Intelligent Connectivity

Unleashing opportunities with the power of 5G, AI and cloud

10 December 2019

Tom Morrod
Research Executive Director

Julian Watson
Principal Analyst

Shen Wang
Senior Consultant

Contents

Table of Contents

Intelligent Connectivity

Unleashing opportunities with the power of 5G, AI and cloud

Julian Watson, Principal Analyst

Preface

Intelligent connectivity refers to:

- The ability of devices, buildings, computers and people to acquire knowledge and skills.
- Their capacity to make decisions and change their behavior based on the acquisition of this knowledge and skills.
- The role of connectivity and collaboration between devices, buildings, computers and people in distributing this knowledge and skills.

Intelligent connectivity is enabled through the collaboration of AI, connectivity (including 5G), cloud and edge and IoT. These ever-evolving technologies will work together to create new, immersive experiences for consumers, address challenges and create opportunities for enterprises and industry. For instance, 5G will enable much faster extraction of data from the cloud onto devices, reducing the need for storage. IoT will yield rich streams on data; AI will generate insights that can improve customer experience. The edge will deliver the rapid response required for applications that need to react to changes in the physical in real time.

The diverse range of players in this space will need to think creatively and collaboratively to understand and predict the future needs of their customers. Their thinking will be guided less by the "push" of product and services, but more by the "pull" of consumer experiences and desired outcomes for enterprise and industry. This thinking will lead to changes in how products are designed and fuel entirely new business models.

Intelligent connectivity

Defining intelligent connectivity

The Oxford English Dictionary (OED), defines intelligence as "the ability to acquire and apply knowledge and skills".

Its three definitions of intelligence demonstrate that not only humans, but also devices, buildings and computers can possess such knowledge and skills:

- Having or showing intelligence, especially of a high level.
- (Of a device or building) able to vary its state or action in response to varying situations and past experience.
- (Of a computer terminal) incorporating a microprocessor and having its own processing capability.

Connectivity refers to:

- The state of being connected or inter-connected.

- (In computing) the capacity for the interconnection of platforms, systems, and applications.

Therefore, the term **intelligent connectivity** implies:

- The ability of devices, buildings, computers and people to acquire knowledge and skills.
- Their capacity to make decisions and change their behavior based on the acquisition of this knowledge and skills.
- The role of connectivity and collaboration between devices, buildings, computers and people in distributing this knowledge and skills.

There are varying degrees and types of **intelligence**. As children, humans start to develop intelligence through the experience of interacting with other children, adults and the physical world. For instance, a child touches a lightbulb and recoils in pain. He learns that the pain he feels on his hand is caused by the heat of the lightbulb. He is unlikely to touch a lightbulb again.

To take another example, a smoke sensor in a building detects the presence of smoke. This automatically triggers an announcement over a public address system requesting office workers to vacate the building. It also automatically triggers a phone call to the local fire department, who attend the scene. The office workers start to vacate the building.

They do this because they understand from past experience that the presence of smoke could mean that there is a fire in the building that could cause them physical harm. In this example, the intelligence of a device (a smoke sensor) is distributing knowledge to office workers who respond to it accordingly.

The fire officers who attended the scene locate the fire and put it out. They inform the building facilities director that the fire was caused by a boiler exploding due to overheating. A new boiler is bought and installed. This replacement boiler includes embedded sensors that monitor gas and water supply, temperature and pressure, and connectivity.

This sensor data is communicated via wired or wireless technology from the boiler (now an **IoT** device) to a **cloud**-based portal, which the building director can access on her laptop or smartphone. Through the portal, she can remotely manage the operation of the boiler and how it transmits information. For instance, she could set the boiler to automatically close down when it reaches a certain temperature. She may also choose to receive an automated text message if the boiler exceeds a temperature threshold.

The above scenario may reduce the likelihood of another fire in the building caused by the boiler overheating. But in this scenario the cause of the boiler overheating has not been established. Neither has likelihood of the boiler overheating been reduced.

In an alternative scenario, the data from the embedded sensors is transmitted to a cloud-based analytical engine. This applies **AI** (**artificial intelligence**) to the data and learns what combination of factors are likely to cause the boiler to overheat and explode. This insight can be applied to automate the operation of the boiler, including the adjustment of settings and automatic shutdown. In this scenario, data flows from the device, the boiler, to the cloud-based analytical engine, and intelligence flows back from the cloud to the device.

Across many industries, applications and use cases, there is a growing requirement intelligence to be applied as close as possible to source of data. In the future, an autonomous car will need to respond in real-time to changes in the physical environment. A physical obstacle may appear in its path. The autonomous car will need to "see" the physical obstacle, identify what it is (a cardboard box or a person) and react accordingly – applying the breaks if it is a person. This is an example of intelligence at the edge, commonly known as **edge computing**.

As previously stated, the term **intelligent connectivity** implies not only the intelligence of devices, buildings, computers and people but also the role of connectivity and collaboration between devices, buildings, computers and people in distributing these knowledge and skills.

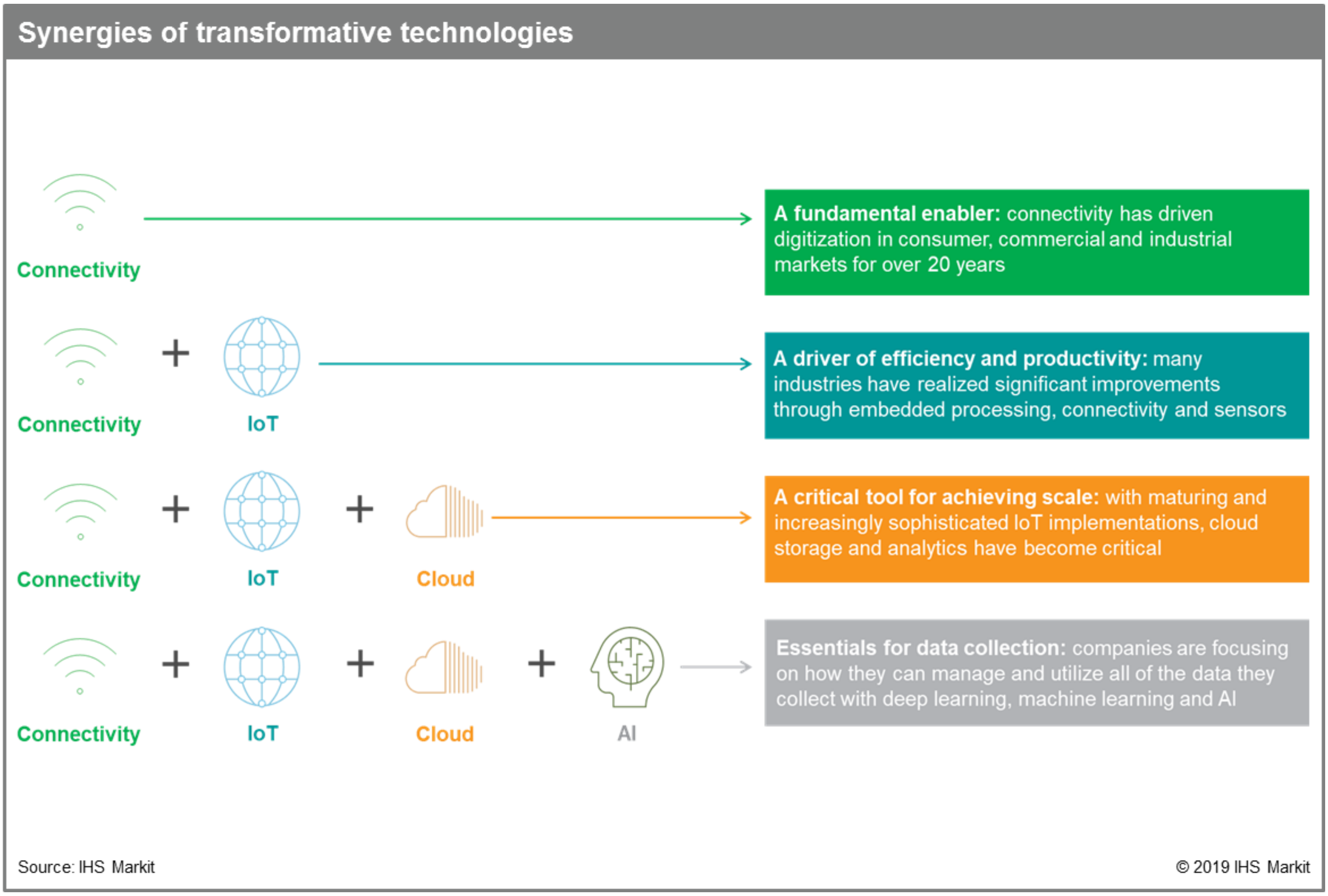
Connectivity is the wired or wireless pipe through which data is moved and knowledge and intelligence exchanged between devices, buildings, computers and people. The wireless connectivity landscape is diverse, ranging from short-range and local area network technologies like 802.11 (Wi-Fi), Bluetooth Low Energy (BLE) and ZigBee to the cellular wide area network (WAN) technologies 2G, 3G, and 4G LTE. The latter technology, 4G LTE, itself consists of several different iterations, serving applications that require deep range and a long battery life, but limited bandwidth (e.g. NB-IoT) and applications that are bandwidth hungry (LTE-Advanced).

