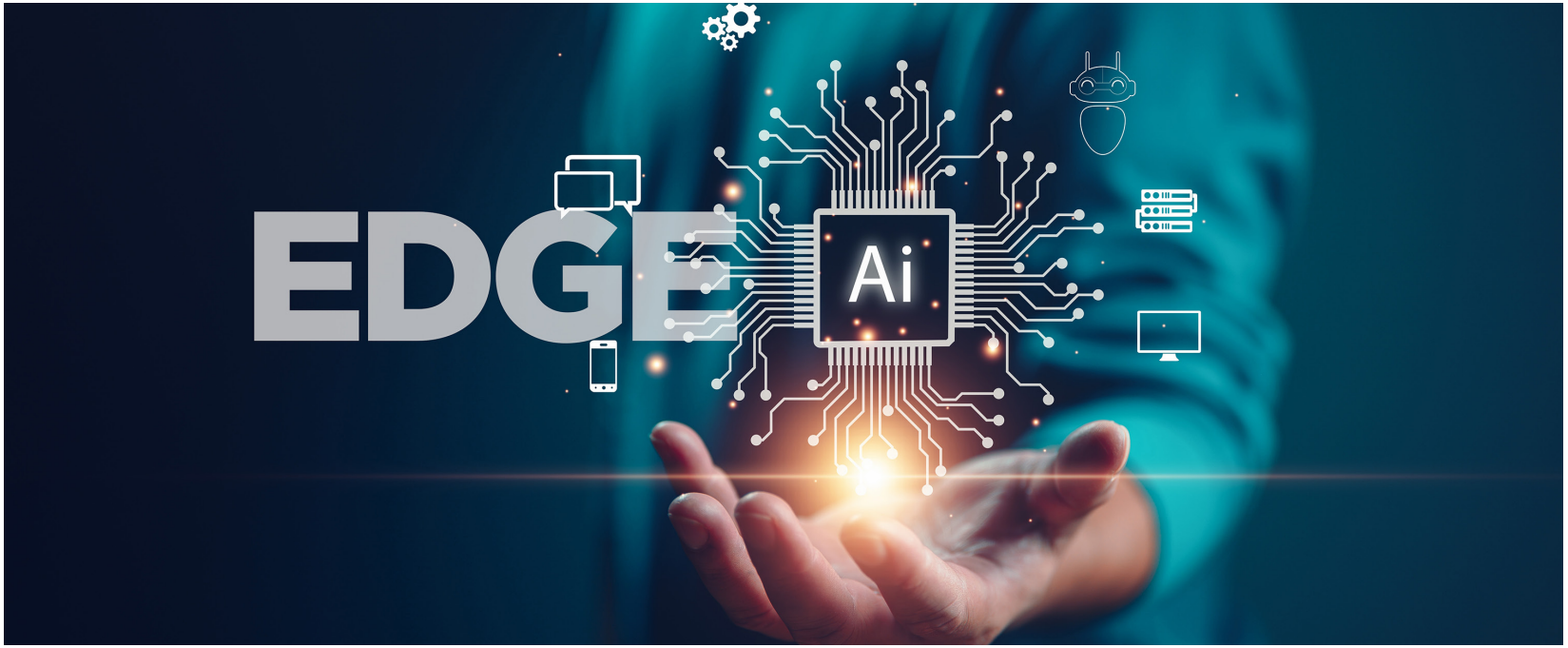
Leading the 5G IoT Race: Key Industry Insights Series 4. Edge AI

By **Robin Duke-Woolley**, CEO, Beecham Research

Edge AI is one of the most notable sectors of artificial intelligence, providing the ability to run AI processes without the latency issues of data transmission to the cloud. Edge AI is enabling greater, more widespread use of AI, providing the means for smart IoT devices to respond quickly to new inputs and to new application opportunities.



1. What is Edge AI and why do we need it?

Edge AI is the deployment of AI applications in devices at the network edge throughout the physical world. It is called "Edge AI" because the AI computation is done near the user at the edge of the network, close to where the data is located, rather than centrally in a cloud computing facility or private data centre.

Cloud AI is where data processing and storage occur in centralised data centers managed by cloud service providers, usually situated far away from end-users. By offering cloud computing service such as servers, storage, databases, etc. over the internet to the cloud, it offers flexible resource accessibility and economics of scale.

