".

At the same time as AI was finding its way, we developed an unprecedented global data infrastructure. Every time you access the internet to read the news, do a search, buy something, play a game, or check your email, bank balance or social media feed, you interact with this infrastructure. It isn't just a physical one of computers and wires, but also one of software, including social networks and microblogging sites.

Data-driven AI both feeds on this infrastructure and powers it – it is hard to imagine

one without the other. And it is hard to imagine life without either of them.

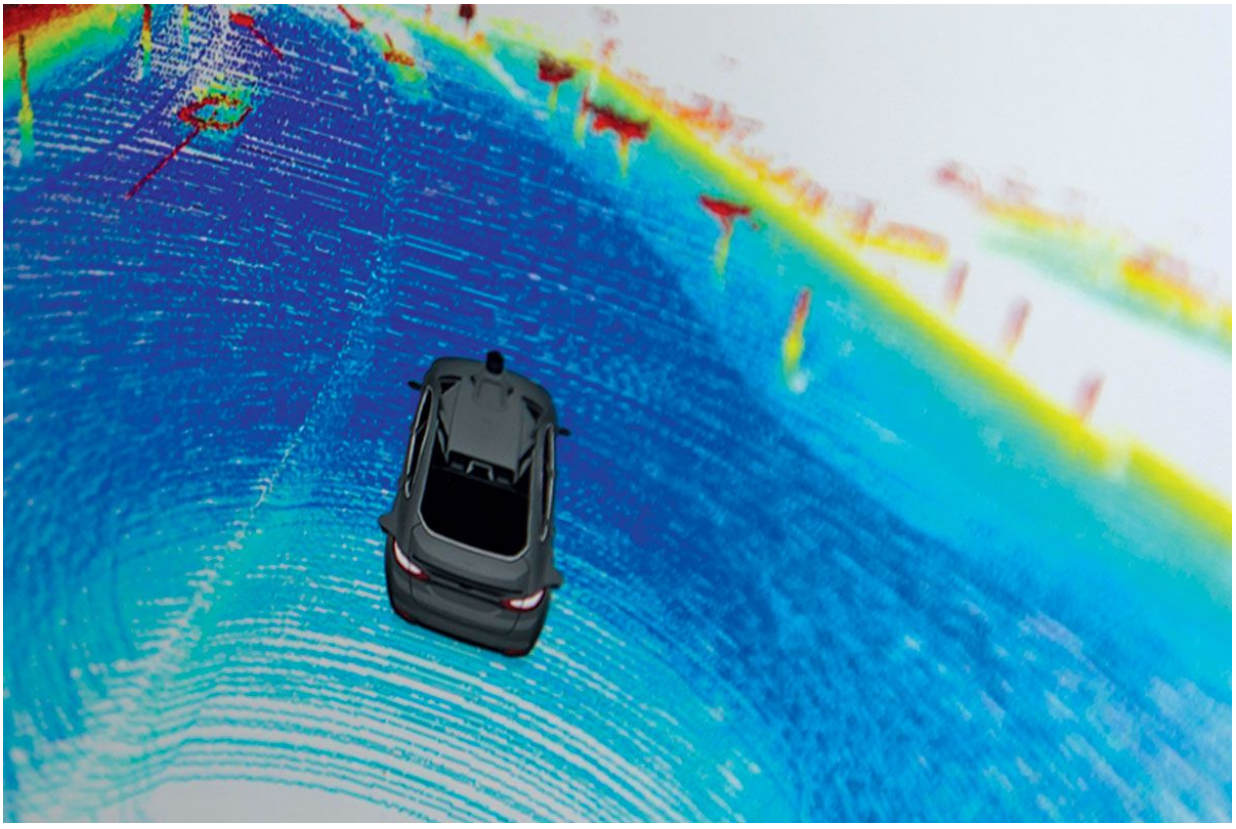
Challenges ahead

This is what makes modern AI a brilliant and powerful technology, but also a fundamentally disruptive one.

The unified data infrastructure is not like any medium invented before. Unlike the copper cables that used to connect people in the telegraph or telephone age, it takes a keen interest in our actions. The medium looks back at us, anticipating our moves, guessing our intents, often trying to serve us better and sometimes to influence us. This gives a whole new meaning to the claim, made by the 1970s communications theorist Marshall McLuhan, that a medium can never be neutral.

The challenges AI might present us with include surveillance, discrimination, persuasion, unemployment and possibly even addiction. Are we prepared?

Intelligent machines need to collect data – often personal data – in order to work. This simple fact potentially turns them into surveillance devices: they know our location, our browsing history and our social networks. Can we decide who has access, what use can be made of the data, or whether the data gets deleted for ever? If the answer is no, then we don't have control.