The latest arrival to the crowded connectivity space is 5G. As we discuss later, 5G is distinct from previous generations of cellular technologies in that it has been designed at the outset to address many different technical requirements, device form factors, applications and audiences.

The building blocks of intelligent connectivity

In the previous section AI, the cloud and edge, connectivity (including 5G) and IoT were namechecked as enablers of intelligent connectivity. In this section, we discuss how these, and other building blocks will work together to create a new wave or consumer, enterprise, and industrial applications and use cases.

The diagram below (**Synergies of transformative technologies**) shows how different transformative technologies are collaborating to address the evolving requirements of people, enterprises and industry. The addition of embedded processing and sensors (IoT) to connectivity has yielded rich streams of data on the status, location and condition of connected nodes or their surrounding environment. The cloud is addressing the requirement to store and apply analytics to such large volumes of data. AI techniques are helping to manage this data and generate useful business insights from it.



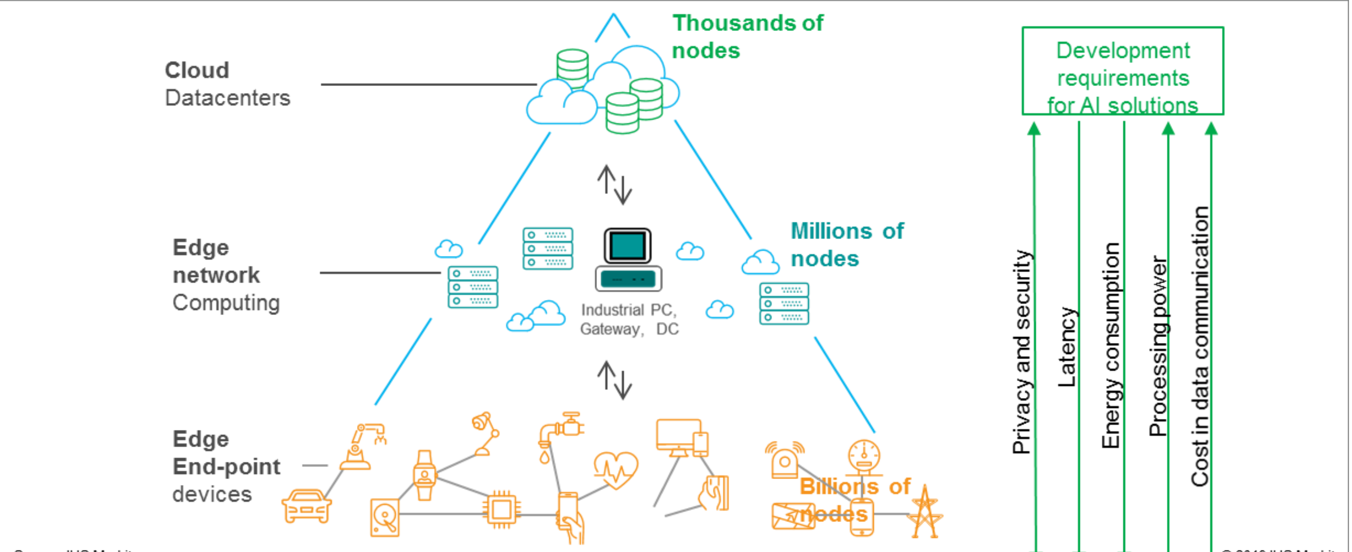
The following schematic (**AI diversification from Edge to Cloud**) show that intelligence in the form of AI can be applied across the cloud (datacenters), the edge network of industrial PCs, gateways and edge servers and edge end-points - the diverse universe of IoT nodes.

In this schematic, edge end-points, such as sensors that measure attributes like temperature and humidity, is where data is first generated and elaborated. Network edge/fog infrastructure gathers the data from edge end-points and elaborates it further before transmitting data onwards to the cloud for storage and further analytics.

The parameters shown on the right (privacy and security, latency, energy consumption, processor power and cost of data communication) will shape the development requirements for AI solutions: namely the application (inference) and training infrastructure required.

Energy consumption is a particular issue for battery-powered edge end-point devices. Performing data-intensive analytical workloads such as training could reduce device battery life and ultimately negate the original investment case behind the IoT/AI application.

AI diversification from Edge to Cloud



Source: IHS Markit

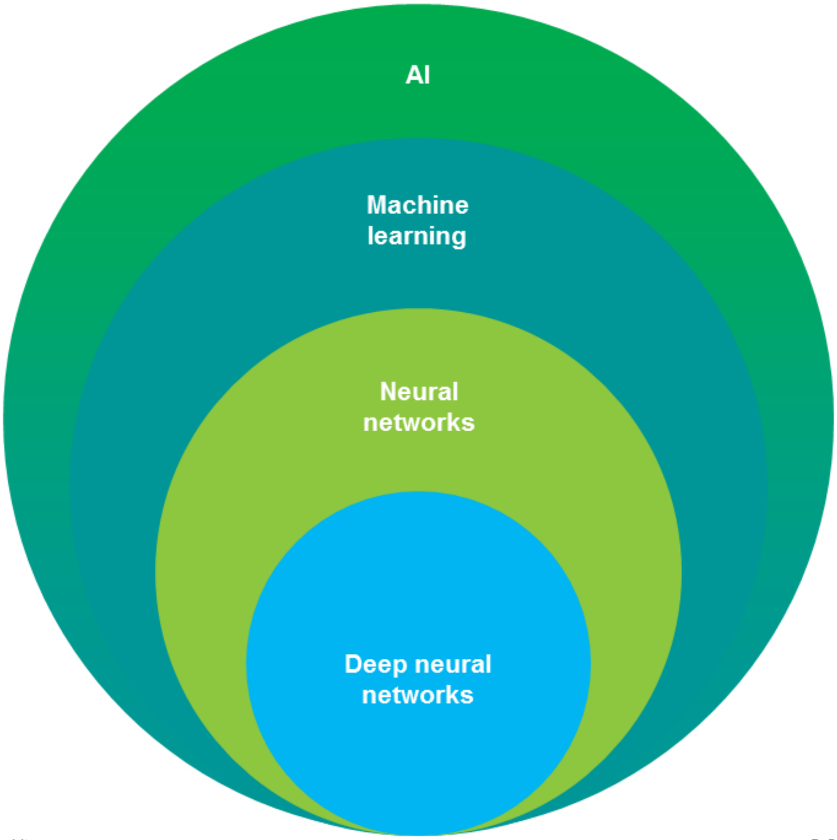
© 2019 IHS Markit

The development of AI

AI generically refers to the body of science that studies how to enable machines to perform independent problem solving, inference, learning, knowledge representation, and decision making. An intelligent machine does not need necessarily to show all these skills, but it must comply with one or any combination of them. More specifically, AI can fulfill four major skills: perception, learning, abstraction, and reasoning

