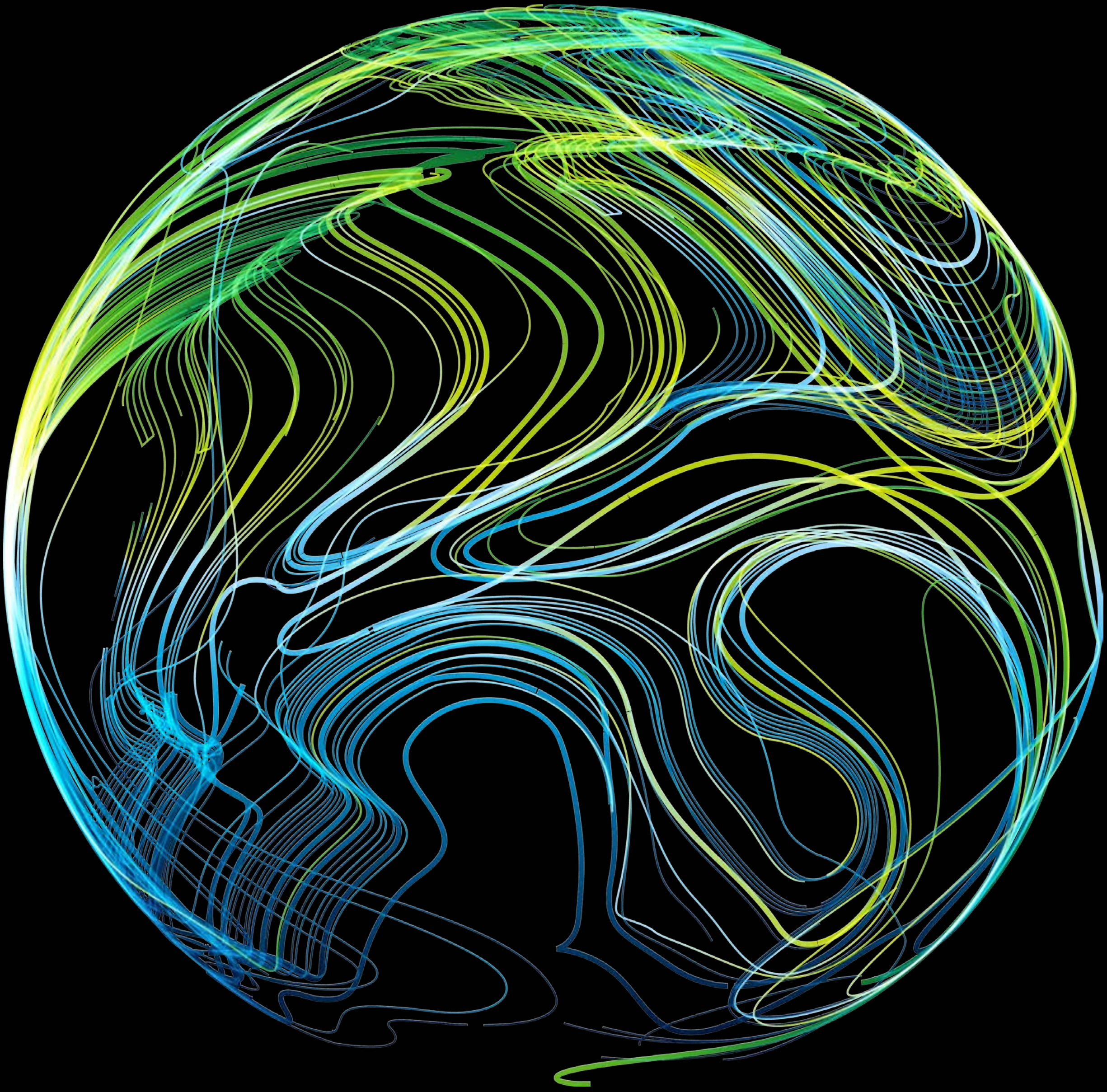
Deloitte.



2017 Technology, Media and Telecommunications Predictions

Middle East edition

Foreword

Welcome to the 2017 edition of Deloitte's Predictions for the technology, media and telecommunications (TMT) sectors.

For the first time in our 5 years of releasing our Middle East edition, we are including predictions for all three sectors together, and not splitting them into different sub-industries. This, by itself, is a reflection of the exciting industry we are in. An industry that continues to blur the boundaries of innovation, and reshape how operators, media players and technology companies collaborate and interact in an increasingly integrated market place.