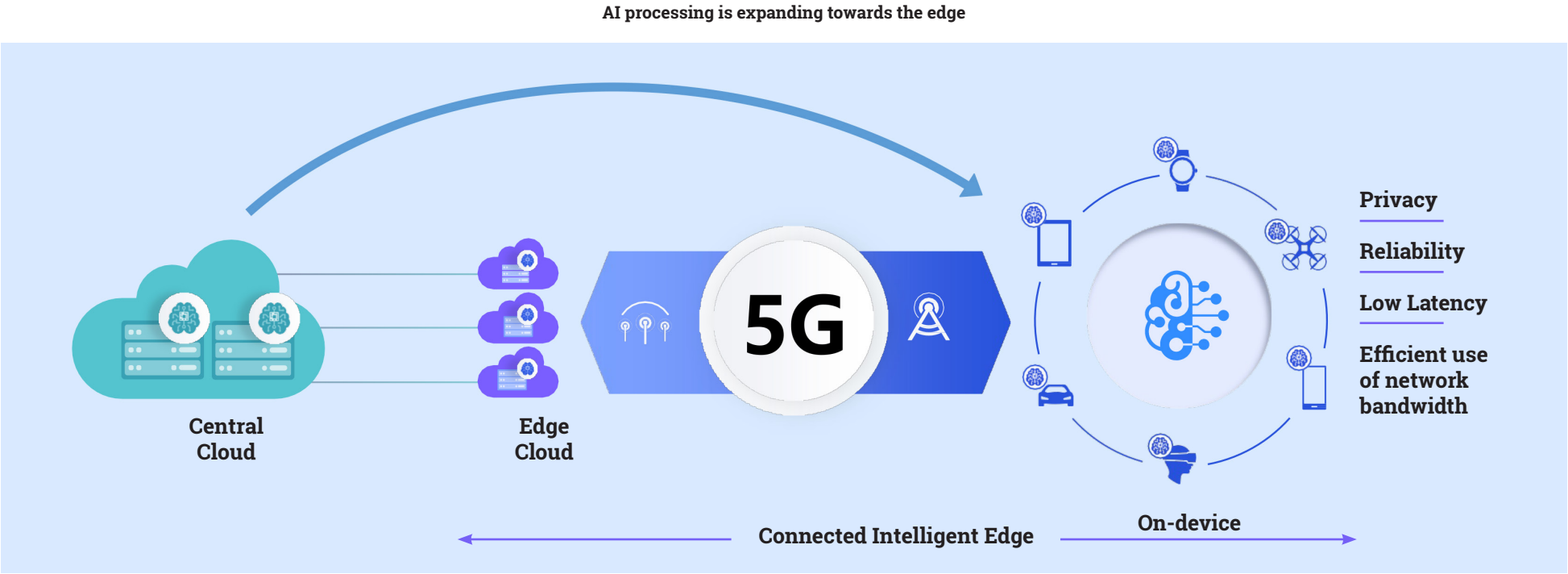


Figure 1: Moving AI from the Cloud to the Edge

Edge AI and Cloud AI are therefore complementary technologies. Particular advantages of Edge AI include the following:

i. Lower data transfer volume
Data is processed by the edge device, and only a significantly lower amount of processed data is sent to the cloud. By reducing the traffic amount across the connection between a small cell and the core network, the connectivity cost can be reduced.

ii. Speed for Real-time computing
Real-time processing is a fundamental advantage of edge computing. The physical proximity of edge devices to the data sources makes it possible to achieve lower latency which improves real-time data processing performance. It supports delay-sensitive applications and services such as remote surgery, tactile internet, unmanned vehicles, and vehicle accident prevention. A diverse range of services, including decision support, decision-making, and data analysis, can be provided by edge servers in a real-time manner.

iii. Privacy and security
Since transferring sensitive user data over networks makes it vulnerable to theft and distortion, running AI at the edge enables keeping the data private. Edge computing makes it possible to guarantee that private data never leaves the local device. For the cases where data must be processed remotely, edge devices can be used to discard personally identifiable information before data transfer, thus enhancing user privacy and security.

iv. High availability
Decentralization and offline capabilities make Edge AI more robust by providing transient services during a network failure or cyber-attacks. In this way, deploying AI tasks to the edge ensures significantly higher availability and overall robustness needed for mission-critical AI applications.

v. Cost and energy efficiency
Moving AI processing to the edge is highly cost efficient because only processed, highly valuable data is sent to the cloud. Sending and storing huge amounts of data to the cloud is costly, while processing locally at the edge saves these costs and is also more energy efficient.

vi. Scalability and Adaptability
Edge AI provides the opportunity for scalable and adaptable solutions that can cater for diverse edge devices and evolving application requirements. It focuses on developing lightweight and efficient AI algorithms that can run effectively on resource-constrained edge devices, ensuring flexibility and compatibility across different hardware platforms. In this way, it provides a natural path for scaling IoT solutions and for enabling mass deployments.

