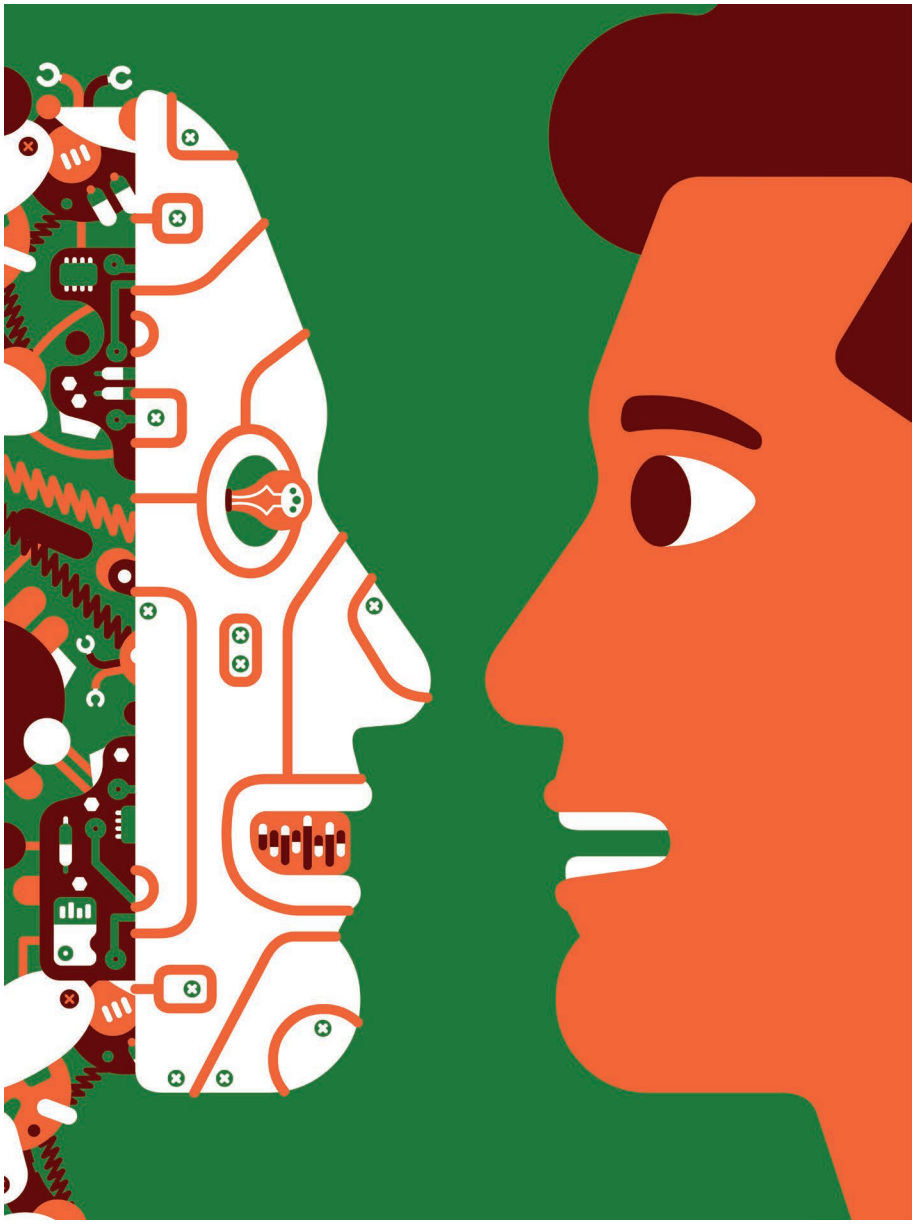
JAN KALLWEJT



AUTOMATION

Meeting people's expectations

The public's view of artificial intelligence might not be accurate, but that doesn't mean researchers can ignore it.

