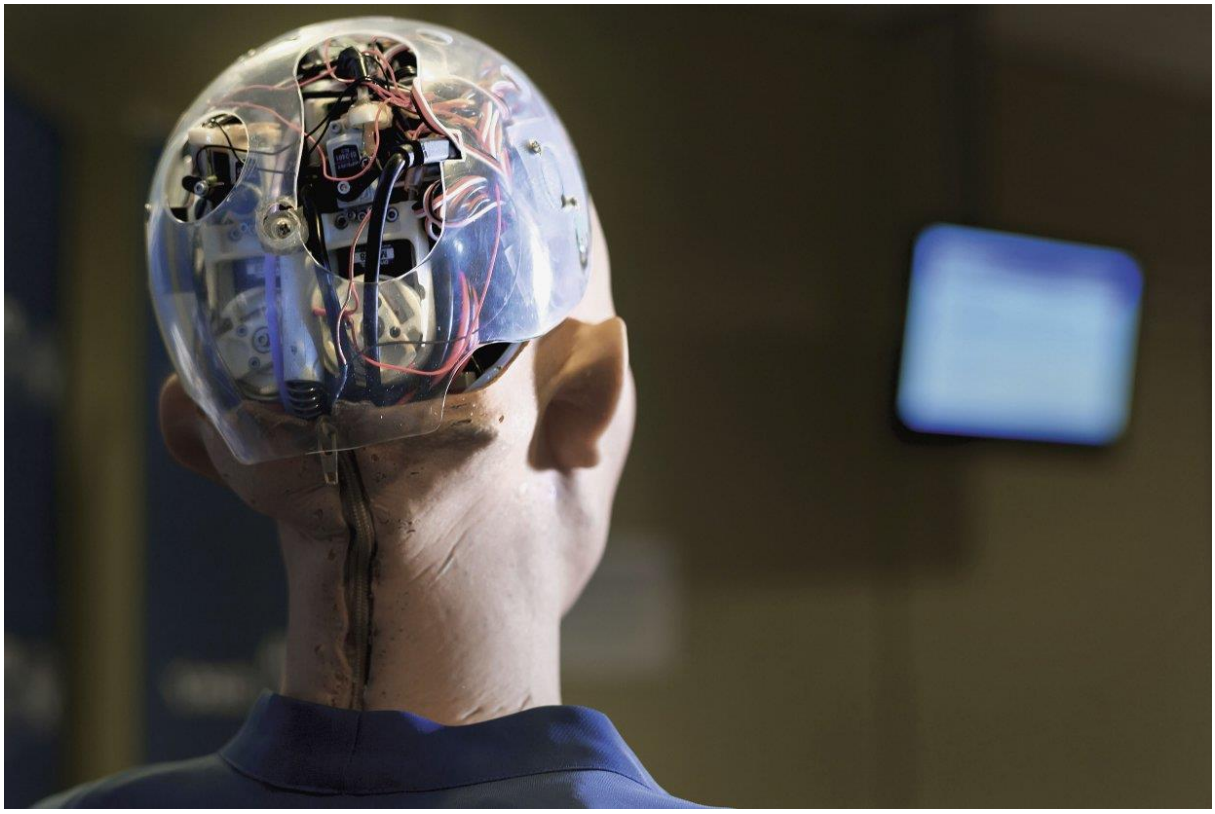
Artificial intelligence begins to transform security landscape

Artificial intelligence has advanced rapidly in line with computer processing power and the proliferation of data. *Professor Hussein Abbass* considers how artificial intelligence and trusted autonomous systems could change the way intelligence and security organisations process and respond to information

In July 1997, *The New York Times* quoted Dr Piet Hut, an astrophysicist at the Institute for Advanced Study in Princeton, as saying that it could be a century, "maybe even longer", before a computer could beat a human in the strategy game Go. In 2016, fewer than 20 years later, an artificial intelligence (AI) program named 'AlphaGo' beat Lee Sedol, a South Korean multiple world champion.



