5G Autonomous Edge
The Next Frontier of Digital Business Innovation

2020

May 25
Leonard Lee
Dean Freeman

Sponsored by
intel®
neXt Curve
# The 5G Autonomous Edge

## The Next Frontier of Digital Business Innovation

### Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXECUTIVE SUMMARY</td>
<td>4</td>
</tr>
<tr>
<td>THE 5G AUTONOMOUS EDGE</td>
<td>5</td>
</tr>
<tr>
<td>WHERE IS THE 5G AUTONOMOUS EDGE?</td>
<td>6</td>
</tr>
<tr>
<td>TRANSFORMATIVE TECHNOLOGY - 5G</td>
<td>6</td>
</tr>
<tr>
<td>TRANSFORMATIVE TECHNOLOGY - ARTIFICIAL INTELLIGENCE</td>
<td>8</td>
</tr>
<tr>
<td>TRANSFORMATIVE TECHNOLOGY - CLOUD COMPUTING</td>
<td>8</td>
</tr>
<tr>
<td>THE NEW PLATFORM FOR BUSINESS INNOVATION AND REINVENTION</td>
<td>9</td>
</tr>
<tr>
<td>UBIQUITOUS INTELLIGENCE (AI + CLOUD)</td>
<td>9</td>
</tr>
<tr>
<td>EDGE CLOUD COMPUTING (5G + CLOUD)</td>
<td>10</td>
</tr>
<tr>
<td>AUTONOMOUS INFRASTRUCTURE (AI + 5G)</td>
<td>11</td>
</tr>
<tr>
<td>AN AUTONOMOUS INFRASTRUCTURE FOR AUTONOMOUS APPLICATIONS</td>
<td>12</td>
</tr>
<tr>
<td>AUTONOMOUS CONTROL LOOP</td>
<td>12</td>
</tr>
<tr>
<td>5G AUTONOMOUS APPLICATIONS</td>
<td>13</td>
</tr>
<tr>
<td>5G AUTONOMOUS EDGE INFRASTRUCTURE</td>
<td>13</td>
</tr>
<tr>
<td>REINVENTING BUSINESSES AND INDUSTRIES</td>
<td>14</td>
</tr>
<tr>
<td>AUTONOMOUS OPERATIONS</td>
<td>15</td>
</tr>
<tr>
<td>AUTONOMOUS SERVICES</td>
<td>15</td>
</tr>
<tr>
<td>AUTONOMOUS INFRASTRUCTURE</td>
<td>15</td>
</tr>
<tr>
<td>INTERACTIVE MEDIA &amp; COMMUNICATIONS</td>
<td>16</td>
</tr>
<tr>
<td>CONCLUSION</td>
<td>16</td>
</tr>
</tbody>
</table>
The Research Team

Leonard Lee, Managing Director, neXt Curve
Mr. Lee is the founder of the research advisory firm neXt Curve. Drawing upon over twenty-five years as a managing partner, principal consultant, and industry analyst with Gartner, IBM, PwC and EY, Leonard has advised and delivered emerging technology and business solutions to leading enterprises across a broad range of industries. His perspective is shaped by extensive experience helping Global 500 companies drive business innovation and value through digital technologies and assisting top technology vendors with their go-to-market strategies for their digital products and services.

Dean Freeman, Research Fellow, neXt Curve
Dean Freeman is a research fellow at neXt Curve focused on semiconductor industry research. He has 17 years of experience as a VP of Research at Gartner where he led the IoT Center of Excellence and the development of the IoT semiconductor forecast. He has worked extensively with leading IoT vendors in developing go-to-market strategies for their embedded device and edge computing offerings. Mr. Freeman also directed multiple aspects of Gartner’s research in semiconductor manufacturing to forecast market share, size and growth of all semiconductor process equipment.
Executive Summary

Imagine…

…a smart factory populated by a fleet of mobile autonomous robots that are able to self-organize and tailor their work on a reconfigurable shop floor bringing about the industrial revolution of mass customization.

…cyber-physical experiences for brick and mortar retail that predictively personalize and orchestrate the offline and online aspects of a customer’s journey revolutionizing the retail experience.

…the next generation of live events and sports media captured in high-definition volumetric video and 3D audio that seamlessly integrates with massive social gaming bringing immersive media into the mainstream.

…smart city infrastructure that supports the real-time coordination of first responder and healthcare services revolutionizing public health and safety.

Thanks to the emergence of the 5G Autonomous Edge, these aspirational use cases are becoming real possibilities. As 5G, artificial intelligence (AI) and cloud computing technologies continue their rapid evolution and converge:

- Intelligence is becoming increasingly ubiquitous
- The power and elasticity of cloud computing extends from large centralized data centers out toward the network edge
- Autonomous digital infrastructures will emerge that massively scale to support new real-time distributed computing architectures

These transformative features of the 5G Autonomous Edge are essential ingredients for the real-time autonomous applications that will open up a new frontier of business innovation and industry reinvention.