Across the global and regional predictions, we believe that the distinction between sectors is fast becoming obsolete. The introduction of dedicated machine learning capability to smartphones is relevant across all industry sectors, not just the technology or telecommunications verticals. The transition to 5G and resulting implications on machine to machine communication is a critical enabler to new technology adoption, starting with self-driving cars. IoT itself is the epitome of this borderless ecosystem with operators and technology companies working closely together to shape the cities and lives of tomorrow. Cybersecurity is an evergreen topic in the region raising threats to media companies and Telcos equally, and requiring cross sectorial regulations and safety measures.

With smart cities and nations so high in the agenda of the Middle East countries, our region is at the forefront of this borderless market place, with regional Telcos talking more about AI and IoT than network expansion. In this day and age, breaking borders, albeit at industry level, is a refreshing twist. 2017 promises to be yet another exciting year for the TMT sector. We wish you all the best for this year and trust that you and your colleagues will find this year's predictions a useful stimulant in your strategic thinking. We look forward to discussing them with you.

Emmanuel Durou
Partner, Head of Middle East TMT industry
Deloitte & Touche (M.E.)

Paul Lee
Partner, Head of Global TMT Research
Deloitte Touche Tohmatsu Limited

Brains at the edge: machine learning goes mobile

Deloitte Global predicts that over 300 million smartphones, or more than a fifth of units sold in 2017, will have on-board neural network machine learning capability[144, 145]. These are computer models designed to mimic aspects of the human brain's structure and function, with elements representing neurons and their interconnections. They will allow smartphones to perform machine learning tasks even when not connected to a network. This functionality will enhance applications including indoor navigation, image classification, augmented reality, speech recognition and language translation even where there is little or no cellular or Wi-Fi connectivity, such as in remote areas, underground or on an airplane. Where there is connectivity, on-board machine learning may allow tasks to be done better and faster, or with more privacy.

Some tasks performed by computers or mobile devices are straightforward: a push of a button on a keyboard is translated into binary information that the processor is programmed to recognize. The letter 'I' provides an example. On a smartphone, when the language is set to English, typing the lower-case letter 'i' alone will prompt the processor to change it automatically to an upper-case 'I', since (in English) the lower-case version almost never exists on its own. This example of auto-correct is programmed, and is not machine learning – although other kinds of auto-correct do in fact use machine learning[146].

But other functions cannot be programmed explicitly in the same way. Recognizing that an object is a face, and whose face it is, in a world of varying light sources, hats and glasses, is remarkably challenging for programmers. Voice recognition and language translation are similarly challenging.

These types of challenges are better dealt with by machine learning – the process by which computers can get better at performing tasks through exposure to data. Up until now, that required massive computational power, the kind usually only found in clusters of energy-consuming, cloud-based computer servers equipped with specialized processors[147]. An example would be computer translation: years ago, translation consisted of looking up a word or two in one language from a stored dictionary, and substituting a word or two in another language. This kind of large-scale statistical machine translation was better than nothing, but far from perfect. By adding neural machine translation, translation is not done piecemeal, but sentences at a time, yielding results that are significantly more grammatical, idiomatic, and easier to understand[148]. In 2016, this was all done in the cloud, not on the mobile device, but one day this kind of translation and other tasks such as recognizing objects in images may be able to be done natively.

Although some smartphones in 2016 were capable of extremely limited machine learning tasks such as recognizing a single face or fingerprint, more powerful cognitive tasks only worked when connected to large data centers. New chips CPUs (central processing units, the traditional 'brain' of computers and mobile devices), GPUs (graphics processing units, historically used for gaming but also capable of machine learning tasks), or dedicated FPGAs (field-programmable gate arrays, a more expensive but more flexible kind of chip that can be reconfigured or re-programmed by the customer after manufacturing[149]) and/or special software emulators (an example would be a social network: the new Facebook app has software that allows phones to run neural network processes using on-board processors in 1/20th of a second[150]) will now be able to provide neural networks at prices, sizes, and power consumption that fit smartphones.

Machine learning on-the-go will not just be limited to smartphones. These capabilities are likely to be found over time in tens of millions (or more) drones[151], tablets, cars[152], virtual or augmented reality devices[153], medical tools[154], Internet of Things (IoT) devices[155] and unforeseen new technologies.

Historically, having gaps in connectivity was not a big deal: if our phones could not provide image classification or indoor navigation, we managed to do without. But as our phones have become more powerful and ubiquitous, they are becoming critical devices in our daily lives and need to be able to perform machine learning tasks all the time, not just most of the time. Translation is only one example. A smartphone-enabled medical device or vehicle-driving application that works all the time may be a matter of life or death, rather than just convenience.