'Sophia', an AI human-like robot developed by Hong Kong-based humanoid robotics company Hanson Robotics, is pictured during the 'AI for Good' Global Summit at the International Telecommunication Union on 7 June 2017, in Geneva, Switzerland. Sophia was granted citizenship by Saudi Arabia at a conference on 25 October in Riyadh, Saudi Arabia. (Fabrice Coffrini/AFP/Getty Images)

1710186

Governments and organisations around the world are attempting to come to terms with the implications of this advance in technology. On 19 October, the United Arab Emirates announced

the appointment of 27-year-old Omar bin Sultan al-Olama as the minister of state for AI; the first such ministerial appointment in history.

The advancement of the technology is also likely to have significant social, ethical, and legal implications. On 25 October, Saudi Arabia granted citizenship to a robot named Sophia that incorporates AI technologies. On 4 November, Japan granted residency in Tokyo to 'Shibuya Mirai', an artificial intelligence program with no associated physical form. In both cases, despite the grants, there are significant questions about the legal status of these creations. Such questions are likely to become increasingly common.

[Continued in full version…]

Human intelligence

Humans are, among other things, information-collection and -processing systems. The human body is loaded with sensors that can collect information and with actuators that can affect the physical world. At the lower levels of cognition, the brain sends and receives signals to and from the body and can undertake what, in the methodology of signals intelligence, might be termed low-level signals analysis. At the higher levels of cognition, the brain can transform collected and processed signals into meaning and make decisions for the body to execute.

Humans have been augmenting their physical capabilities since the development of the first stone tools. In the modern world these tools can be operated physically, as when driving a car, or remotely, as with a pilot flying an unmanned aerial vehicle. However, these capabilities can also be exercised through the creation of technologies that can operate autonomously, as with driverless cars.

China's 19-year-old Go player Ke Jie considers his next move during the second match against Google's artificial intelligence programme AlphaGo in Wuzhen in eastern China's Zhejiang province on 25 May 2017. AlphaGo defeated Ke over a three-match series. (STR/AFP/Getty Images)

1710182

Human cognition is imperfect, with limited memory, finite cognitive resources, and biases. Accordingly, as with their limited physical capabilities, humans have sought to augment their cognitive capabilities with tools.

[Continued in full version…]

How AI works

To replicate human autonomy in cognition and action, computer programs need to be able to encode information about the world into a formal representation that computers can make use of. They also need to be able to update this information by making new observations and finding and assessing evidence. These systems also need to be able to generate new information from existing information or through information accumulated by interacting with the environment. Developing systems capable of engaging in this activity is the goal of practitioners in the field of AI.