There are several forms of AI. When machines are required to show specific self-learning skills, IHS Markit implicitly refers to a subset of AI called "machine learning" (ML). In this extent, ML is a set of algorithms that gives machines the ability to automatically find and learn patterns by feeding them with data, without explicitly programming them. In the basket of ML fall several technologies and techniques to which neural network (NN) and deep learning (DL) also belong to. NN and DL refer to computational models that try to emulate the structure and workings of a human brain, including process phases like training and inference.

Subsets in AI algorithms



Source: IHS Markit

The "training" phase sees large volumes of data fed into a computer's NN brain. The data may relate to sounds, images or some other type of information. The computer learns to recognize specific objects and identify patterns and discover how to identify objects in a variety of situations. The resulting trained model is then applied at the inference phase to provide output (e.g. recognizing a specific image) on-real time input. As cited before, in human analogy, a child learns that touching a lightbulb will cause pain (training), so he avoids touching the next lightbulb he sees (inference).

IHS Markit notes many industry verticals are actively investigating the potential application of AI across a plurality of use cases, such as prescriptive & predictive maintenance and process control in manufacturing and production, human-machine interface (HMI), video analytics and image analysis, and autonomous machines.

But AI is currently in its infancy: we are just into the so-called "**narrow**" or "**weak AI**" today. At this "**weak**" stage AI is "weaker" than humans or is equal or superior but just on limited tasks and senses. Most applications today have reached this point with four basic abilities: perception, learning, abstracting, and reasoning.

The further development of AI to the more advanced stage of "**Generic – multi-modal AI**" (comparable to human intelligence applied in all senses) and beyond will be aided by progress in the following areas:

- **Scale**: instant access to thousands of machines and sensors to perform complex and granular data analytics.

- **Performance**: fast handling of differentiated AI workloads with adequate power consumption.
- **Quality**: structured data and system training via accurate and reliable models and appropriate input data.
- **Custom**: flexible and adjustable solutions to different use cases—scalable platforms to ensure profitable business cases.

Nevertheless, there are several challenges on the horizon that could hinder the development and ultimate adoption of AI. Foremost among these are data privacy/security and ethics.

Future forms of AI could be more intelligent than human beings. As such, it may become impossible to predict, let alone control, the actions of an AI-enabled device that is superior in intelligence to a human. In the future, unmanned devices and robotics may be sufficiently intelligent to be able to make autonomous decisions that could result in harm being done to human beings.

Ethics in AI is a very tough problem to address, and there are two conflicting reasons. On one side is the pervasive influence that AI will have in the global society and the entire industry, including military. On the other side is the mix of cultures, traditions, and values that are extremely diversified across races and religions in different countries. In other words, on one side we need to tackle ethics and AI on a global basis. On the other side, the differences and peculiarities present in the world represent a huge challenge even in the definition itself of what the word ethics might mean.

AI in the cloud or at the edge

AI in the cloud or at the edge? This binary, somewhat rhetorical question, is frequently posed, but fails to take into account:

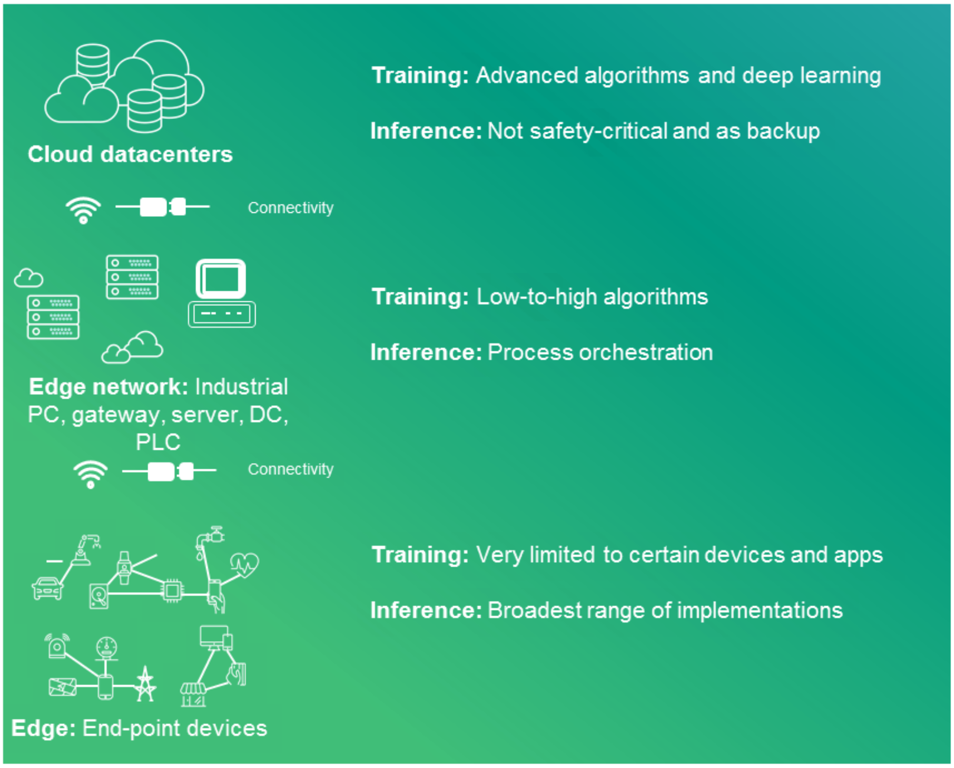
- The inter-dependence of cloud and edge architectures.
- That many real-life applications and use cases will need a hybrid cloud-edge approach.

Whether compute is concentrated in the cloud or at the edge will depend on the requirements of each application, illustrated in the schematic **AI diversification from Edge to Cloud**, namely:

- Privacy and security
- Latency
- Energy consumption
- Processor power
- The cost of data communication

As the graphic below illustrates, the cloud (datacenters) will perform the most intensive workloads (training). By contrast, the broadest range of inference implementations will be performed on edge end-points. Many, though not all of these nodes, are battery operated devices designed specifically for low total cost of ownership (TCO) use cases. STMicroelectronics, Renesas and NXP are integrating ML capability into new MCU (microcontroller) product families that are optimized for such power-constrained, low-cost IoT devices.