BY JEFF HECHT

When all mail used to arrive by post, sorting out the junk was a minor chore. But when mail began to be sent digitally, computers started churning out so much junk e-mail that it was necessary to automate the act of identifying the real mail — the 'ham', in Internet jargon — from the torrents of spam. But clever spammers soon found ways to avoid the spam filters, so e-mail providers turned to artificial intelligence (AI) to outsmart them. These days, AI systems are trained to tell spam from ham by being fed enormous amounts of both, and they defend your inbox without you having to lift a finger.

This is one way in which AI has slipped into our daily lives almost without us noticing. In the early 2000s, there were fears that spam was on the verge of killing e-mail. Machine learning saved the day because it excels at recognizing the kinds of pattern that are inevitable in mass mailings. It is not infallible; databases must be updated to keep up with spammers' new tricks, and users still have to occasionally check that the spam filter hasn't swallowed something important. But machine-learning systems are the best tools we have.

Our inbox is not the only place where we can unknowingly encounter AI. Machine learning is used by streaming services to recommend which film we should watch next; it tries to nudge us towards our next likely online purchase; and it identifies people and objects in our photos so that we can find them again more easily. Many people now engage computers in spoken conversation on a daily basis. And Google's AlphaGo program used machine learning to beat human masters of the 3,000-year-old board game Go, to widespread media acclaim.

But as the list of applications grows, so too does our awareness of AI in our lives. As a result, technologists pushing for the next big thing in automation now face sterner questions about what the public really wants. A century of science fiction has coloured public opinion of AI to such an extent that a common reaction to its growing prominence is one of fear. Now the small group of companies that are investing billions of dollars in using machine learning to drive cars — a much more ambitious and risky task than simply recognizing patterns in e-mails — find themselves having to address the question of how to deal with the public's perception of AI.

A DIFFICULT JOURNEY

The automotive industry is one of the hottest areas for the use of AI, but it has experienced some difficult times. High oil prices in the late 2000s hurt sales of highly profitable SUVs and trucks, which had been a mainstay for US car-makers in particular. In the wake of the 2008 financial crisis, General Motors and Chrysler (now Fiat Chrysler Automobiles) needed an US$80-billion bailout from the US government. Now, rapid advances in AI and the development of improved sensing systems are offering the chance to introduce something dramatic and new. Many established car manufacturers see self-driving vehicles as the next step in their recovery, and some of the biggest names in technology are now entering the automotive sector.

Advocates of autonomous vehicles say that the technology can make people's lives easier by taking the strain out of the rush-hour commute and by bringing vital mobility to people who are unable to drive. They also argue that replacing the human driver — "the nut behind

the wheel", as some people put it — with AI would greatly reduce the toll of road accidents, which kill more than 1 million people a year according to the World Health Organization.

The development of self-driving vehicles is proceeding rapidly. Test cars created by Waymo, which is a sister company to tech giant Google, have so far driven more than 10 million miles on public roads, and Waymo recently began the commercial operation of a fleet of driverless taxis in Phoenix, Arizona. General Motors plans to introduce its own 'robo-taxis' in 2019, and Ford is targeting 2021 for the launch of a similar service.

The automotive and technology industries think that automotive cars are ready to roll, but the public is not yet ready to ride. In April 2018, a survey from pollsters Gallup found that only 9% of US adults would use a driverless car as soon as government regulators certify them as safe. Another 38% said they would wait a while after the vehicles' introduction before getting on board. Meanwhile, 52% said they never want to use a self-driving vehicle.

The gulf between industry confidence and public caution is ominous for those who are developing the technology. The fear with which many people approach AI may at least partly reflect a history of AI experiments going wrong in unnerving ways. In 2016, for instance, a Microsoft chatbot that was designed to engage in playful conversation on the social-media site Twitter began spouting racist and misogynistic abuse just 16 hours after going live. The behaviour arose after some Twitter users decided to feed it such sentiments. The following year, researchers at the Massachusetts Institute of Technology (MIT) in Cambridge showed that shaping and patterning objects in specific ways could fool an image-recognition system developed by Google into classifying a 3D-printed turtle as a rifle, regardless of which angle it was viewed at. And Mike Sellers, an AI researcher and game designer at Indiana University Bloomington, recently recounted an experience in the early 2000s in which two experimental AI bots he created for the US defence agency DARPA decided that a third bot would be good to eat.