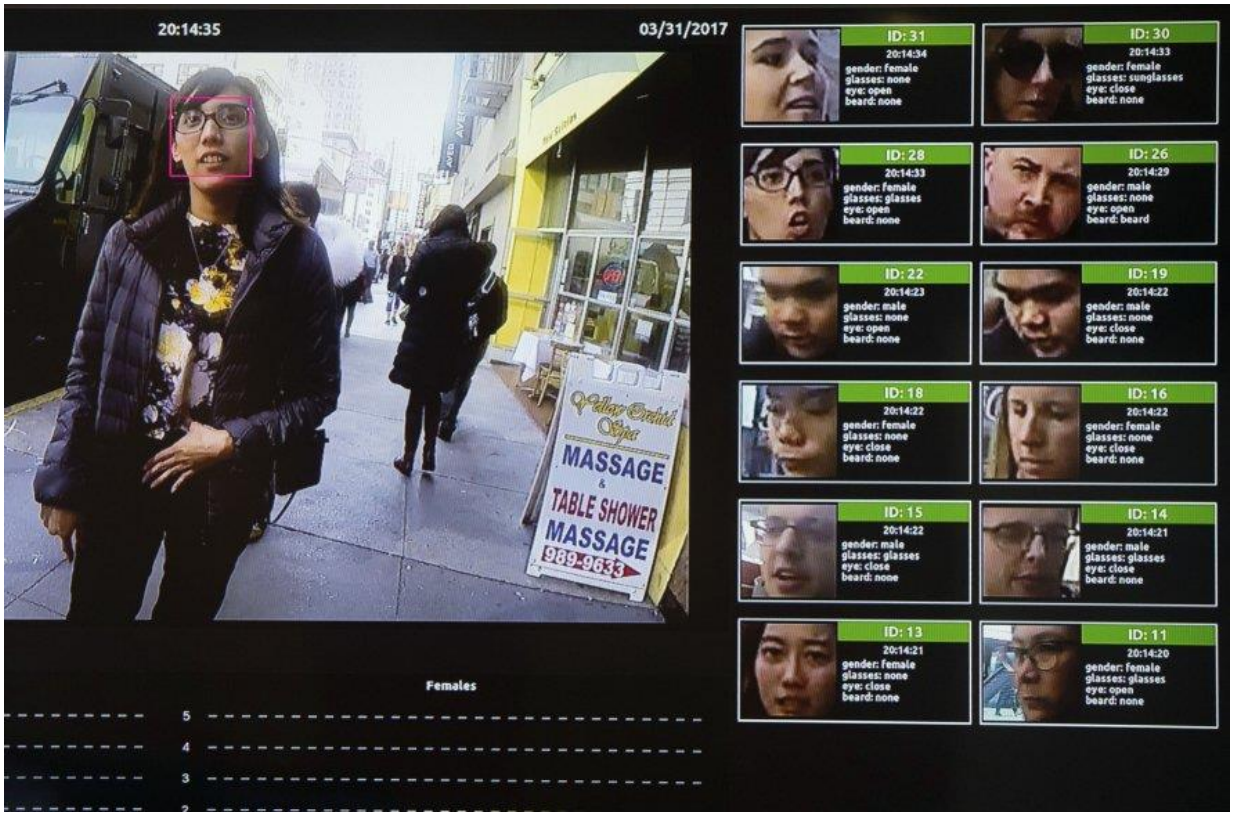
The term 'artificial intelligence' was introduced in a workshop in 1956 by a group of four influential scientists: John McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon. AI is primarily a field of computer science but is multi-disciplinary, drawing on work in the fields of mathematics, statistics, and cognitive sciences.

The classic branch of AI was the development of 'expert systems'. The objective was to transform everything humans know about the world and how the world works into simple condition-action pairs, for example: 'if the source for the data is unreliable, then the data is unreliable'. Although for most humans this is a fairly short intuitive leap, to a computer it is a logical relationship that needs to be specified in its programming. These pairs are encoded into the computer program as symbols (for example, computer code or words in English). The program then combines them into structures and performs operations on them based on defined rules to produce new expressions.

The challenges with this form of AI are manifold. First, it is a costly and time-consuming exercise to define the large number of rules that the system requires to operate. Second, these rules must be kept up to date, which is likely to be similarly resource-intensive. Third, and consequently, this form of AI does not scale easily. As such, expert systems are now primarily used in niche areas, including recommender systems, such as those used by popular on-demand music and television services to make suggestions based on a user's previous choices.

A second branch of AI overlaps with experts systems but deviates in the objective in that it enables systematic planning and reasoning about these plans. In the example of a group of robots that need to navigate through an urban area to find victims after a disaster, the AI planner will identify which robot will do what and how each of them will go about achieving their own assignment. Such a system would still require the manipulation of symbols as with the first branch. However, it could delegate lower-level processing to non-symbolic systems (see below). Although a planner system would still be limited by the scope of the symbols and rules introduced by its designer, the combination of symbolic and non-symbolic systems offers AI a way to mimic the rational mind.

A third branch of AI is machine learning. This branch is concerned with developing systems capable of acquiring knowledge beyond what is already known. The basic idea is to design an AI that improves itself with experience. The nature and form of this experience defines different types of learning: supervised, unsupervised, and reinforcement learning.