Different use cases for training and inference implementation



Source: IHS Markit © 2019 IHS Markit

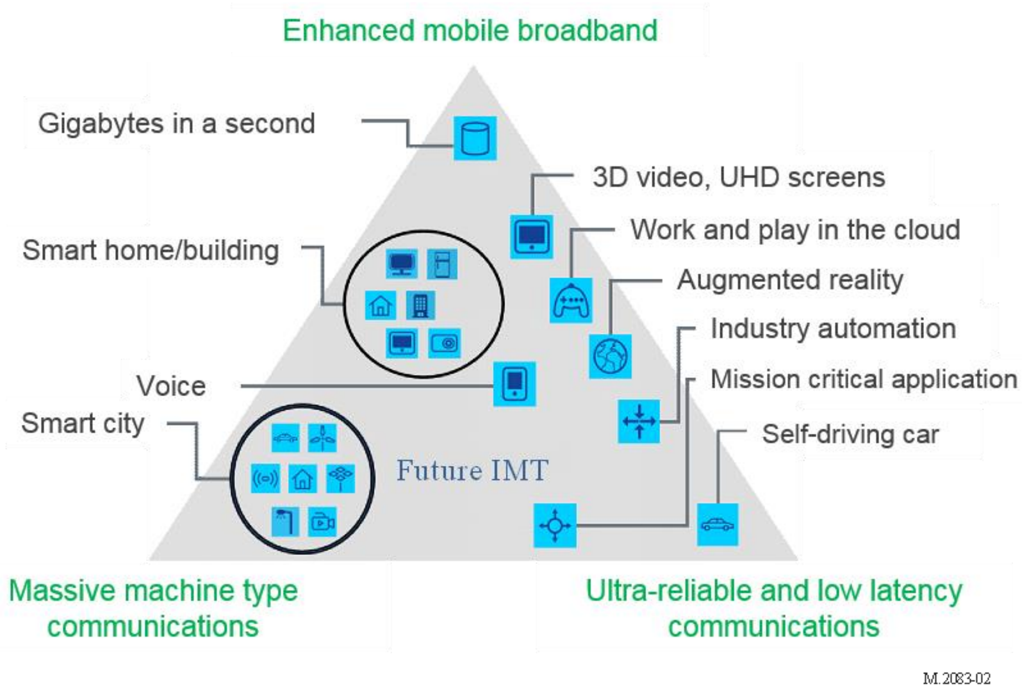
The graphic also points to the wired or wireless connectivity that links up edge end-points to the edge network and the edge network to the cloud. 5G will play an increasing important role here. IHS Markit expects that 5G will be integrated into several IoT gateway products before the end of 2019 to provide enhanced downlink and uplink bandwidth for commercial and mobile applications.

Global 5G development

5G defined

5G, as its name suggests, is the fifth generation of mobile technology. Unlike previous generations of mobile technology, 5G has been designed from the outset to serve the requirements of not only consumers and enterprise, but industry verticals too.

The graphic below shows the 3 broad types of use cases that 5G is designed to serve: enhanced mobile broadband (eMBB), Massive machine type communications (Massive IoT), and Ultra-reliable and low latency communications (URLCC), which is also commonly referred to as Mission Critical Services (MCS).



Source: ITU-R M.2083-0 9/2015

The technical requirements for eMBB, Massive IoT and MCS are diverse. eMBB use cases like 4K video streaming and cloud gaming are bandwidth-intensive. Massive IoT use cases such as smart metering typically requires deep coverage and extended battery life. Industrial automation, an MCS use case, demands very low latency and very high levels of availability. 5G's targeted performance characteristics seek to address these diverse use cases through 1–10Gbps connections to endpoints in the field (bandwidth); 1 millisecond end-to-end round-trip delay (latency); perception of 99.999% availability or *10*-5 packet loss rate and up to 10-year battery life for low-power, machine-type devices.

3GPP Release 15 introduced the first set of 5G standards. 5G non-standalone (NSA) NR (New radio) specifications were completed in December 2017, six months ahead of schedule, due to a strong push from various stakeholders that wished to deploy 5G as soon as possible. 5G standalone (SA) NR specifications, also part of 3GPP Release 15, were completed in June 2018.

Today's commercial standards-based 5G networks are based on 5G NSA NR, which leverages existing LTE networks, and primarily serve eMBB use cases. Work items addressing Massive IoT and MCS requirements and a host of other features, such as multicast/broadcast, positioning and C-V2X (Cellular V2X) will come in Releases 16 and 17, which are currently set for completion in 2020 and 2021 respectively. This standardization work will gradually facilitate the expansion of the 5G ecosystem of chipsets, modules and devices and pave the way for industry verticals to develop proof of concepts (POCs).

Network slicing will come with 5G SA NR

IHS Markit's Evolution from 4G to 5G Service Provider Survey – 2019 suggests that in 2020 around a third of 18 leading service providers (accounting for half of the world's telecom capex and revenue) will move to NG SA NR. This move will pave the way for 5G network slicing.

IHS Markit defines network slicing as "a logical partition or a virtual piece of a physical network, including fixed and mobile, physical and virtual, that has connections, capacities (bandwidth), and characteristics (e.g. Service Level Agreements or SLAs) that make the slice look and act like a physical network to the user, programme or process that is using it".

3GPP has identified a list of more than 12 network slicing requirements. These include the ability to dynamically create slices for different scenarios and devices, the ability to access different slices, that each slice will not negatively impact another, and guarantees around flexibility, customisation, and API access.

Network slicing is not a new concept, but 5G's very high bandwidth and very low latency capabilities make it relevant for a range of scenarios that have exacting performance requirements. Potential consumer applications leveraging 5G network slicing include video and broadcast (advanced formats and resolutions – 4K, UHD, 360/VR immersive video, as well as live broadcast events and interactive content) and gaming: advanced multiplayer mobile gaming, VR and AR games, VR esports and cloud gaming.

This opens opportunities for telecoms operators beyond monetising data volumes, into such areas as guaranteeing service levels for content and cloud providers.

It's a 5G world, already!

It's a 5G world, already! The year 2019 has witnessed the first wave of standards-based 5G commercial launches. According to the Global mobile Suppliers' Association (GSA), by October 2019 50 operators had launched 3GPP-compliant 5G commercial services across 27 countries. GSA also noted that 328 operators in 109 countries were investing in 5G.

Based on IHS Markit's assessment during the first half of 2019, there are three types of fundamentally different 5G deployments: large-scale, small-scale and laggards. Large-scale deployments of over 10,000 5G NR/gNBs have occurred in China and South Korea. In Australia, UK, Saudi Arabia, Switzerland, UAE, and the US there have been small-scale, tactical deployments of 100s of 5G NR/gNBs. Among the laggards are operators in countries such as Argentina, Colombia and Uruguay where LTE is underdeveloped and may be leapfrogged by 5G.

All existing 5G commercial deployments are based on 5G NSA NR; a variety of different spectrum bands have been utilized. IHS Markit's Evolution from 4G to 5G Service Provider Survey – 2019 found that most existing 5G NR commercial deployments and service launches and those planned for the rest of 2019 will use mid-band spectrum (3–6GHz), particularly 3.5GHz and 4.5GHz.

This is the most widely used spectrum in Asia Pacific and EMEA. 6–39GHz spectrum, which includes the mmWave bands of 28GHz and 39GHz, is the next widely used spectrum (e.g. deployments from AT&T and Verizon in the US). Sub-1GHz and 1–3GHz represent the least commonly utilized spectrum. As standards develop and more spectrum becomes available for 5G, an increasing number of operators will deploy 5G in more than one band to build out capacity and coverage.

5G may be in its infancy, but the development of the 5G device ecosystem has been ramping quickly. As of mid-November 2019, according to the GSA, 72 vendors had announced a total of 183 5G devices. These came in no less than fifteen different form factors, such as smartphones, indoor and outdoor CPE (customer premises equipment), laptops/notebooks, robots, drones, enterprise routers, IoT routers and dongles/adapters. Of the 183 announced 5G devices, over 40 are commercially available.

And China is driving it

China has been preparing for commercial 5G launch for several years. City-focused trials in 2018 have been followed by massive trials this year. In Early November 2019, China's three mobile operators China Mobile, China Telecom and China Unicom, launched commercial 5G services – a couple of months ahead of schedule. IHS Markit expects that by the end of this year the three will deploy some 100,000 5G base stations, rising to one million in 2020. By the end of this year, 5G is set to be available in 50 cities across the country.

