HOW OPERATORS CAN MAXIMISE IOT REVENUE IN 2024

Whitepaper

TELCO & CONNECTIVITY

JUNIPER RESEARCH

1.1 Introduction

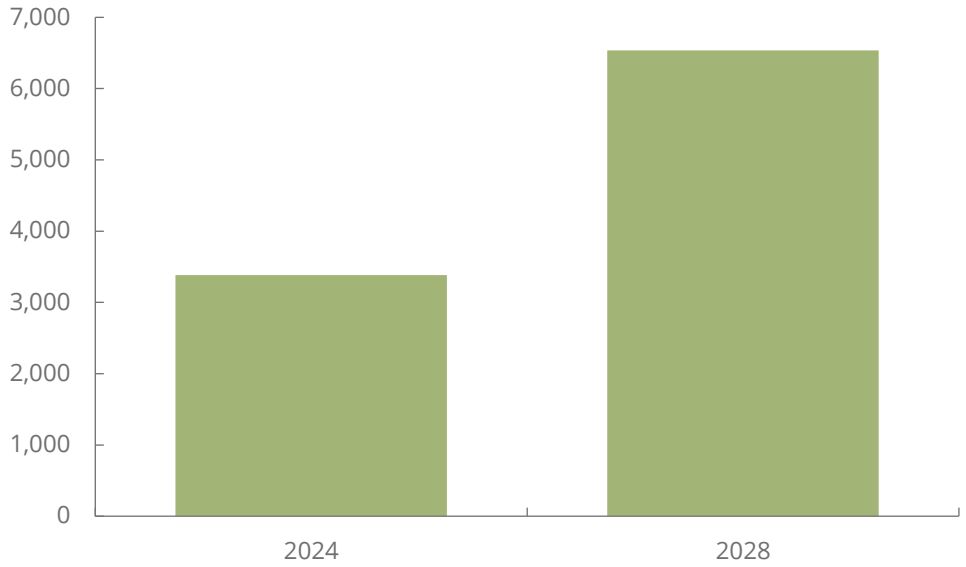
Juniper Research defines cellular IoT as:

'The connections, interfaces, and infrastructure allowing data to be transferred between machines, and between machines and individuals. Cellular IoT includes the management systems, interfaces, and portals needed to extract value from the machine data.'

The total number of cellular IoT devices will grow from 3.4 billion in 2024 to 6.5 billion in 2028. This represents an increase of 93%. However, the average operator revenue per cellular IoT device per month is forecast to decline over the same period of time.

This decrease will be driven by the stabilisation of 5G IoT pricing, the growing number of low-revenue LPWA (Low-power Wide-area) connections, and efficiency improvements to networks.

Figure 1: Total Number of Cellular IoT Devices (m), 2024 and 2028



Source: Juniper Research

As a result, connectivity represents a limited proportion of the total revenue generated, with the GSMA estimating connectivity to represent only 10% of total revenue from IoT. This is expected by Juniper Research to continue to fall, as average revenue per connection per month falls over the next four years.

Consequentially, operators are unable to solely rely on connectivity for IoT revenue, with most developing their own IoT management platforms, and value-added services. Likewise, IoT vendors have sought to expand their offerings to include these platforms and services to diversify revenue from IoT hardware shipments.

1.2 Key Monetisable Opportunities in Cellular IoT

This section will provide analysis, insight, and recommendations for how operators and IoT vendors can best maximise revenue from their cellular IoT services over the next four years.

1.2.1 3GPP Releases RedCap

In 5G Release 17, 3GPP introduced a new tier of IoT devices known as RedCap (Reduced Capability) or NR-Light (New Radio Light). Its introduction aims to bridge the capability and complexity gap between current 5G tiers, with an optimised design for mid-tier use cases.

RedCap devices can support 150 Mbps and 50 Mbps in downlink and uplink respectively, with support for lower transmit power. RedCap utilises narrow bandwidths, and provides lower-order modulation, and optional support for half-duplex, frequency-division duplexing.