In some cases, the way in which AI experiments are reported can stoke unnecessary fear. A blog post in June 2017 from AI researchers at Facebook described their work in which two chatbots attempting to negotiate with one another began using code words in conversation. The post led to news stories that created the impression of researchers hurriedly pulling the plug on their experiment before it got out of control. Fears might also have been inflated by warnings from public figures such as entrepreneur Elon Musk that AI could soon become so powerful that it would rule the world.

HOW HUMAN?

Some of the fear of AI probably stems from the idea that machines and humans are not so very different. "If you see a machine beat the world champion at Go, you can think that the system is so much better than you that you have no chance," says Steve McGough, a data scientist at Newcastle University, UK. What many people don't understand, he says, is that those machines could not beat humans at anything else.

From the earliest speculation in fiction, machine intelligence was envisaged as being similar to human intelligence. One of the first science-fiction robots was a mechanical man created by writer L. Frank Baum in his sequels to *The Wonderful Wizard of Oz* in the early 1900s. Called Tik-Tok, it had no emotions, but in other respects it had human-like intelligence. And Isaac Asimov, a biochemist who was better known for his science fiction, based his 'three laws of robotics' in the 1942 story *Runaround* on the assumption that the machines would think like humans. The story went on to inspire many of the early pioneers of AI. After hearing of Asimov's death in 1992, Marvin Minsky, who co-founded an AI lab at MIT in the early 1960s, wrote that, after reading *Runaround*, he "never stopped thinking about how minds might work. Surely we'd someday build robots that think."

The first boom of AI in the 1950s was partly based on the work of Alan Turing, creator of the eponymous 'Turing test' in which a human is asked to judge whether they are communicating with another human or a machine. In the 1960s, "there was lots of work being done to get machines to do the kinds of things that humans could do", says Nils John Nilsson, who at the time was working on AI and neural networks at what is now the US non-profit organization SRI International in Menlo Park, California. But progress stalled, and by the late 1970s, funding had dwindled in what became known as the 'AI winter'.

All that changed in 1997 when IBM's Deep Blue computer defeated world chess champion Garry Kasparov. Deep Blue won through the sheer brute force of its massive computing power: it calculated every possible scenario in the three minutes it was allowed for each move. Since the turn of the millennium, that brute-force approach has been applied to the technology of machine learning, in which huge volumes of data are analysed to look for patterns. Depending on the data set used, this can allow AI systems to identify images or devise a winning strategy in a game that has seemingly endless permutations. But its mastery of any particular task is not transferable. AlphaGo beat the world champion at Go, but, McGough says, "if you changed one of the rules, the grand master would be able to take all their experience and use that to win". Victories at chess and Go might create the impression that computers are now our cognitive equals, "but we're a long way away from that", he says.

"The human brain can solve a problem it has never seen before. Nothing in machine learning is like that."