China's three mobile operators are deploying 5G NR in NSA and SA mode, but plan to expedite the transition to SA in 2020. According to their 1H 2020 financial presentations, China Mobile plans to reach 50,000 5G BTS (base station transceivers) by the end of 2020 and China Telecom and China Unicom both of up 40,000.

Alongside deploying the world's most expansive 5G network, China's operators have been working with partners to stress test various commercial, industrial and consumer applications. China Unicom has worked with Beijing's 301 Hospital and Fujian Medical University on 5G remote surgery. In December 2018, China Mobile subsidiary Migu broadcast the 12th Annual Migu Music Awards, which is the first worldwide 5G+ true 4K music awards broadcast – using network slicing.

The role of AI in 5G

In the future, mobile networks will become more complex. 5G will co-exist with legacy networks for many years to come. Operators will deploy 5G in multiple spectrum bands and serve many different applications, from those requiring high bandwidth and low latency to those needing deep coverage and low power. Beamforming, massive MIMO (multiple-input and multiple-output) and network slicing will be among the various techniques used to increase throughput, coverage and capacity and address the performance requirements of specific applications.

AI and ML have the power to identify patterns from huge volumes of data – including from mobile networks and devices. This learning can be applied, for instance, to recognise the patterns leading up to events such as a spike in network traffic. This, in turn, can be used to allocate additional capacity to meet increased demand on the network and prevent an unplanned downtime. There are several other ways in which ML might be applied.

These include:

- Low latency data processing and transmission.
- Multi-access edge computing (MEC), particularly for applications that require very high bandwidth and very low latency.
- Designing and managing dedicated network slices.
- Beamforming improves the throughput available to a device. ML algorithms can help the device to connect to the best beam from within the cell or a neighbouring cell.
- Massive MIMO refers to the large number of antennas (at least 32) in the base station antenna array. Massive MIMO is designed to increase throughput, network capacity and coverage by reducing interference from neighbouring cells. AI and ML can be used to optimise beamforming by forecasting user distribution and dynamically managing antenna resource.

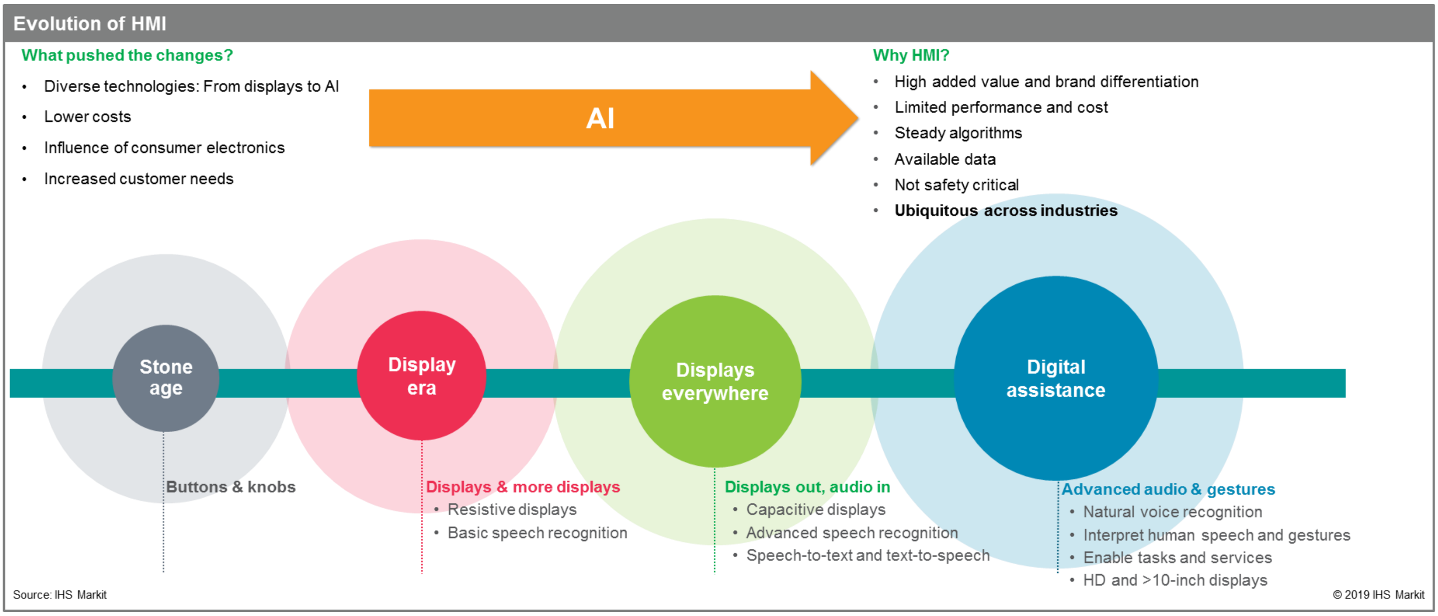
The changing nature of human-machine interaction

The rise of the digital assistant

One of the facets of **intelligent connectivity** is "the role of connectivity and collaboration between devices, buildings, computers and people in distributing this knowledge and skills". In other words, humans have a symbiotic relationship with devices, buildings and computers.

The concept of human-machine interaction (HMI) has been around for many years, but popular awareness of HMI has spiked in recent times by the strong adoption of digital assistants, such as Amazon Alexa, that have the ability to learn and respond to natural language. Using voice, a person can put into action any number of tasks, from switching on a light, playing a song from a twinned music streaming service over the smart speaker or even placing an online shopping order. Of course, enabling such a range of "skills" involves much more than developing AI-capable hardware and software; also, technical interoperability with other devices and integration with online services. The convenience of voice is driving the adoption of digital assistance way beyond the home to embedded infotainment systems in the car, hotel rooms, offices and hospital waiting rooms.

Digital assistants are designed to react to verbal commands or requests made by humans. Increasingly, the capability of HMI will extend to other forms of active or passive communication such as gesture/emotion recognition and even into areas such as recognizing and responding to the signs of driver drowsiness.



The enduring utility of smartphones

In IHS Markit's view, the rise of the digital assistant has not demised the key role smartphones play in the life of consumers. Different models of smartphones may vary greatly by screen size and resolution, audio quality and amount of storage and degree of processing power. But they also share some common attributes: embedded cellular, Wi-Fi and Bluetooth connectivity, sufficient storage to host multiple applications and a display good enough for reading, watching video and interacting with applications. Unlike tablets and laptops, they are small and light enough to be carried easily by users.

These common attributes currently give smartphones a central role in the consumer IoT space. All of the following of the gateway, controller/receiver and display functions are things consumers take for granted in their daily lives.

At the center of consumer IoT is the human – the smartphone allows a person to control devices in the home, stream music to earbuds, remotely monitor the home and discover insights from other devices. As such, a smartphone can play at least three roles in the consumer IoT ecosystem.

First, as a gateway for wearables such as smartwatches and health monitors, which are most commonly Bluetooth-connected. The Bluetooth-connected smartphone receives data from the wearable (e.g. number of steps, hours slept and heart rate) and pushes it out to the cloud for analysis. It then sends data back to display on the wearable and acts as the gateway conduit for firmware updates. Smartphones also provide basic connectivity to earbuds

Second, as a controller/receiver: integrated Wi-Fi allows the smartphone to control, through installed apps, smart home features such as lighting, heating and home appliance settings. The smartphone user can perform these tasks either at home, connected to the home router/hub or outside of the home in areas with cellular or Wi-Fi connectivity. Beyond the home, smartphones can receive and act on alerts from motion sensors and intruder alarms.