Table 2: RedCap Capabilities

Capability	Description
Battery Life	RedCap use cases target a battery life of a few years for industrial sensors, and one to two weeks for wearables.
Data Rate	The most data-intensive RedCap use cases rely on 150Mbps in downlink and 50Mbps in uplink.
Latency	RedCap targets latency of less than 100ms for industrial sensors, and less than 500ms for video surveillance.

Source: Juniper Research

In September 2023, Ericsson, Qualcomm, and Vodafone achieved the first RedCap RAN data sessions. It utilised Vodafone's live 5G test network 'CREATE' in Spain, along with Qualcomm's Snapadragon X35 modem. Juniper Research expects the commercial roll-out of RedCap to have begun by the midway point of 2024, with 5G RedCap chipsets already available.

RedCap will be progressed with Release 18, with a focus on positioning performance requirements and enhancements. It is also expected that the release will include details on device-to-device communications over 5G side link interfaces.

Currently, RedCap software is already commercially available for time division duplex and frequency division duplex, in the low and mid-frequency bands, with the first RedCap modules expected to become available in 2024. Juniper Research expects early adoption of RedCap to be for use cases such as CCTV (Closed Circuit TeleVision) cameras, low-cost routers, and industrial sensors and machinery.

Later deployments of RedCap will be more technologically advanced, including augmented reality devices, and high-end wearables such as smart watches. As a result, RedCap is expected to have the greatest impact in the following market verticals:

- Smart Cities – applications such as smart grids, environmental sensors, predictive maintenance, and high-resolution surveillance will be supported by RedCap, with the tier enabling enterprises to cut back on unnecessary power and resource consumption.
- Manufacturing and Mining – as RedCap enables reliable connectivity and mobility to devices, it can be utilised to support wearable devices, and sensors throughout manufacturing and mining operations, enabling greater efficiency, and expanded communication between workers and machines.
- Healthcare – RedCap devices' mix of mobility and throughput enables operators to provide reliable connectivity for healthcare use cases, such as health monitors, and health implants.

In order to monetise these uses of RedCap, operators will create specific subscription models, plans, and bundles, which provide enterprise with access to network slices optimised for RedCap. This will increase revenue, but also introduce further complexity into the cellular IoT market.

Operators and IoT vendors must address this complexity through developing their RedCap capabilities. Specifically, Juniper Research recommends that operators and IoT vendors develop RedCap deployment analytics that provide enterprises with detailed information on where RedCap can improve their IoT operations. This must also extend to recommendations for connectivity providers, and device configurations to further optimise network operations.

1.2.2 Digital Twins in Cellular IoT

Juniper Research defines a digital twin as:

'A real-time replication of a real-world object, providing a virtual image of a physical entity or system.'

Juniper Research expects it to become increasingly complex and difficult for operators, vendors, and enterprises to understand and manage operations efficiently.

Digital twins provide a solution to this complexity and difficulty enabling the monitoring, design, simulation, optimisation, and prediction of the behaviour of physical systems.

Juniper Research recommends that operators and IoT vendors integrate digital twins into their IoT platforms, with a specific focus on enabling operators' understanding cellular IoT focused private network deployments.

Digital twins will also be especially effective in use cases, such as agriculture, where there are a large number of sensors for data collection, which can then be used to replicate a physical environment such as a super farm. Therefore, operators and IoT vendors with large client bases in these markets, or seeking to expand their offering to enterprises in these segments must make the development of digital twins in IoT management platforms a priority.