A display shows a facial recognition system for law enforcement during an Nvidia GPU Technology conference, which showcases artificial intelligence, deep learning, virtual reality, and autonomous machines, on 1 November in Washington, DC. Advances in graphics processing unit technology are supporting the development of increasingly powerful machine-learning systems. (Saul Loeb/AFP/Getty Images)

1710180

By way of example, a system may be designed to determine whether a particular image in a dataset contains a car. In the case of supervised learning, the system would be provided with a dataset of images that have been individually labelled as containing a car or not. The supervised-learner system would take this dataset and enter into a training cycle until it 'learned' the concept and could then apply it to a set of unlabelled images.

In unsupervised-learning systems, the machine-learning algorithm is guided by a mathematical cost function designed by a human user to group or summarise data, in this example to enable it to group similar cars together. The third form of learning is reinforcement learning, whereby the algorithm learns from interactions and feedback signals that 'reward' or 'punish' the algorithm in question each time it makes an assessment.

At present, the majority of successes in the world of machine learning fall under the first form: supervised learning. The drawback of supervised learning is that it requires significant amounts of data for the system to 'practise' on. However, given the rapid increase in the amount of data being generated, this issue does not present an insurmountable challenge.

[Continued in full version…]

Proliferation of sensors

The sensors and data analysis tools that used to be the exclusive preserve of a small number of intelligence organisations only a decade or two ago are increasingly available to the general public.

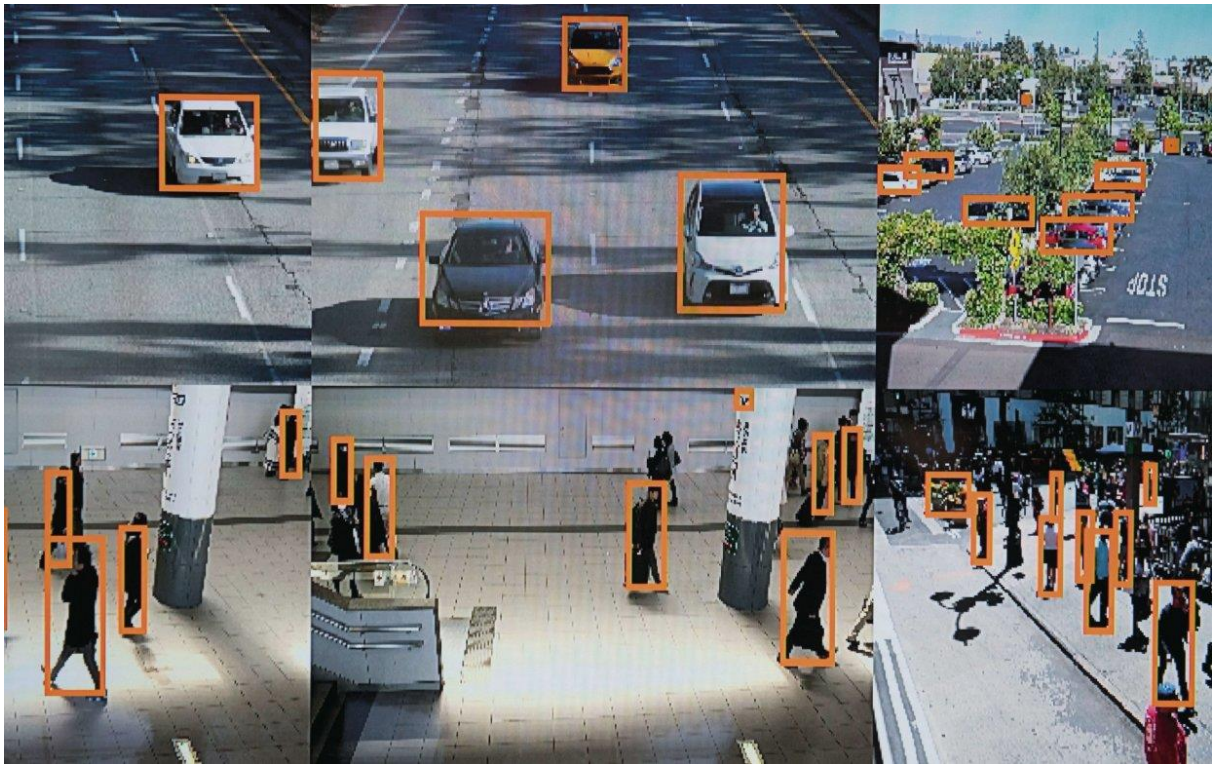
A modern smartphone is equipped with an array of sensors, including: microphones, cameras, accelerometers, gyroscopes, magnetometers, infrared light detectors, light sensors, and even radiation sensors. All of this collected information is processed digitally and it is likely that it will increasingly be stored remotely or shared between a wide range of users.