Third, as a display for data and video from other devices: The larger size and greater resolution of smartphone screens allow them to display richer data sets, such as the visualization of weekly fitness activity and sleep quality, than could be shown on smaller smartwatch screen. Live footage and alerts from video cameras at home, tracking pets or dependents, can also be accessed remotely via smartphone apps.

The smartphone is not the only device that supports some of these functions. Tablets share many of these capabilities, but less than half are cellular-enabled, thus limiting their suitability to control and receive alerts from smart home devices when the user is away from home. Digital assistants can be instructed by voice to control lighting, heating and other systems when the user is at home. Such services or skills can either be discovered or added through an accompanying smartphone app – so digital assistants tend to work with rather than compete with, the smartphone.

Although the current state-of-the-art smart phone already possesses powerful communication, entertainment, and photography functions, innovation is not stopping. In the future, the basic functions of smartphones are expected to be improved continuously and comprehensively, in terms of 5G capabilities, high display refresh rates such as 90Hz, wider color gamut in display, and imaging systems reconstructed from image sensors. These advanced functionalities, in turn, will continue to strengthen the important role of smartphones in the 5G, IoT ecosystem.

In the longer-term, the smart home will evolve from one controlled by people to one in which artificial intelligence allows smart lighting, heating and security systems to learn the behavior of home dwellers, detect anomalies and autonomously act on this data.

Unleashing opportunities with the power of 5G, AI and cloud

Defining the opportunity

AI, the cloud and edge, connectivity (including 5G) and IoT are not, in themselves, markets, but the foundations for products, services and solutions for vertical industries and experiences for consumers.

As discussed in previous sections, AI, the cloud and edge, connectivity and IoT frequently work in tandem to fulfill customer requirements. For example, at the most basic level:

- **IoT** nodes with embedded sensors generate streams of data.
- Wired or wireless **connectivity** (including 5G) facilitates the transmission of this data from IoT nodes to the edge and cloud, and instructions back to the nodes.
- Data is initially collected at the **edge**; the **cloud** then hosts and applies deeper analytics, in the form of AI, to this data.
- **AI** algorithms learn how to identify objects and patterns from large data sets. This knowledge can be used to address business challenges and create new opportunities.

A familiar adage in the IoT is that "no single company can do everything". This adage equally applies to AI, the cloud and edge and connectivity. Semiconductor innovation is critical to the future development of all of these technologies. But the semiconductor industry itself is highly fragmented: Qualcomm plays a particularly important role in 5G connectivity. NVIDIA is a leading player in the AI arena, particularly for demanding automotive applications. STMicroelectronics, Renesas and NXP are integrating ML capability into new MCU product families that are optimized for power-constrained, low-cost IoT devices.

Telecoms operators deploy and operate the public network connectivity that allows for the flow of data from many types of devices to the cloud and back. With 5G and multi-access edge computing (MEC), which will see the deployment of application servers on infrastructure like base stations, operators have the opportunity to develop solutions that for customers that demand low-latency compute at the edge. AWS IoT Greengrass and Azure IoT Edge represent moves by the leading cloud vendors AWS (Amazon) and Microsoft to bring local compute and ML inference to edge devices. Across many IoT verticals, systems integrators apply the glue that joins multi-vendor hardware, software and systems and allows for the ingestion of IoT data into enterprise CRM and ERP (enterprise resource planning) databases.

Ultimately, the opportunities for the supply-side vendors will be shaped by their ability to meet the requirements of the demand-side: different consumer demographics and enterprises of different sizes across a multitude of verticals. Broadly speaking, enterprises of all types should be focused on remaining relevant in markets where technology innovation and globalization continue to disrupt traditional business models. Again, the main incentives for investing in IoT apply to the outcomes companies want from AI, the cloud and edge, namely:

- Increased efficiencies/reduced cost.
- New source of revenue, often involving the transition from a product-centric to a service-centric business model.
- An enhanced customer experiences.
- Compliance with regulatory requirements, e.g. data privacy, security and safety.

These objectives are not mutually exclusive: measures designed to increase efficiencies may underpin the later transition to an "as-a-service" business model.

Defining what consumers want, given their diverse characteristics in terms of age, gender, language, culture, income etc., is no easy task. Cost will always be an issue and has been a major factor in inhibiting demand for wearables like smart glasses and VR/AR headsets. Nevertheless, consumers of all types are increasingly looking for enhanced experiences: high-quality 4K streaming and 360 viewing of sports events are among the first 5G smartphone applications gaining in popularity.

In the minds of consumers, the growing importance of "experience" compared to device "look and feel" or monthly data requirements mean that device OEMs (original equipment manufacturers) need to shift their thinking from developing compelling products to facilitating great experiences over the lifespan of the device.

This thinking will have an impact on device design. For instance, 5G's high bandwidth and low latency capabilities will enable it to extract information from cloud-based applications, such as a CRM (customer relationship management) database, with little delay. By integrating 5G into laptops and notebooks, OEMs may consider reducing on-device storage and processing in exchange for improving battery life. Similarly, smartwatch developers will weigh the pros and cons of adding more advanced sensors into wearables and shifting storage to the cloud.

The opportunities and challenges of intelligent connectivity

Amid all of these complexities, what are the opportunities for intelligent connectivity and what challenges do vendors face? At the outset, it's important to recognize that many enterprises on the demand-side lack the internal skillsets to develop, in-house, new products, services and solutions leveraging the capabilities of AI, the cloud and edge, connectivity (including 5G) and IoT. Instead, they will be focused on the desired outcomes: return-on-investment (ROI) and compliance. Achieving these outcomes can be challenging. For instance, the technical complexity and unforeseen costs associated with deploying, securing and managing IoT devices has frequently prevented IoT projects moving from POC stage to deployment at scale.

In the future, new products, services and solutions will leverage many, if not all, of the building blocks of intelligent connectivity: AI, the cloud and edge, connectivity (including 5G) and IoT. Below, we map out some promising applications that will leverage these capabilities, as well some significant challenges.

Consumer opportunities: smartphones, digital assistants and entertainment

The consumer opportunity is huge and varied. According to IHS Markit forecasts, by 2023 around a quarter of the 1.7 billion mobile handsets shipped globally will integrate 5G. Increasingly, smartphones will be AI-capable: IHS Markit projects that every two out of three smartphones will have prebuilt AI hardware and features by 2025. Even at the early stage of AI, it is driving an increasing number of attractive features like facial recognition and unlocking, smile and people identification in photos and the enhancement of images. Features like emotion recognition and user profiling are in the development pipeline.

Digital assistants in the form of smart speakers will continue to be in demand. IHS Markit estimates that nearly 70 million were shipped in 2018 and the global installed base of smart speakers will rise from over 100 million that year to over 800 million in 2023.

AI-capabilities will be increasingly integrated into consumer entertainment devices. The aim will be to deliver more personalized, immersive experiences to adults and children living in the family home. For instance, AI-capable cameras, twinned with smart TVs or streaming boxes, will use facial recognition to launch personal viewing preferences and recommendations. AI will also be able to identify key moments, such as goals, in recorded sports programs. This will allow time-constrained sports fans to skip the boring periods of play and fast forward to the action.

Such personalization will make it easier for parents to prevent their children from watching age-restricted content. In these examples, security and data privacy are critical. Consumers are understandingly wary of allowing their facial images, behavior and preferences to be transmitted to and stored in the cloud. Therefore processing data at the edge will not only facilitate rapid facial recognition: it will also ensure that personal data is ringfenced.