The integration of digital twins will enable IoT management platforms to provide a wide range of services to enterprises including:

- Troubleshooting – digital twins can be used to replicate previous network failures, which can then be studied to identify the specific configuration errors which lead to failure.
- Anomaly Detection – digital twins can be used to provide a vision of expected network behaviour, which can then be cross-referenced with actual network behaviour to identify anomalies.
- Planning – digital twins produce accurate real-time performance estimates, enabling operators to predict when a network will run out of resources. This information allows users to create accurate and effective priority schedules for network upgrades, and expansion.
- Hypothetical Scenario Analysis – users can utilise digital twins as a sandbox to experiment with potential network configurations, and how networks will cope with change in user behaviour, such as data consumption.

In order to effectively provide digital twins as part of their IoT management platforms, Juniper Research recommends that operators and IoT vendors offer Digital Twins-as-a-Service solutions, reducing the technical expertise required to benefit. In turn, this will expand the addressable user base for digital twin services, reducing the costs of network and IoT management.

Moreover, operators and IoT vendors must ensure that enterprises can easily and directly manage their IoT devices from the digital twin user interface. This will simplify the management process, increasing efficiency.

Alongside integrating digital twins into their IoT management platforms, operators must also seek to implement them into their own cellular networks. This will enable them to troubleshoot potential network issues and anomalies, plan future investments, and conduct hypothetical scenario analyses on their own networks, ensuring that quality of service is retained despite increasing complexity and difficulty due to the growing number of IoT devices, and traffic.

In order to develop their digital twin offerings, Juniper Research recommends that operators and IoT vendors enable enterprises to create automatic responses and processes to specific scenarios to further improve the efficiency of IoT management.

eSIMs (Embedded Subscriber Identity Module) and eUICC (Embedded Universal Integrated Circuit Card)

Juniper Research defines an eSIM as:

'The hardware component of the eSIM that allows for remote SIM provisioning and the ability for a connected device or sensor to switch between network operator profiles.'

Juniper Research defines an eUICC as:

'The software component of the eSIM that allows for remote SIM provisioning and the ability for a connected device or sensor to switch between network operator profiles.'

eSIMs are soldered directly into devices, protecting them from damage or being lost. Further benefits of eSIMs include:

- Security – eSIMs are not physically swappable, and therefore not vulnerable to SIM swap attack, and identity theft.
- Versatility – eSIMs enable users to store multiple cellular profiles in single SIM; enabling users to easily switch between different profiles.
- Size – eSIMs are smaller than traditional SIMs; enabling their usage in smaller IoT devices.

eSIMs are vital to cellular IoT use cases dependent on roaming, as the lack of roaming agreements for 5G and LTE-M networks prevents effective roaming services. By leveraging eSIMs, enterprises can easily ensure that their IoT devices retain constant connection to networks, regardless of where they travel domestically and internationally.

Moreover, enterprises are able to pay local prices for connectivity, rather than roaming charges which are higher. This locality also enables enterprises to bypass restrictions on permanently roaming IoT devices, such as those in Brazil, Canada, China, India, Singapore, and the US, allowing for continued mobility. Consequentially, eSIMs are key to use cases such as international asset and fleet tracking.

In order to simplify enterprises' usage of eSIMs, operators and IoT vendors must integrate remote SIM provisioning platforms into their IoT management platforms. A remote SIM provisioning platform enables enterprises to change a SIM's primary cellular carrier, or profile remotely. By integrating platforms together, operators and IoT vendors will be able to unify the process of managing IoT connectivity, thus improving efficiency and value, and thus revenue.

A key challenge to the impact of eSIMs on the cellular IoT market is the limitations of NB-IoT support for SMS communications. The majority of operators do not support SMS communications with NB-IoT; eliminating the usage of eSIMs. This is as eSIMs are reliant on SMS communications to control OTA (Over the Air) profile transactions.

However, in May 2023, the GSMA announced it would address this issue with new eSIM specifications for IoT devices.

i. GSMA Releases New eSIM Specification for Constrained IoT Devices

The GSMA's specifications introduce a new entity known as eIM (eSIM IoT Manager) which provides a simplified architecture for eSIM profile switching, thus enabling effective eSIM usage in IoT devices. The development of test specification, compliance, and certification is expected to be completed by the mid-way point of 2024.