These advantages offer substantial enhancement of IoT applications and, as a result, the opportunities are huge. Figure 2 shows Beecham Research estimates drawn from several different analyst forecasts – all pointing towards high growth of over 24% per annum in global Edge AI revenues during the period 2022 through 2030.

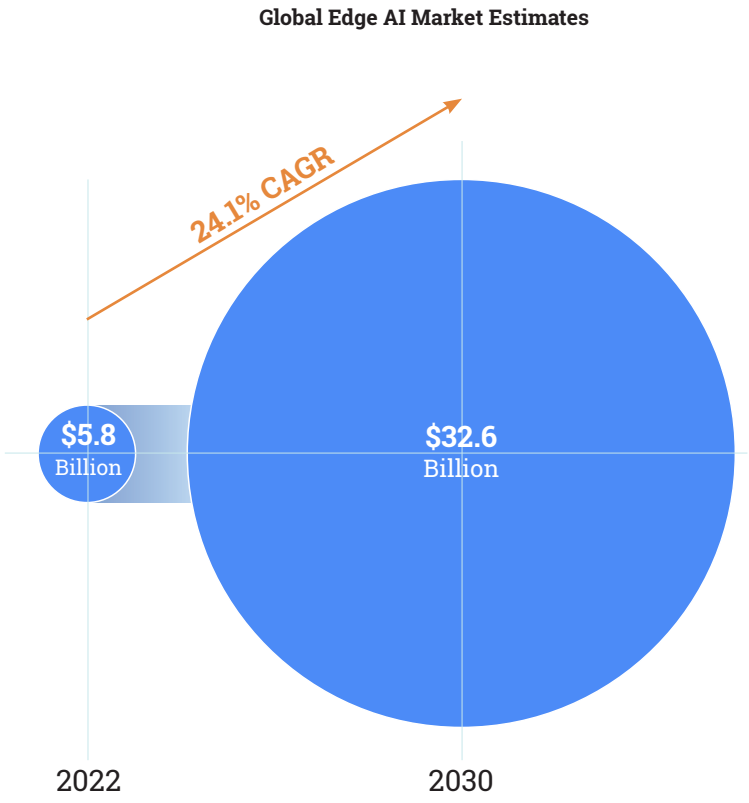


Figure 2: Global Edge AI Revenue Projection

(Source: Beecham Research)

2. Where will Edge AI be used?

Edge AI will be used in all IoT sectors for a wide range of applications that will evolve over time. To give an idea of these, Figure 3 shows 9 application scenarios where Edge AI is already evident.

Figure 3: Key Application Scenarios for Edge AI

A brief outline for some of these is as follows:

- **Smart Homes.** Edge AI allows smart home devices to perform voice recognition, natural language processing, and activity recognition locally. It enables faster response times, enhanced privacy, and local automation of various tasks within the smart home ecosystem.

- **Healthcare Monitoring.** Edge AI enables real-time monitoring and analysis of health data from wearable devices, such as heart rate monitors or glucose sensors. It allows for immediate detection of abnormal patterns, personalized healthcare recommendations, and timely alerts for medical interventions.

- **Smart Manufacturing.** Edge AI plays a vital role in smart manufacturing applications, such as predictive maintenance, quality control, and robotics. It enables real-time analysis of sensor data, anomaly detection and autonomous decision-making at the edge, improving operational efficiency and reducing downtime.

- **Outdoor Robots and Human-Robot Collaboration.** Edge AI plays a major role in enabling human-robot collaboration applications by bringing AI capabilities directly to the edge devices and robots involved in the collaboration. It allows robots to analyze and process data in real-time, make intelligent decisions, and interact with humans efficiently, without relying heavily on cloud resources. The capabilities of outdoor robots are especially important for outdoor operations in harsh and hazardous environment, where human safety can be protected by using robots. Edge AI facilitates various aspects, such as perception, decision-making, and interaction. For example, edge AI algorithms deployed on robot platforms can enable real-time object recognition, tracking, and gesture recognition, enabling robots to understand and respond to human commands and actions.