Similarly, freely or readily available software can be used by the owner of the device to process this data in a way that was not possible at the start of the 21st century. A minimally technical user can access automatic translation, the ability to search by an image, to convert speech to text, or vice versa. Deep neural networks can be loaded on to a smartphone to automatically detect and identify objects in an image. All of these capabilities are likely to have been considered top development priorities for leading signals intelligence agencies in the 1950s and 1960s, and are now freely available on commonly used devices.

[Continued in full version…]

Trusted autonomy

Although advances in AI and autonomous systems present a potential response to this trend, they also raise profound operational and ethical challenges. Responding to these challenges requires developing systems that are 'trustworthy'. Trusted autonomy requires additional layers of assurance above and beyond the requirements for machine cognition. In a hypothetical trusted smart autonomous system (TSAS), these layers would fall into two categories: technological and cognitive assurances.



A display shows a vehicle and person recognition system for law enforcement during the Nvidia GPU Technology conference in Washington, DC, on 1 November. A trusted smart autonomous system could theoretically be developed, using vehicle and facial recognition, to anticipate and prevent vehicle-impact attacks. (Saul Loeb/AFP/Getty Images)

1710181

The technological performance covers the physical and cyber security of the TSAS (ensuring that it cannot be subverted by hostile actors), along with the reliability, robustness, agility, and other elements required to ensure that the technological performance of the system is fit for purpose. The cognitive performance covers risk, transparency, ethics, legal compliance, social embodiment, and other elements to ensure the compatibility of TSAS with its organisational and social environment.

The significance of these requirements and the underlying challenges that they are intended to address would be greatest in life-or-death situations. Some more advanced weapons systems are already autonomous for part of their mission. For example, the US AGM-158C Long-Range Anti-Ship Missile (LRASM) uses autonomous guidance algorithms during its terminal phase to pinpoint specific targets. However, in these cases a human operator has already made the decision to deploy the weapon.

In contrast, *Jane's* assesses that the development of autonomous systems capable of making decisions about weapons deployment, either offensively or to respond to threats to human life, is technologically possible. Indeed, it is likely that trustworthiness is the main reason that such systems have not become available.

[Continued in full version…]