Under the new specification, subscription management is handled by the eIM, eliminating the need for integration of the SM SR (Subscription Manager Secure Routing) with provisioning servers, reducing costs. Instead, the eIM relays profile

download messages between the provisioning server, terminating the provisioning server's protocols for a protocol that is suitable for low-powered IoT devices, such as CoAp (Constrained Application Protocol) over DTLS (Datagram Transport Layer Security) over UDP (User Datagram Protocol). The new specification permits any protocol stack between the IoT device and eIM, providing it meets security requirements.

This support for new protocols and the creation of an intermediary between the provisioning server and IoT devices will enable eSIM use over NB-IoT networks, and on constrained IoT devices, such as sensors that lack the capabilities necessary for eSIM usage.

Operators and IoT vendors must ensure that their IoT management platforms are capable of supporting the new specifications, by embedding the eIM as part of their IoT management platforms. This will enable the same protocol stack for device and data management to be used for profile download and management, reducing the complexity of eSIM management.

Juniper Research further recommends that operators and IoT vendors focus on the creation of autonomous profile management tools, which will enable profile switching based on parameters predetermined by the enterprise. Automation will be critical to the effective use of eSIMs in IoT as the vast number of devices used will make individual profile management impossible for enterprises.

1.2.3 Federated Learning in Cellular IoT

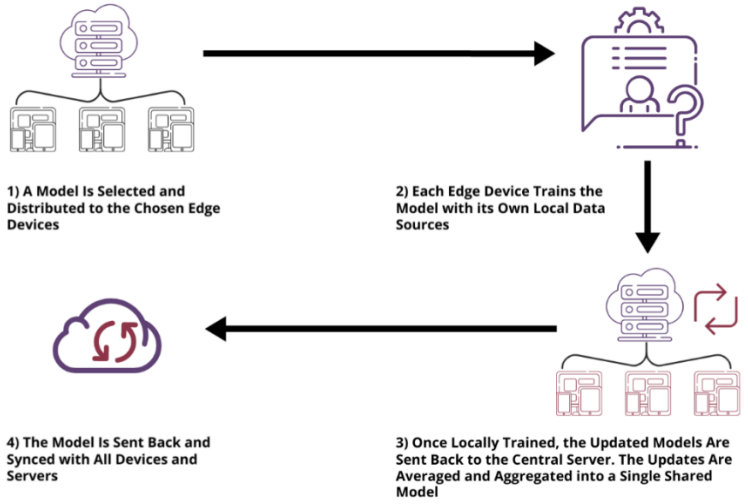
Deep learning is a subset of machine learning based on neural networks, where multiple layers of processing are used to extract higher-level features from data. Federated learning is itself a subset of machine learning. Juniper Research defines federated learning as:

'A decentralised approach to training machine learning models that does not require data to be exchanged between participating parties.'

Federated learning trains models on raw data at the network edge. These localised models are then aggregated to build a shared global model. This provides three key advantages over traditional machine learning approaches, which are:

- Scalability – federated learning allows a deep learning model to be simultaneously trained on multiple datasets. As a result, the development and deployment time of effective models is significantly reduced, allowing for models to be scaled quickly.
- Cost-effectiveness – as federated learning trains machine learning models at the edge of networks, and only transfers trained parts of the aggregate over the network, there is a significant reduction in the volume of data needed to be transmitted when compared to machine learning models centralised to a single server or cloud. As a result, users experience significantly lower costs for data transfers throughout the development of their machine learning models.

Figure 3: The Process of Federated Learning



Source: Juniper Research

- Security and Privacy – federated learning allows collaboration on machine learning models without compromising proprietary data or privacy concerns, with each user training their version of the central model individually. The newly trained model's configuration is then summarised and encrypted, prior to being uploaded to the central server. As a result, users can be confident in the security and privacy of their own data throughout the collaboration process, removing a major barrier to collaboration.