- **UAVs.** Edge AI enables a whole raft of new capabilities for UAVs. Critical technical challenges of UAVs are autonomous navigation, power management, security and privacy, formation control, computer vision, and communication. Autonomous navigation can be further divided into localization and mapping, path planning, and collision avoidance systems.

3. Outdoor Robots

To illustrate the potential growth for these, the following looks at one of them – outdoor robots – in more detail.

As shown in Figure 4, outdoor robots is a relatively small business at present with a lot of experimenting taking place in terms of application development and shape. Global revenues amounted to around $163 million in 2022, with expectations for strong growth over the next decade to $701 million in 2032 – an annual growth rate of 15.7%. There are three principal types of outdoor robot:

- Delivery Robots, including UAVs and autonomous vehicles that are increasingly being used for last-mile deliveries.
- Security and Inspection, including patrolling secure premises, inspecting infrastructure and pipelines for potential faults and harsh terrain.
- Guide, including use in hospitality and retail environments to guide visitors and provide information.

Of these, the largest and accounting for more than 50% in value is Delivery. Security and Inspection is the next largest, with Guide being the smallest category.

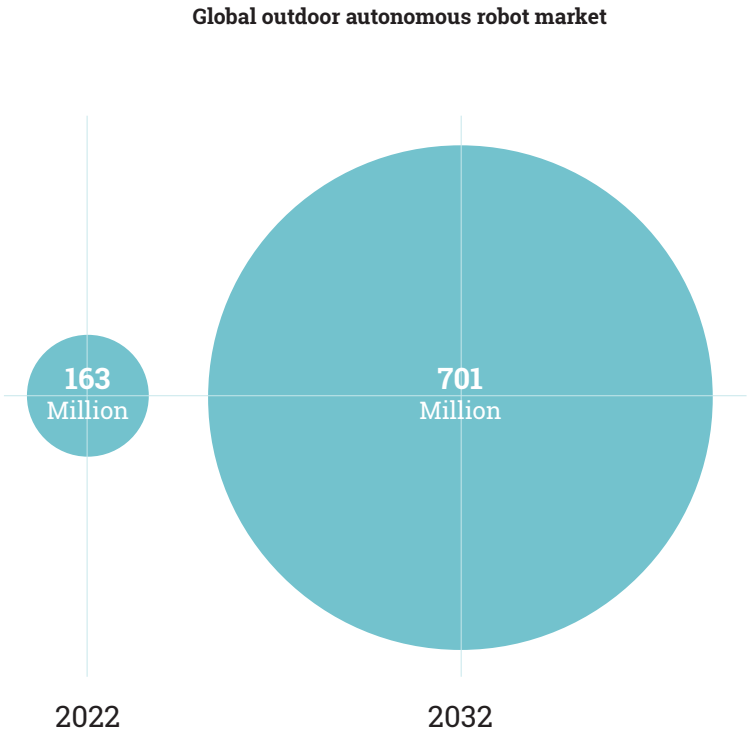


Figure 4: Outdoor Robots Revenue Forecast, with 15.7% CAGR in the period

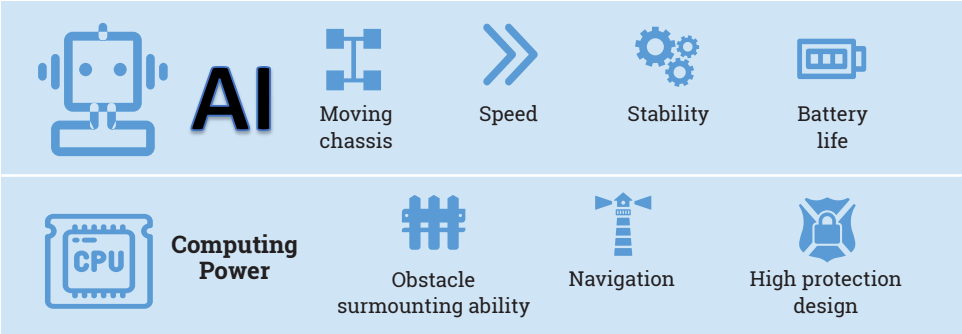
(Source: Grand View Research)

The main application areas are developing rapidly and include factory parks, ports, the power industry, airports, mining, geological prospecting, street patrol and smart agriculture.