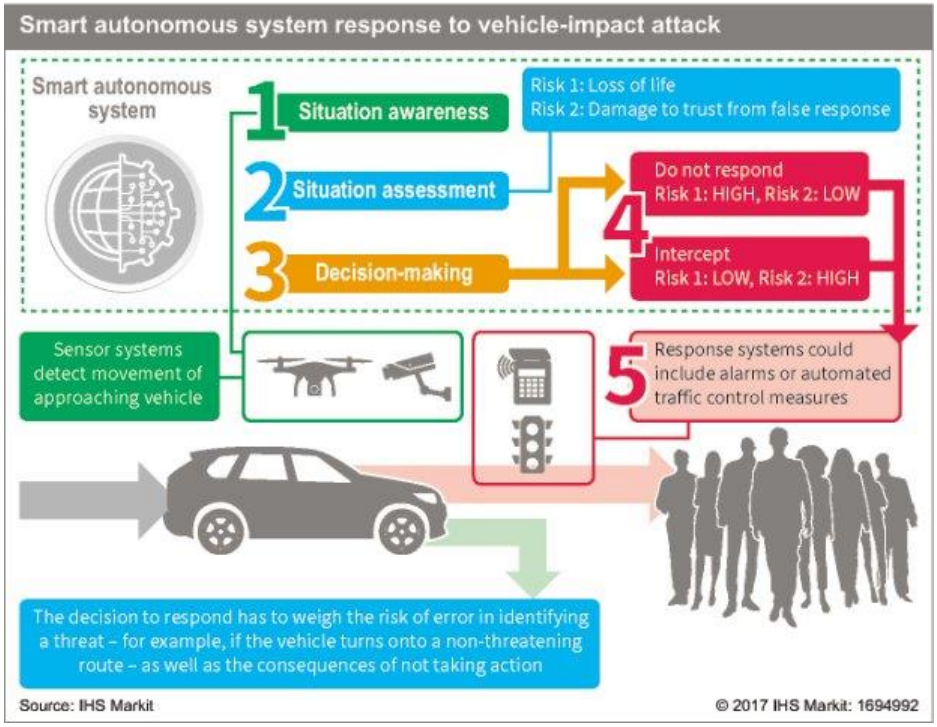
Autonomous response to a vehicle-impact attack

It takes 22.5 seconds for a van travelling at a speed of 80 km/h to travel a distance of 500 m. This is how much time it took the attacker in the 17 August vehicle-impact attack in Barcelona, Spain, to kill 16 people and injure 100 more. Such an attack could be planned on the spot, use a vehicle that anyone with a driving licence and enough money could hire, and last less than a minute.

These types of events are beyond the capabilities of existing intelligence services and the classic intelligence cycle. In the scenario considered above, it is unlikely that the time between intent and execution would allow for any form of human interference.

In practical terms, unless the attacker were already under full-time surveillance (meaning that in turn they had already been identified as a top-tier threat by an agency), preventing such an attack once it had begun would be practically impossible. Accordingly, as they become increasingly practical, the question of whether to deploy a TSAS to interdict such activity will become increasingly relevant.

It is conceivable that a TSAS – an 'eye in the sky' – could be developed that identified a subject from surveillance cameras, detected the acceleration of the car as an abnormal behaviour in a pedestrian area, predicted the trajectory the car was travelling along, assessed the casualty risk, and identified the best course of action. This action could be as simple as sounding an alarm or activating automated traffic defences – such as rising bollards in pedestrian areas.



Smart autonomous system response to vehicle-impact attack. (IHS Markit)

1694992

[Continued in full version…]

OSINT, AI, and counter-intelligence

The discipline of open source intelligence (OSINT) covers a wide range of activities, but an increasing proportion of OSINT involves online sources. In part this is because of a growing focus on social media intelligence (SOCMINT). Twitter is particularly valuable for intelligence collection as it offers real-time broadcasting of information. Twitter messages may appear to convey a limited amount of information with their current limit of 140 or 280 characters. However, with half a billion tweets per day, the least sophisticated form of analysis can be very powerful. The inclusion of imagery and geotagging metadata provide avenues for deeper exploitation.

A growing number of private companies provide services that make use of Twitter data. Comparatively simple data-summarisation and machine-learning techniques can be used to mine Twitter data to discover the most frequent words used within groups discussing a topic. This form of summarisation is marketed as a way of conducting sentiment analysis or for predicting the trajectory of an event as it unfolds. Many of these companies will work with governments to provide these services and it is highly likely that at least some intelligence and security organisations are engaged in similar projects. There are two implications for the work of intelligence organisations from this situation – one positive and one negative. The positive implication is that the advancement of AI raises the prospect of increasingly sophisticated and hence – potentially – useful analysis of sources such as Twitter.

[Continued in full version…]

On the web
Inter-state competition intensifies in high-performance computing
Technological revolutions threaten nuclear security
Picking your brains – The appliance of neuroscience

Author
Professor Hussein Abbass is an academic expert on artificial intelligence and trusted autonomy.