Federated learning is ideal for implementing machine learning in cellular IoT, as traditional machine learning applications are restricted, due to data and privacy, concerns, and regulations.

Moreover, data quality can vary and network availability can become unstable, meaning the traditional machine learning approach's centralised model training can reduce model quality, and increase the probability of errors. In contrast, the decentralised approach of federated learning enables operators to isolate poor quality traffic, and insulate models from potential errors.

Further federated learning will provide operators and vendors with greater flexibility in the development of machine learning models, enabling them to cater to the heterogeneity in cellular IoT devices and use cases. For example, localised models can be used to account for the difference between different geographic models in the larger model, as well as to create optimised processes which cater to different IoT device capabilities, such as processing power.

Despite the advantages of federated learning, its application in cellular IoT faces several significant challenges, including:

- Statistical Heterogeneity – this refers to the distribution of data volume and class distribution variance among clients. This can increase model training latency and accuracy, decreasing the capacity of operators and IoT vendors to efficiently use federated learning.
- Client Heterogeneity – this refers to differences in client resources, such as storage and computation capacity, which are frequent between IoT devices. These differences result in different training rates between devices, threatening delays, and the bottlenecking of transmissions.
- Communication Inefficiency and Cost – due to the cost of communications, and privacy concerns, data generate by each client remains local, meaning that clients must transfer information or model updates frequently during training. This exacerbates inefficiency communications, especially when numerous clients are uploading updates to the service. This inefficiency further increases the costs to operators and IoT vendors for communications between clients and their federated learning server.

These challenges are exacerbated by the ever-increasing number of clients, an issue which will persist as the usage of IoT continues to grow over the next five years. Consequentially, the evolution of client selection methods will be critical in enabling the role of federated learning to grow in cellular IoT.

A client in federated learning refers to data sources, such as IoT devices, smartphones, and connected vehicles. Federated learning models are divided into two main client categories, which are:

- Cross-device – models drawn from millions of clients, such as individual IoT devices, and edge nodes, each of which typically storing local data. Clients are not expected to participate throughout the entire training process.
- Cross Silo – clients are typically a specific enterprise or organisation, with a limited number of participants, and large data volumes. Each client is expected to participate throughout the entire training process.

The client selection process uses a predefined criteria to select participating clients in each model training round. This can be used to create optimised subsets of clients for stages of model training. The implementation of client selection processes has been demonstrated to accelerate convergence, lower communication costs, as well as optimising final models.

Table 4: Federated Learning Client Selection Methods

Client Selection Method	Description
Random Methods	The traditional client selection method, where all clients will have the same probability of being selected for model training.
Greedy Methods	The clients with high-level quality grades and low expenses are selected. A heuristic method is used to characterise the quality rate of each client.
Clustering Methods	Clients that train the model are clustered according to their attribute similarities, such as resources, allocated data, and location.
Multi-armed Bandit Methods	Predominantly used to identify the cause if repeated discovery of a situation.

Source: Juniper Research

In order to maximise the impact of client selection methods, Juniper Research recommends that operators and IoT vendors develop a dynamic approach, with adaptive client selection and gradient compression. Adaptive client selection is a method which is cognisant of the training progress of clients. Gradient compression is a technique that reduces the communication overhead and bandwidth usage of distributed training frameworks. This approach will enable different compressions to be assigned to individual clients, accounting for specific characteristics such as time varying, and computing capabilities.

The grouping of adaptive client selection and gradient compression will enable operators and IoT vendors to address key challenges to the use of federated learning in cellular IoT, such as client heterogeneity, and communication costs.

Operators and IoT vendors must also introduce diversity considerations into client selection, picking representatives out of the available clients. By encouraging diversity, redundant communication and under-representation can be reduced, therefore improving the quality of the aggregate model, and reducing communications expenditure. Given the heterogeneity in IoT clients, this diversity will be critical to effective usage of federated learning.