Cloud gaming and VR/AR encompass several use cases where 5G, cloud/edge and AI could play a significant role. Cloud gaming allows users to access games whenever they want on a screen of their choice. This has long been held up as the theoretical end game for the distribution of games content and services – a global market worth over $100 billion in 2018 according to IHS Markit.

The ability to access from any screen lowers the barriers to entry for consumers to access high-end experiences by removing up-front hardware costs – an expensive console or gaming PC. 5G will allow for instant access on compatible smartphones to games hosted in the cloud; network slicing incorporating edge compute will address the need for latency of <100 millisecond latency for multiplayer online battle arena (MOBA) and latency of <20 millisecond latency for highly interactive games. AI could potentially play several roles in cloud gaming, from enhancing the quality of game streaming over time through analysis of previous streams and allowing users to generate virtual scenes.

The lack of high-end headsets has also limited the potential for broadcast VR – which, for instance, allows users watching sporting events to choose camera angles or point at players to receive stats. Cloud streaming over 5G potentially reduces the need for costly hardware, although gaps in 5G coverage will mean that home audiences will not always be able to access such immersive experiences. However, dedicated 5G coverage at the venue or at a second scene (such as an entertainment venue) represent interesting new revenue opportunities.

Video surveillance opportunities

IHS Markit forecasts that the market for AI-enabled video surveillance equipment will grow from a sub-$1 billion opportunity in 2018 to a $10 billion opportunity by 2023. The enablers and drivers of such growth are numerous:

- The industry transition from standard to high definition video.
- Growing demand for intelligent analytics.
- Smart and safe city initiatives.
- The development of AI deep learning (DL) analytics, facilitating a significant leap in the level of accuracy and reliability in object and behavior classification.
- The increasing integration of connectivity into video surveillance cameras, providing massive scaling opportunities for the surveillance market.

Video surveillance provides an interesting case study into the complementary functions of the cloud and the edge. A hybrid approach is being developed where analytics workloads are more distributed using a mix of smart cameras at the edge combined with centralized servers and cloud-based analysis.

This means some analytics can be run at the camera, for example, crowd monitoring, counting, and object detection, to save bandwidth and relieve the computing capacity at the back-end. This trend is expected to grow further thanks to the availability of more advanced semiconductors with low-power capabilities. More powerful centralized analytics could be used to run processor-intensive applications such as feature extractions of human/vehicle and object searching. This approach has gained interest in the Chinese market, where AI deployments have been the greatest and where projects are large and can benefit from the distribution of compute and analytics the most.

Automotive opportunities

AI, the cloud and edge, connectivity and IoT are top of mind for both car manufacturers and the broader automotive supplier ecosystem. There are several reasons for this.

- **Consumer behavior is rapidly evolving**: The growing popularity of mobility-as-a-service providers such as Uber and Lyft is a disruptive force to the traditional business models of automakers – selling vehicles. By the end of last year Uber had 91 million monthly active platform consumers worldwide (and 3.9 million drivers), while Lyft surpassed 20 million active riders (and 2 million drivers) earlier this year. Volvo is one of several automakers to develop their own mobility-as-a-service offering in a bid to counter this disruption. More broadly, automakers are seeking to differentiate through adding features that make the in-car experience enjoyable and journeys less troublesome.

- **Data from the connected car is fueling new business models**: Data generated from cars and their drivers are opening up tangible new business models. One example is usage-based (or pay-as/how-you drive) insurance (UBI). Unlike traditional motor insurance, on which premiums are based on things like age, gender, credit score and the number of no-claims years and related personal and demographic factors, usage-based premiums are based on actual driver behavior.

  UBI is enabled via the installation of a cellular-connected black box in a car to track driver behavior (such as acceleration, braking, cornering, and distance, time and route of travel). Various insurers are also developing and applying algorithms to identify distracted driving due to device usage and other forms of behavior. For the consumer, UBI can result in lower premiums. Potential benefits for the insurer can be quicker and less costly claims management, a reduction in fraudulent claims and greater scope for positive engagement with the user (e.g. driving tips and incentives).

- **We are on the road to autonomy**: The vision of the future sees fully autonomous vehicles eventually being able to make their way safely from A to B. That means among other things, not getting lost or mounting the curb into the pavement; stopping at traffic lights; indicating when changing lanes or turning; and breaking to avoid collision with other cars on the road and pedestrians crossing the road.

All of this requires cars to be able to detect and recognize many different types of objects in different types of conditions: night or day, cloudy and raining or sunny and bright. Such requirements are prompting the entire automotive supply chain, from semiconductor suppliers up to OEMs, to seriously invest in AI and DL. Some of the challenges they need to address are algorithm development for tasks such as recognition/prediction of actions; power consumption to be optimized for electric vehicles (EVs) and, inevitably, safety.

As of today, ML is already performing some functions in a few vehicles. In infotainment HMI, most of the speech recognition technologies rely on algorithms based on neural networks running in the cloud. New premium brands have also launched a hybrid approach offering, on top of cloud-based support, embedded hardware able to perform voice recognition in the absence of wireless connectivity.

In ADAS, deep learning applies to realms like camera-based machine vision systems and radar-based detection units, driver drowsiness, and sensor-fusion electronic control units (ECUs). Through sophisticated radar systems and multiple cameras, cars learn how to self-park, detect pedestrians or other objects, and sense dangerous situations to avoid collisions.

In driver drowsiness, deep learning detects cognitive distraction caused not only by irregular facial or eye movements but also by erratic driver behavior and biological patterns. Here, algorithms analyze time-series data about the car (e.g. steering) and data on the driver (e.g., heart rate), to detect driving behavior deviating from appropriate or expected patterns. The vehicle then issues a warning in the form of an alert or the car's AI directly intervenes and takes control to, say, steer the vehicle back to its lane.

Factory opportunities

The pressure to increase capacity and improve efficiency and yield has driven the implementation of factory automation for many decades. On paper, the combined capabilities of AI, the cloud and edge, connectivity and IoT would deliver above and beyond previous improvements in capacity, efficacy and yield.

However, key challenges to adoption of these transformative technologies are the long equipment life of installed equipment; an often-cautious approach to introducing new technologies and concern over the security of mission-critical factory data and equipment. Nevertheless, there is growing recognition in the industry that AI and other technologies can bring tangible benefits across many areas and applications: maintenance, testing and simulation, quality control, autonomous machines/robotics, customer interface, inventory management, equip ment inspection, and process control.

The question facing industrial companies is not if, but how to integrate these new technologies in their operations. On the one hand, introducing limited new functionality to existing systems could be achieved through retrofitting. On the other, more costly investment in greenfield facilities and new equipment leveraging AI, IoT and other technologies will deliver richer functionality and greater benefits. This requires deep thinking and long-term strategy and ROI planning. But considering the huge number of potential use cases depending on industry verticals and subdomains, a wide spectrum of implementation options will be available between these two extremes.

From the implementation perspective, edge computing continues to gain a lot of attention in IIoT (industrial IoT). The need for ensure data security and meet high performance and requirements for real-time operations are driving this interest. There are a number of different potential hosts for edge computing in the factory environment, from industrial PCs (IPC) programmable logic controllers (PLCs) to on-premise datacenters and routers and gateways. IHS Markit forecasts that manufacturing industries will account for over 5% of the spending for Edge datacenters, by 2022.

A note on 5G challenges

It's a 5G world, already. Commercial launches are spreading across the planet. The ecosystem of devices is growing fast. South Korea's operators, which have deployed 5G in the 3.5GHz and the 28GHz bands, have attracted over 3 million 5G subscribers since launching commercial services in April 2019. This is despite some customer disquiet over quality of service, particularly regarding initial latency degradation when users switch from LTE to 5G (up switching). Operators at the vanguard of 5G have reported early customer engagement with 4K video streaming, cloud gaming and XR (e.g. 360 degree viewing of sport events).