There are several key reasons for this expected growth:

1. Increasing government initiatives and investment in robot manufacturing companies.
2. Increasing demand for autonomous last mile delivery.
3. Huge growth in e-commerce and online food industry.
4. Rising demand for security and guide robots.
5. Growth in hospitality and healthcare industries

Key Technologies for Outdoor Robots



Layout of Outdoor Robots

Figure 5: Key Technologies for Outdoor Robots

The technologies required for these robots are illustrated in Figure 5. These combine computing power and decision-making through use of AI with the mechanics to move, sense and navigate.

Fibocom's approach to robotics focuses on three main technology areas – sensing, motion and intelligence. These are referred to as the three robotic core technologies:

- Firstly, there are the components used to realise sensing – different types of sensors, cameras, then ways to detect the environment and other objects. The computing and AI activities associated with these are vSLAM (visual Simultaneous Localization and Mapping), used for mapping, path planning, positioning features and other functions.
- Secondly, the components for motion and navigation. These include the steering engine, electric machinery, torque control, motion control and overall power management. The computing and AI activities associated with these are related to Navigation, including object recognition, avoidance, slope control and boundary distinguish.
- Thirdly, the elements for intelligence and decision-making. Components for this include CPU/GPU/DSP/NPU, the Robot OS, operator optimization and deep learning. The activities are headed under Performance and include speech processing, image recognition, machine learning, Natural Language Processing (NLP), User Interface interaction and the decision-making system.

Fibocom's solution is for a Robot Intelligent Chassis upon which application providers can then mount their overall robot shape (working body) and application (interaction body) for their individual needs. The functional structure of the robot is then divided into 5 parts:

- Motion body
- Sensor body
- Decision-making body
- Working body
- Interaction body

The core of the robot project is the perception/sensing body and the decision-making body. Different chassis (motion body) and operation body (working + interaction) constitute different types of robot. This maximises the common elements across all robot types.

4. Fibocom module range and case studies for Edge AI

Fibocom offers a set of three modules for Edge AI:

(a) Smart Module SC171 for PCBA AI Solution for Robotics – 5G Smart high level solution

The SC171 module supports 5G sub-6GHz connectivity and is based on Qualcomm's QCM6490 SoC. This supports the industrial grade PCBA solution with high computing power, suitable for complex indoor and outdoor scenarios, and can be applied to various types of robots such as logistics vehicles, patrol inspection robots, guiding robots and more.

Advantages:

- The PCBA solution with SC171 series supports 12 TOPS of AI computing power, which has excellent AI performance, supports 4K high-definition video, supports up to 5 cameras running simultaneously, and efficiently calculates and processes data.
- It has a variety of peripheral interfaces, including 8 * USB, 4 * MIPI, 8 * UART, 2 * SPI, 2 * LAN, 2 * CAN, and more ports, supporting intelligent and multiple application scenarios.
- Industrial-grade design: Industrial-grade Ethernet and various industrial-grade connectors.

(b) Smart Module SC138 – 4G Smart medium level solution

The SC138 module supports 4G connectivity and is based on Qualcomm's QCM6125 chipset platform. It has early customers in the Chinese market for devices including: Patrol robot for smart grids, industrial campus, logistics warehouse, hospital and more.

(c) Smart Module SC126 – 4G Smart entry level solution

The SC 126 module also supports 4G connectivity and is based on Qualcomm's QCM2290 chipset platform. An early customer is a Hong Kong stock listed Robot manufacturer, with an automatic lawn mower for home use.

Fibocom 4G/5G Smart Module Series

SC171 **SC138** **SC126**

Figure 6: Smart AI module SC171 with high computing power and PCBA Solution

5. One-on-one interview with Fibocom

Robin Duke-Woolley
CEO
Beecham Research

Shirley Tang
Director of MC
Product Marketing
Fibocom

Robin Duke-Woolley (RDW), CEO of Beecham Research, interviewed Shirley Tang (ST), Director of MC Product Marketing at Fibocom, to understand in more detail how Fibocom is introducing EdgeAI to the IoT market.

RDW: Looking at different application areas, What is it that Edge AI is doing? Let's look at three quite different sectors as examples – Smart Home, Healthcare and Smart Manufacturing

ST: Edge AI is the concept of applying artificial intelligence (AI) capabilities to the edge computing environment. It combines edge computing and artificial intelligence technology to enable intelligent data processing and decision-making on edge devices. Edge AI utilizes machine learning, deep learning, and other AI technologies to enable edge devices to autonomously perceive, understand, and respond to data without the need for continuous reliance on cloud or central servers. This can achieve real-time intelligent decision-making, intelligent control, and intelligent interaction, bringing a higher level of intelligence to edge environments.