Juniper Research further recommends that operators and IoT vendors focus on developing the capabilities of their federated learning servers. Whilst the federated learning process occurs in individual clients, the federated learning server is still critical. The server in federated learning is responsible for coordinating the learning process and aggregating models. The distribution of updated models to clients and device security is also handled by the server. Consequentially, it is essential that federated learning servers are able to provide high data transmission rates, and processing speed.

Operators and IoT vendors can invest in and develop a number of solutions for improving data transmission, and processing speed, including:

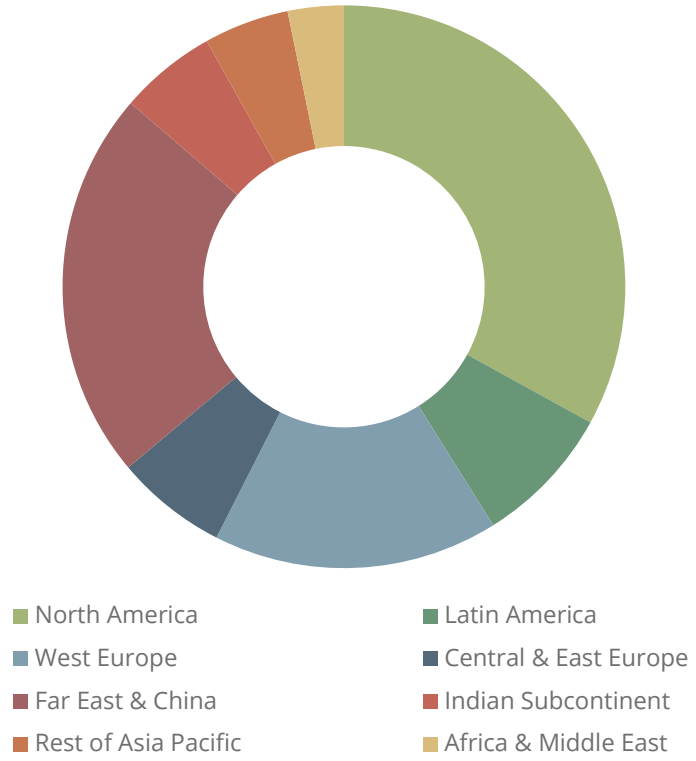
- Asynchronous Updates – asynchronous update mechanisms refer to where the sever aggregate the model based on available data, rather than waiting for all clients to finish training. The implementation of asynchronous updates will enable operators and IoT vendors to optimise server resource distributions, and account for device heterogeneity, by preventing delays, server overloads, and bottlenecks due to low-capability devices.
- Model Quantisation – most numbers in machine learning libraries use a high-precision floating point representation of 32 bits, but this is not always necessary during the training phase. Model quantisation is a model size reduction that converts these this representation to low-precision floating point representation, such as 16 bit or 8 bit. These representations require less memory, enabling more frequent client updates, as well as accelerated data transfers, and data processing.

1.3 Market Forecast Summary: Total Number of Cellular IoT Devices in 2028

Juniper Research found that global number of cellular IoT devices will increase from 3.4 billion in 2024 to 6.5 billion by 2028. However, the study predicted that this 90% growth in connections will require the deployment of new services that enable the efficient automation of IoT device management and security.

- The research identified intelligent infrastructure management solutions that enable IoT users to automate the configuration of devices, security processes and connectivity in real-time, as key to handling the large increase in cellular data. Global cellular IoT data generated will grow to 46 petabytes in 2028, up from 21 petabytes this year, leading to further investment in IoT automation services, such as federated learning.
- The majority of machine learning models are trained via data sources that are stored in a single location, thus making opportunities for fraudulent players a simpler task. In response, it is recommended that operators transition to federated learning models, a subset of machine learning that leverages a decentralised data approach to minimise the chances of data fraud over IoT networks.
- Federated machine learning limits the exposure of sensitive IoT data, thus reducing the threat of data breaches. As the number of cellular IoT connections continues to grow it is imperative that platforms and operators both ensure data is secure in transition and on device. Failure to achieve this will dissuade IoT users in industries with sensitive data from using a cellular IoT-based approach to connectivity.