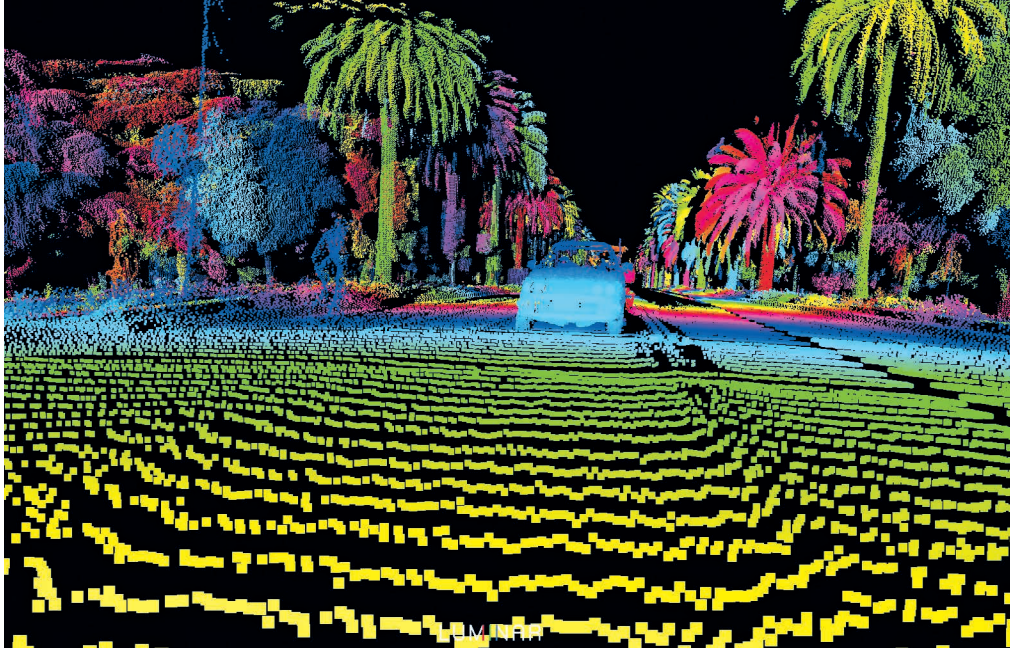
"The human brain can solve a problem it has never seen before," says Greg Hullender, a specialist in machine learning and natural language, who has been a research scientist for both Amazon and Microsoft. "Nothing in machine learning is like that. Everything is built for a specific problem." This crucial difference leads many AI specialists to scoff at both warnings of a robot takeover and claims of a future AI utopia. "Both are wild exaggerations," says Peter Stone, an AI researcher at the University of Texas at Austin.

CAR TROUBLE

In the case of autonomous vehicles, public resistance is fuelled by more than the fear of being overthrown by an artificial intelligence. In the United States, people who love to drive were aghast when advocates of self-driving cars argued that reserving the roads for robots would improve safety. Some are

KYODO NEWS/GETTY

Ride-hailing company Uber tested a fleet of self-driving Volvos in Pittsburgh, Pennsylvania, in March 2018.



Autonomous vehicles use laser radars called lidars to create a map of the world around them.

even organizing an opposition. In January 2018, endurance driver and writer Alex Ray launched the Human Driving Association to preserve humans' right to drive themselves. By October, some 4,000 drivers had joined the group. Car enthusiast and businessman McKeel Hagerty, whose company is the largest insurer of classic cars in the United States, believes he will need to attract 6 million members to his similar 'Save Driving' campaign in order to preserve human driving.

Crucially, however, people are worried about safety, despite statistics showing that human error is responsible for the majority of fatal car accidents. Most people think that their driving is better than average, so they worry less about their own driving than that of other people or robots. And self-driving cars have already been involved in accidents that have undermined public trust.

The first fatal accident came in 2016, when a Tesla enthusiast in Florida crashed into a truck while using the car's limited 'Autopilot' driver-assistance feature. Investigators found that the main causes were the Tesla driver ignoring repeated warnings to place his hands on the wheel, and the truck driver making an unsafe turn.

In March 2018, a crash in Mountain View, California, raised more-troubling questions. Running under Autopilot, a Tesla accelerated to above the speed limit and steered into a barrier, causing collisions with two other cars, according to a preliminary report by the US National Transportation Safety Board.

That crash came just a few days after the first fatality involving a fully self-driving car. A vehicle being tested by ride-hailing company Uber one night struck and killed a pedestrian who was walking a bicycle across a road in Tempe, Arizona. The details that emerged were unsettling. The car's sensors detected the woman and bicycle six seconds before the crash, but after five seconds the car's AI had still not decided whether to alert the safety driver. The car was moving at 69 kilometres per hour when the driver looked up from her phone just in time to see the car hit the pedestrian.