Driverless cars "see" better than ever, but mistakes can't be ruled out
Jeff Swensen/The New York Times/Redux / eyevine

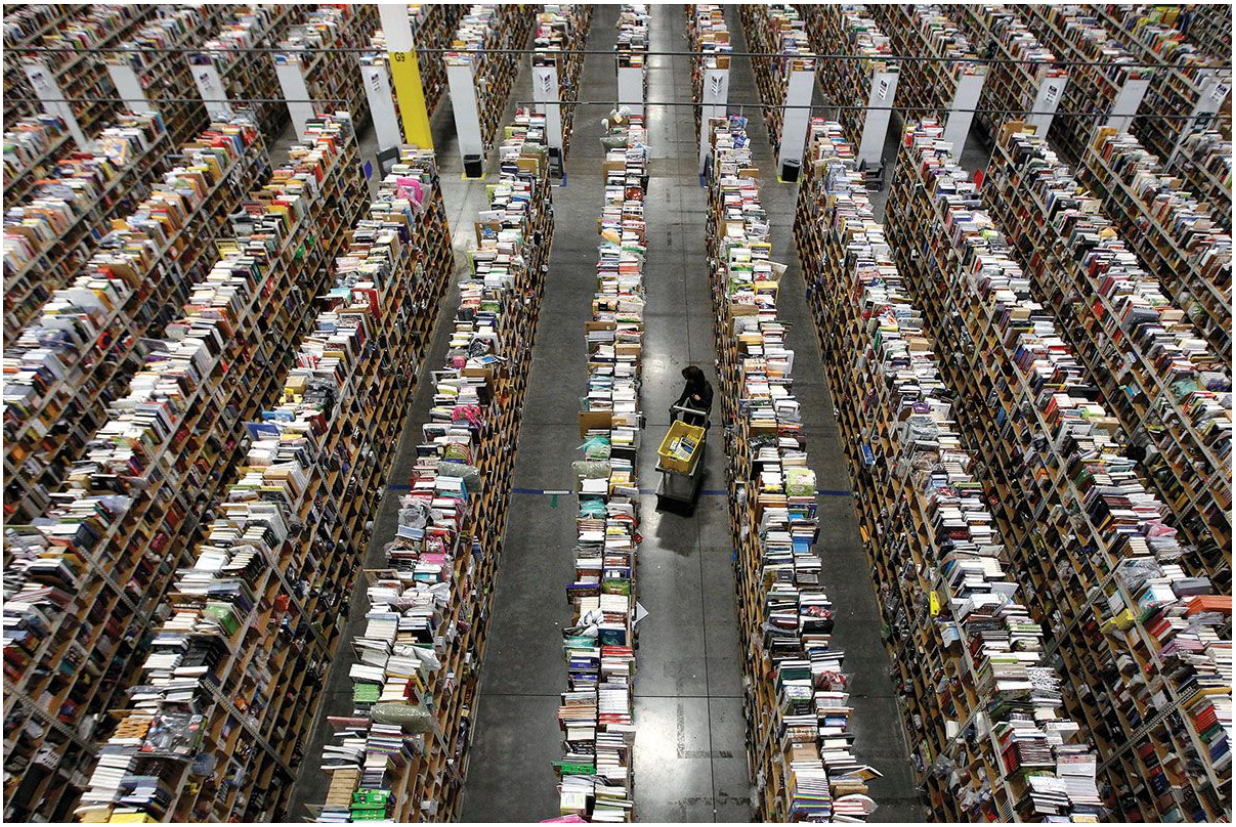
AI's capability to make predictions is useful for insurance, loans and policing. But the quality of those predictions will depend on subtle design choices and on the way the information used to train it is collected, which creates a very real risk of implicit and unintended discrimination. A recent investigation by ProPublica, for example, claims to have uncovered a bias that would disadvantage African Americans in the software used in many US courts to make parole decisions. Another case has been reported where

different job ads were targeted at different ethnic groups. Both starkly illustrate the unintended effects of the complex interaction between algorithms and data.

Another concern is persuasion. The business model of many AI companies is advertising, which means getting people to click on specific links. Research on how to steer users is well under way. The more the machines know about us, the better the job they can do of nudging us. Predictive interfaces might even induce addiction in vulnerable users, by actively rewarding them with the juiciest content that the web has to offer. This is something that needs to be carefully studied.

Employment will be affected too, as AIs learn from us (quite literally) how to do certain jobs, either because they watch how we do them, or because we are paid to generate their training data.

The emergence of internet crowdsourcing allows businesses to automatically outsource micro-tasks that require human intelligence, by posting them on websites or apps where workers can choose the tasks they want to accept. In a way this works just like Uber, but for tasks other than driving, and is mediated by a computer system. Typical tasks would include transcribing handwriting or labelling images.



AI will increase the automation of warehouses like Amazon's
Ralph Freso / Reuters

This also creates a workforce directly managed via computers, and defines a set of tasks that are the ideal candidate for automation. Indeed, many of those task-workers are actually generating or annotating the data being used to train their AI replacements.

At the same time, we can expect many call centres and warehouses to be increasingly automated within a decade.