Figure 5: Total Number of Cellular IoT Devices (6.5 billion), Split by 8 Key Regions, 2028



Source: Juniper Research

Order the Full Research

Discover a must-read assessment of the current and future cellular IoT market plus key research developments in this latest report. Delivering invaluable insights into latest technical specifications, such as the GSMA's IoT eSIM, and Release 17 and 18 from 3GPP, the research reveals vital strategies for working with satellite operators and system integrators. With an evaluation of 60 key countries' readiness for cellular IoT monetisation over the next five years, the report also features a Competitor Leaderboard which provides a comprehensive analysis of 18 leading IoT management platform vendors in the cellular IoT market.

Key Features

- **Key Takeaways & Strategic Recommendations:** In-depth analysis of key development opportunities and key findings within the cellular IoT market, accompanied by strategic recommendations for operators and IoT vendors.
- **Market Outlook:** Deep dive evaluation of the trends and technologies shaping the monetisation of cellular IoT; outlining different strategies required for network technologies, the development and introduction of key cellular IoT standards, and trends in cellular IoT deployments.
- **Benchmark Industry Forecasts:** Business overview for operators and IoT vendors including four-year forecasts for the total number of cellular IoT connections, total operator revenue from cellular IoT, total data generated by cellular IoT, total number of cellular IoT modules shipped, total expenditure on cellular IoT, and total cellular IoT market value. Split by the following network technologies: 2G and 3G, 4G, 5G, and LPWA.
- **Juniper Research Country Readiness Index:** Rigorous assessment of the current and future cellular IoT market status in 60 countries; leveraging Juniper Research's bespoke forecast data for the global cellular IoT market.
- **Juniper Research Competitor Leaderboard**: Key player capability and capacity assessment for 18 IoT management platform vendors, via the Juniper Research Competitor Leaderboard including:
    - AT&T
    - Cisco
    - Huawei
    - Microsoft
    - Telefónica Global Solutions
    - Vodafone Business

What's in this Research?

1. **Market Trends & Strategies:** Top-level report evaluating key technologies, specifications, and software in the cellular IoT market.
2. **Future Market Outlook:** Deep dive evaluation of the trends and technologies shaping the monetisation of cellular IoT; outlining the different strategies required for network technologies including as 4G, 5G, and LPWA, the development and introduction of key cellular IoT standards, and trends in cellular IoT deployments.
3. **Five-year Forecasts:** Extensive forecasts for the total number of cellular IoT connections, total operator revenue from cellular IoT, total data generated by cellular IoT, total number of cellular IoT modules shipped, total expenditure on cellular IoT, and total cellular IoT market value. These are split by the following network technologies: 2G and 3G, 4G, 5G, and LPWA.
4. **Interactive Forecast Excel:** Highly granular dataset comprising of 34,424 datapoints; allied to an interactive scenario tool; giving users the ability to manipulate Juniper Research's data (Interactive XL).

5. ***harvest* Digital Markets Intelligence Centre:** 12 months' access to all the data in our online data platform, including continuous data updates and exportable charts, tables, and graphs (ONLINE).

Publication Details

Publication Date: March 2024

Author: Alex Webb

Contact: For more information contact info@juniperresearch.com

Juniper Research Ltd, 9 Cedarwood, Chineham Park, Basingstoke, Hampshire, RG24 8WD UK

Tel: UK: +44 (0)1256 830002/475656 USA: +1 408 716 5483 (International answering service)

http://www.juniperresearch.com