The accidents in Arizona and California amplified fears that autonomous cars pose a threat to human safety. A survey conducted by Cox Automotive shortly after the accidents showed that public awareness of autonomous cars increased by more than 20% from 2016 to 2018, but over the same period, the proportion who said they would never buy a fully self-driving car grew from 30% to almost half. Respondents cited safety as a crucial issue.

The automotive industry is now relying on a charm offensive to turn things round. In the aftermath of the Arizona accident, Uber shut down its test facility in the state and said it would change the way it tested driverless cars. Meanwhile, before its vehicles were available to the public, Waymo embarked on an education campaign alongside organizations such as Mothers Against Drunk Driving and the US National Safety Council. And in late 2017, computer-chip-maker Intel recruited basketball star LeBron James to appear in advertisements to promote self-driving vehicles.

THE HUMAN FACTOR

The accidents involving self-driving cars have focused attention on a serious weakness in the plan to develop automotive autonomy in five stages, from low-level driver-assistance technologies, such as cruise control, to vehicles that can drive unsupervised. The fly in the ointment is human behaviour.

Tesla currently offers level-2 autonomy, which means that there are multiple systems to aid the driver in stopping, accelerating and staying in lane, but the driver must keep their hands on the wheel and their eyes on the road. Yet any car enthusiast driving a Tesla is sure to take their hands off the wheel at some point to see how well the car drives itself. And the crash in Florida, in which the driver had his hands on the wheel for just 25 seconds in a 37-minute period, shows that some will place even more trust in the vehicle's systems than that.

Level-3 autonomy, which is intended to be the next step up, is defined as a system in which the car is in control but can turn the wheel over to its human driver at any time if it encounters a problem it cannot solve. But many specialists now doubt that level-3 autonomy is viable. "Humans are terrible at remaining attentive if we're not doing anything," says Chris Bessette, program manager for autonomous driving at the Charles Stark Draper Laboratory in Cambridge, Massachusetts. A wandering mind leaves the human driver without the awareness of the situation that is needed if the car cannot decide what to do, and Bessette doubts whether control can be transferred effectively. Waymo developed and tested a level-3 system in 2012, but abandoned it after finding that human drivers "over-trusted the technology and were not monitoring the roadway carefully enough to be able to safely take control when needed".

Now, some in the automotive industry are changing gear and aiming to jump directly to level 5: fully autonomous vehicles that can perform to people's best expectations. Cars from General Motors will not have pedals or manual steering, and will operate only in limited urban areas that have been well-mapped to guide the self-driving cars. This is a big step beyond the automated people-movers already used in airports and theme parks, which run on guideways isolated from other traffic. The robo-taxis will be in complete control, and the rider will not be expected to pay any attention to the road.

The focus on urban driving is important. Cars in densely populated urban areas typically travel at half the speed they reach on larger highways. This reduces the stopping distance, as well as both the energy and the likelihood of crashes. The slower speed also reduces the workload for the AI, because it does not need to map as far ahead as it would at a higher speed. There are plans to take self-driving cars off the road if there is bad weather that could impair the system's ability to navigate the streets safely.

The use of robo-taxis will answer several questions about self-driving cars. Will the savings from not having to pay a driver offset the extra costs needed to automate the car? And how well will the sensors and software perform on busy city streets? The trials also provide an opportunity to test new technology, particularly to cut costs and extend the range of the expensive laser radars (lidars) that are needed to map the car's environment in detail to a distance of 200 metres or more, which will be required when travelling at higher speeds away from the city centre.

But the biggest test might be whether people will be willing to ride in them. From Google Glass to the US Atomic Energy Commission's 1958 proposal to use nuclear explosions in construction projects, history is littered with hyped technologies that failed to catch on. As AI takes on more-visible roles in our daily lives, its developers might soon find out whether the public would like this one to go the same way. ■

Jeff Hecht *is a freelance writer based in Newton, Massachusetts.*