Like all of the technologies featured in this report (AI, the cloud and edge, connectivity and IoT) 5G is fast evolving. Similar to most of them, it is at a somewhat early stage of development. Significant progress is needed to improve and increase coverage, add new features like network slicing and to hit challenging technical objectives around very low latency and very high availability. The 3GPP roadmap provides some clarity over what to expect from Release 16 and 17, not least C-V2X (vehicle-to-vehicle/infrastructure/pedestrian/network), one of the stepping stones towards an autonomous future.

Technology readiness is not the sole determinant of the adoption and ultimate success of a technology, but one of many factors. 5G is an interesting case because it has been pitched from early on its' development as being potentially transformative for many industry verticals – some of which have little legacy of adopting cellular. These are the main non-technology factors that IHS Markit believes will influence the rate of adoption of 5G in industry verticals:

- **The legacy adoption of cellular**. Cellular adoption has been historically strong in automotive (vehicles) but weak in the smart home and industrial automation. In the latter area, machine connectivity is dominated by private wired technologies. Cellular adoption has been hampered by the need to meet strict performance requirements and the perception that public mobile network connectivity may increase the risk of sensitive data being exposed beyond the factory walls. Nevertheless, there has been growing interest in recent times in private cellular for certain applications and 5G.

- **Regulation and compliance**. Drones are a good example of the impact of regulation. Companies across the agricultural, energy, logistics and transportation sectors have shown particular interest in using drones to perform a variety of tasks. But restrictions on beyond-visual-line-of-sight (BVLOS) flights has, to date, limited use cases and the requirement for wide area network connectivity.

- **Equipment/device replacement rates**: Device/equipment life is lengthy in many industrial use cases (industrial automation, metering and remote asset monitoring). Some of these devices are connected with alternative technologies – such as the unlicensed low-power wide area network (LPWAN) technology LoRaWAN. In some cases, there will be a retrofit opportunity for unconnected assets.

- **Use cases, ROI and business models**: Establishing viable use cases, demonstrating ROI and business model fit are perhaps the main challenges facing 5G as it seeks to penetrate new industrial opportunities. Determining ROI can be particularly challenging owing to the complexity involved in integrating different hardware and software from multiple vendors, deploying devices, managing them in the field over years, and ingesting IoT data into internal enterprise systems. Lengthy POCs can be expected.

Historically, telecoms operators have monetized connectivity based on usage – minutes of voice, number of messages, amount of data consumed. But recurring, OPEX-based pricing does not match the capex-centric budgeting of many companies in the industrial space. They may be willing to move towards more flexible budgeting methods. However, they may prefer to pay for guaranteed performance and service levels rather than a traditional metric like usage.

In the conclusion we'll discuss how vendors in the AI, the cloud and edge, connectivity (including 5G) and IoT can address some of these challenges through business model flexibility and open collaboration.

Conclusion – developing the ecosystem for intelligent connectivity

Intelligent connectivity implies the collaborative efforts of AI, the cloud and edge, connectivity (including 5G) and IoT to bring value to enterprises and consumers. This value may refer to lower operational costs, access to new sources of revenue, an enhanced customer experience or better compliance with regulatory requirements.

The supply side consists of many thousands of vendors and developers working across one or more of these technologies. Ultimately, their success will be shaped by their ability to understand the current and future pain points and opportunities of the demand-side (different consumer demographics and enterprises of different sizes across a multitude of verticals) and execute effectively on well thought out strategies.

It's no small task to make generic recommendations that apply to many different use cases, but the following principles are broadly relevant:

- **Think solutions and experiences, not products**. At the outset, product design should be shaped by customers' current and future requirements. Physical products are not the end-game, but the enabler of great consumer experiences and the solution to enterprise and industry pain points and opportunities. No single company holds the crystal ball on customer needs. Therefore, collaboration across the supply chain and ongoing engagement with customers is critical.

  The focus on solutions and experiences, coupled with integrating the capabilities of intelligent connectivity technologies, could fundamentally change human interaction with physical devices and pave the way for new, smaller, form factors in the coming decades. For instance, AI that reacts to non-verbal behavior in the home could mean that things as diverse as lighting, heating and entertainment are automated. Providers will need to think carefully how far automation goes – humans may appreciate reduced hassle but will not want to cede too much control to machines.

- **Solutions need to evolve over time**. This thinking is not new. Updates to device firmware and software and applications add incremental improvements in features and address ever-changing security threats. But in the future, customer will expect more than incremental improvements to their experience over a device's lifespan. Companies across all verticals are facing ever more pressure to make their products and processes sustainable, which will mean enabling great experiences over longer device lifespans. The ability of AI to recognize how users consume services and the quality of experience will play a key role in optimizing how these and delivered and creating new personalized experiences. In other words, AI will help to help configure the other building blocks of intelligent connectivity: connectivity, the cloud/edge and IoT.

- **Take the open road**: The engagement of third-party developers has been a key factor behind the success of Apple's App Store and Google Play In the connectivity space, open protocols such as Bluetooth, Wi-Fi and cellular, rather than proprietary technologies, have seen the most success. Proprietary technologies tend to stifle interoperability and restrict the scope for vendors to benefit from innovation outside of their companies. Indeed, internal silos within companies can restrict data sharing, collaboration and the development of strategically important projects.

  Openness and collaboration, both within companies and with partners, is the best route to understanding and predicting customer pain points and opportunities. This means, among other things, open APIs, developer support initiatives and flexibility. Solution vendors will need to increasingly meet customer demand for a cloud-agnostic or multi-cloud approach and not restrict customers to their own software, but partner with innovative new players to add in new features to their solutions.

- **Business models need to fit and be flexible**: Traditional service providers like telecoms operators and OEMs developing service-based solutions on top of their products need to tailor their business models to the process or outcome they support for the customer. To give two examples from the industrial sphere. First, an industrial company that requires very high bandwidth and availability and very low latency to monitor and control a piece of machinery will primarily be concerned that connectivity is reliable and resilient.

Therefore, a pricing model that reflects these outcomes, rather than simply data volumes, would be a better fit. Second, another industrial company monitoring a remotely located piece of equipment wants to receive an alert if it starts malfunctioning. The frequency of communication and amount of data transmitted is modest. For the industrial company, the value in connecting the equipment, communicating and analyzing the data, will be in knowing when to send a technician to prevent unplanned and costly downtime.

The vision of 5G sees many new different types of devices and experiences, such as multiplayer cloud gaming and AR/VR, directly enabled by mobile technology. In the past, advanced gaming experiences have relied on high-cost, dedicated devices. The low latency, high bandwidth characteristics of 5G will enable multiplayer cloud gaming on the smartphone form factor. These technical facets will also allow for more immersive VR experiences than were achievable through Wi-Fi connected headsets.

Device makers and their content and telecom operator partners will need to think creatively about how to monetize hardware, connectivity and experiences. For example, telecom operators have the opportunity to bundle third-party content, monetize the creation of audiences from their customer base for streamed entertainment events and offer Network-as-a-Service (NaaS) for events that require dedicated bandwidth. AI will be foundational for many of these initiatives: from identifying audiences for streamed events and determining the level of network performance required. Content providers will explore options such as zero-rating deals with telecoms operators, at least for a promotional period. Device makers will explore the potential for bundling the price of connectivity upfront, into hardware cost.

Contacts

Tom Morrod
Research and Analysis Executive Director
Tom.Morrod@ihsmarkit.com

Julian Watson
Principal Research Analyst
Julian.Watson@ihsmarkit.com

Wang Shen
Senior Consultant
Shen.Wang@ihsmarkit.com

IHS Markit®