Deloitte Global predicts that over 300 million smartphones, or more than a fifth of units sold in 2017, will have on-board neural network machine learning capability.

These are computer models designed to mimic aspects of the human brain's structure and function, with elements representing neurons and their interconnections.

From the core to the edge: a brief history of distributing intelligence

Moving intelligence from the core to the edge of networks (closer to the end user) has always created profound changes over time, even if the initial applications were less than exciting.

In the 1970s and 1980s most enterprise computing consisted of large mainframes and minicomputers in the basement of office towers, while employees worked on dumb terminals composed of cathode ray tube monitors (called 'green screens') and keyboards that had no on-board processing power[156]. In the history of computing, processing and memory were scarce and expensive. As a result, the IT architecture was highly centralized, with all of the intelligence at the core, and with the various peripheral devices being relatively dumb.

In the 1980s, processing and memory became exponentially more affordable, and these functions were pushed to desktop PCs. At first, the proposed applications allowed accounting, spreadsheets such as Lotus 1-2-3 (hugely popular in the 1980s[157]), and word processing to be done a little better than in the aforementioned centralized way, but over time the PC revolution ended up having many more interesting and important uses and effects.

Equally, pushing intelligence onto laptops in addition to desktops created entirely new markets. As intelligence moved closer to the end user with smartphones, the early and obvious applications were the ability to browse the web and emails even when away from our PCs. The last decade has shown just how much more powerful and transformative smart devices that fit in our hands can be, from apps to better cameras to language translation.

Based on this trend, we can expect that moving a special form of intelligence, machine learning, to the edge will be transformative. It will create opportunities and even industries that we cannot even begin to imagine. Perhaps we will have to ask our phones just what these new inventions will be.

The bottom line

As mobile devices become more capable of performing machine learning tasks, there are interesting telecom implications. Performing tasks such as image recognition on-board should reduce the amount of data that needs to be uploaded by consumers. That said, this effect is likely to be small compared to activities such as watching or uploading video, which may require thousands of times as much data, and are largely unaffected by on-board machine learning capabilities. However, reducing the amount of data to be transferred (and latency) is much more important in potential IoT applications and analytics[158]. Furthermore, carrying out machine learning on-board is inherently more private and secure[159].

Smartphones are increasingly becoming a critical tool as part of disaster relief[160]. With machine learning, they can be used by foreign aid workers to translate languages or assess medical requirements in real time. At present the mobile machine learning device must connect to far-off data centers – but can only do so provided the cellular network is working. While wireless networks are becoming more resilient, in the most severe emergencies towers can be knocked down and networks can be so congested as to be unusable; they may also lose power when stand-by generators run out of fuel[161]. In emergencies such as this, mobile devices able to perform machine learning tasks without connectivity would be a significant gain.

In the near term, most of the on-board machine learning capacity will be on consumer devices such as smartphones and tablets. But over time the applications for IoT devices may be more transformative. Autonomous vehicles will need to have machine learning capacity all the time, not just when cell signals are strong. At the speeds cars travel on highways, making decisions on-board would offer vitally lower latency: at 130 kilometers per hour, or 36 meters per second, every millisecond counts! Achieving lower latency could also be a reason to use mobile machine learning chips or software in jet engines, medical devices, or even oil and gas pipelines.

Medical devices that dispense insulin or detect epileptic seizures need to recognize patterns and respond in real time, regardless of connectivity. Drones with on-board machine learning are on the market today, and it is imaginable that every device from smart tractors, jet engines to horizontal oil drills will be able to benefit from on-board processing. As an example, the oil and gas industry already uses machine learning (carried out on mainframes) in downhole drilling data analysis[162]. It is possible that pushing this intelligence still further down the hole to the drill head would make for even deeper deep learning.

Another of our 2017 Predictions looks at the role of compromised IoT devices in Distributed Denial of Service (DDoS) attacks[163]. IoT devices are not regularly scanned for malware, nor are they easily patchable. The malware can be removed, but unless the password is changed they are likely to be reinfected soon: perhaps as soon as 98 seconds[164]. As of late 2016, chip vendors were already suggesting that on-board machine learning could detect zero-day malware (that is, previously unknown), and detect or classify suspicious or anomalous behavior[165]. On-board machine learning therefore has the potential to protect the devices in our lives and might even help turn the tide against the growing wave of cyberattacks.

In the near term, most of the on-board machine learning capacity will be on consumer devices such as smartphones and tablets. But over time the applications for IoT devices may be more transformative. Autonomous vehicles will need to have machine learning capacity all the time, not just when cell signals are strong.

Designed and produced by The Creative Studio at Deloitte, London. J10517