Smart Home: A lot of sensors, cameras, smart home appliances and other devices need to be collected for data and do data analyzing, processing, and feedback.

Healthcare: Real time monitoring and remote healthcare to provide patients with more convenient and accurate medical services; Intelligent diagnosis and decision-making assistance; Timely emergency rescue and crisis management.

Smart Manufacturing: Through AI technology, it is possible to monitor and analyze various aspects of the production process, identify problems in a timely manner, and provide solutions; it can analyze and predict a large amount of data, providing a reasonable production plan; as well as to do equipment maintenance and supply chain management.

RDW: With AI being done at the edge, what do you need the connection to the cloud for? What are the characteristics of the connectivity required compared with AI in the cloud?

ST: For higher performance computing and storage capabilities to meet the needs of large-scale data processing and model training; Because the data and algorithms done at the edge are limited, it is non-complex and non-large-scale applications, mainly to improve the data processing and decision-making efficiency. The connectivity is faster and responds more quickly and can also save more bandwidth compared with Cloud AI connectivity.

RDW: How much computing power does EdgeAI need? What does that mean for types of applications that can/cannot be addressed?

ST: basically 1 TOP for entry level applications (sweeping robots, lawn mower, smart speaker), more complicated scenarios require more powerful computing capability:

- Smart IVI, PDA: 3TOPS to 9 TOPS
- Video conference, Live streaming: more than 10 TOPS
- Machine vision: 50TOPS
- Automatic driverless cars: more than 100TOPS

RDW: What degree of local learning can be achieved at the edge? Will this increase over time?

ST: actually, most of the daily AI applications can be achieved at the edge such as simple speech recognition (smart speaker, remote controller), facial recognition within a certain range (such as access control for a company), limited route planning; as well as some indoor robots, such as sweeping robots and lawn mower.
AIGC Big Model Training: needs cloud AI computing
For Vertical industries and segmented fields: can be done at the edge

Yes, this will increase over time, because the CPU, GPU and NPU will become increasingly advanced and powerful; and applications in vertical industries and segmented fields will also become increasingly powerful.

RDW: Which application scenarios do you think will be most important in the next few years?

ST: Robotics, Smart manufacturing, Healthcare and Automatic driverless cars

RDW: The trend is towards outdoor robots from indoor robots. Why is that?

ST: Objectively speaking, most of the environments we live in are forests, lakes, oceans, and deserts. There is a lot more 'outside' to contend with. More specifically there are:

- Intensification of the aging trend
- The problem of difficult employment
- Some harsh working environments
- Improve work efficiency in some environments
- Ensure the safety of workers

RDW: What are the facilities for outdoor robots that need to be built into the cellular modules?

ST: Mainly for GPS positioning, Real-time data / video uploading, data processing and remote monitoring, as well as Real-time emergency response.

RDW: Fibocom offers SC171, SC138 and SC126 cellular modules. What are the main differences between these?

ST: Modules SC171 (5G Smart) and SC138 – medium, SC126 – entry (4G Smart). From entry level to high level, from 4G to 5G, CPU and GPU performance, and TOPS are more and more powerful. So, the applicable product forms and application scenarios are also different.

RDW: What is Fibocom's Solution for Robot Intelligent Chassis?

ST: For robotics, Fibocom provides the body and chassis to our customers which we call our core robotics solution. This has three parts – sensor body, motion body and decision-making body. we call it Robot Intelligent Chassis, which is the core of robotics. And then different structural shells are put on to form different shapes for different robots.

6. Summary

Edge AI revenue is now expected to develop quickly over the next few years, with average growth of over 24% per annum. Applications being enabled by Edge AI in all IoT sectors and these will evolve further over the next few years. Fibocom is addressing this growing need with three different modules – SC171 for high level 5G applications, SC138 for medium level 4G applications and SC126 for entry level 4G applications. SC171 can support 12 to 13 TOPS of processing capability, while SC138 can support around 1.5 TOPS and SC126 around 0.3 TOP. Case studies for each of these working in the field are already available. These modules also support Open CPU, or module plus MCU as preferred by the customer.

In view of its perceived importance, Fibocom has invested further in outdoor robotics. The company now offers a Robot Intelligent Chassis comprising a Motion body, a Sensor body and a Decision-Making body. This chassis is suitable for use as a common robotics element for a wide variety of different robot applications, with the customer then adding the end application and shape.