I do not believe that we yet have the legal and cultural tools to handle these and many other challenges. Who do we turn to if an intelligent algorithm denies us parole,

medical treatment or a diploma? Are we prepared for our character and trustworthiness to be ranked just like our credit history, as some countries are proposing? Do we want the state to have access to our online activities and knowledge of our preferences? Do we want our children to spend their online time in the company of persuasive machines, designed to steer their behaviour in a given direction? What happens to society if large numbers of people are put out of work?

Artificial intelligence has come a long way from its early days in academic laboratories. It is now being integrated into our lives, and promises to improve them. We might not call it AI once it is deployed, but we can expect benefits in fields ranging from healthcare to transportation, from communications to schooling.

And research is not slowing down. The machine-learning paradigm has been effective in addressing many areas like vision and speech processing, and it is likely that future AI will also find a way to integrate some top-down reasoning methods descended from earlier approaches. What will come after that may surprise us again.

As our AI efforts continue to open up new possibilities, we can imagine seamless conversations with machines, fluent real-time translation of speech, and many useful ways to automate our houses and cars.

But we might want to resist the temptation to introduce AI into as many domains as possible, at least before the cultural and legal framework evolves. Widespread adoption of AI brings remarkable opportunities, but also potential risks. Contrary to popular belief these are not existential risks to our species, but rather a possible erosion of our privacy and autonomy.

So as we finally enjoy the benefits of six decades of research in AI, with machines joining us in our everyday lives, we should celebrate – but also tread carefully.

The winters of AI discontent

Emergent technologies are often subjected to hype cycles, sometimes due to speculative bubbles inflated by excessive investor expectations. Some examples are railway mania in the UK in the 1840s and the dot-com bubble in the 1990s.

Artificial intelligence is perhaps unique in having undergone several hype cycles in a relatively short time. Its slumps of optimism even have a specific name: AI winters (see timeline below). The two major winters occurred in the early 1970s and late 1980s. Both were caused largely by the withdrawal of public funding as progress stalled.

AI is now in a renewed phase of heightened optimism and investment. Unlike in previous cycles, however, AI today has a strong – and increasingly diversified – commercial revenue stream. Only time will tell whether this turns out to be a bubble.

1950 Alan Turing publishes the seminal paper "Computing machinery and intelligence". Its opening sentence is "I propose to consider the question, 'Can machines think?'"

1956 The term "artificial intelligence" is coined at a workshop at Dartmouth College

1959 Computer scientists at Carnegie Mellon University create the General Problem Solver (GPS), a program that can solve logic puzzles

1973 The first AI winter sets in as funding and interest dry up

1975 A system called MYCIN diagnoses bacterial infections and recommends antibiotics using deduction based on a series of yes/no questions. It was never used in practice

1987 Second AI winter begins

1989 NASA's AutoClass computer program discovers several previously unknown classes of stars

1994 First web search engines launched

1997 IBM's Deep Blue beats world champion Garry Kasparov at chess

1998 NASA's Remote Agent is first fully autonomous program to control a spacecraft in flight

2002 Amazon replaces human product recommendation editors with an automated system

2007 Google launches Translate, a statistical machine translation service

2009 Google researchers publish an influential paper called "The unreasonable effectiveness of data". It declares that "simple models and a lot of data trump more elaborate models based on less data" (**IEEE Intelligent Systems**, vol 2, p 8)

2011 Apple releases Siri, a voice-operated personal assistant that can answer questions, make recommendations and carry out instructions such as "call home"

2011 IBM's supercomputer Watson beats two human champions at TV quiz game **Jeopardy!**

2012 Google's driverless cars navigate autonomously through traffic

2016 Google's AlphaGo defeats Lee Sedol, one of the world's leading Go players

You win some…

Ben Hider/Getty

One of the most celebrated successes of machine learning (see main story) came earlier this year when an algorithm called AlphaGo defeated South Korean master Lee Sedol at the game Go – something none of its programmers could come close to doing themselves. AlphaGo combined various machine-learning methodologies to analyse databases of more than 30 million Go moves, as well as playing thousands of games against itself. A similar strategy earlier allowed IBM's Watson supercomputer to win at the TV quiz game **Jeopardy!** (pictured above).

Given the right data, it seems that machines can improve their intelligence a great deal. But we should remember that machine learning is a statistical exercise, and therefore it can always fail.

In recent years we have also seen some blunders caused by machine learning. Last year Google apologised after one of its products automatically labelled photos of two black people "gorillas"; this year Microsoft had to withdraw a conversational bot called Tay because it had learned offensive language. In both cases it was not a failure of the algorithm, but of the training data that had been fed to it.

This year also saw the first fatality linked to a "driverless" car, when a driver put a Tesla on autopilot and it failed to detect a trailer on the road. The conditions were unusual, with a white obstacle against a light sky, and the computer vision system simply made a mistake. I do not expect it to be the last one as many companies move into that market.

On the other hand, there are countless stories that do not end up in the news, simply because the AI systems are doing their work as expected. They include search engines, online shops and semi-autonomous cars.

This article appeared in print under the headline "Intelligence reinvented"

Nello Cristianini is professor of artificial intelligence at the University of Bristol, UK

Magazine issue 3097, published 29 October 2016