Imagine… Enterprises will be able to realize learning and adaptability based on a mutualistic blend of human intellect with intelligent automation. They will be able to invent autonomous products/services and business capabilities that will take digital transformation to its next incarnation, the autonomous enterprise.

Businesses innovating on the 5G Autonomous Edge will have access to the untapped deluge of data at the point it is generated. With thoughtful application of AI processing at the edge and the help of ultra-low latency 5G networks, these pioneering organizations will be able to acquire the highly granular view of their business and achieve the real-time orchestration and control they need to innovate and lead their industries.

The next stage of business innovation and industry reinvention begins at the 5G Autonomous Edge.
The 5G Autonomous Edge

In recent years, edge computing has become the topic du jour. While technologists can wax poetic about how edge computing will change how we build and deploy software applications and digital services forever, what does it mean for businesses? Frankly, the benefits and opportunities are not that clear. After all, edge computing has been around for 20 years in the form of CDNs (Content Distribution Networks) which cache web content and media across the Internet to improve online experiences. Those in the industrial world will remind us that they have been doing edge computing in plants and factories for ages. So, edge computing is nothing new.

Then, what is everyone getting excited about?

The excitement about edge computing stems from the rapid convergence of three transformative technologies: 5G, cloud computing and artificial intelligence (AI). As these technologies continue their rapid evolution and intersect, a new computing model is emerging that will change how we can architect the next generation of distributed software applications and cyber-physical systems. We call this new frontier of edge computing the 5G Autonomous Edge.

In principle, the 5G Autonomous Edge is about new possibilities for where you can place cognition, intelligence and automated control across the edge thanks to the advent of 5G mobile communications and the continued advancements in cloud computing technologies. It is also part of a broader evolution of edge computing from its current content and data-oriented character toward one that is cloud based and increasingly inclusive of AI compute. As much as cloud computing has changed the way that we look at IT services and how we develop and deploy applications, 5G Autonomous Edge will change the way that we will look at cloud computing, telecommunications and digital transformation.

As industries continue to become more digital, interactions with customers, suppliers and partners are becoming faster, on demand and increasingly mobile. Consequently, businesses require faster and more responsive systems able to deliver truly real-time digital services, experiences and business capabilities anytime and anywhere to a wide range of computing devices and the billions of IoT (Internet of Things) endpoints that will blanket our world.

The 5G Autonomous Edge will enable the infrastructures that will deliver the near-zero latency and highly reliable orchestration and control required of safety and mission critical industrial applications, as well as novel interactive media and communications services. In many ways, it will serve as a powerful platform for innovation that harnesses the transformative capabilities of 5G, cloud computing and AI.

To technologists, the 5G Autonomous Edge represents a revolutionary inflection point for edge computing. For businesses, it’s a new frontier for innovation and business reinvention. It means taking the leap from being a digital business to becoming an autonomous enterprise by making the organization a learning and adaptable system. Autonomous enterprises will be able to access the untapped deluge of data at the network edge. With some thoughtfully applied AI processing at the edge, these pioneering organizations will be able to acquire the highly granular view of their business and the real-time orchestration and optimization needed to be competitive and innovative in our digital future.
Where is the 5G Autonomous Edge?

There are many interpretations and varieties of “the edge”. One thing is certain, there is no agreement on where it starts and where it ends. The 5G Autonomous Edge is not defined in terms of the network edge, the IoT edge, the thin edge or any union thereof. It spans the real-time connectivity domain where data communications latencies of less than 5 milliseconds required for mission and safety critical mobile applications are achievable.

The 5G Autonomous Edge spans from the so-called far edge, which is devices, the base stations and access networks they connect to, and everything in between such as on-premise data centers and local area networks. It is in this domain that we will see the unique blend of near-zero latency communications and fast-expanding AI compute capabilities running across a cloud-native infrastructure poised to revolutionize edge computing. It is most likely that the elusive 5G killer app will emerge from the 5G Autonomous Edge.

Transformative Technology - 5G

5G is a fascinating and pivotal evolution of communications technology (CT). It promises to transform industries and the way we live with extreme wireless connectivity characterized by enhanced mobile broadband (eMBB), massive machine type communications (mMTC) and ultra-reliable low-latency communications (URLLC). Much of the fanfare about 5G emanates from the next-generation RAN (Radio Access Network), also known as 5G NR or 5G New Radio, which takes advantage of mm Wave spectrum and transformative radio technologies such as massive Multiple-Input Multiple-Output (MIMO), dynamic spectrum sharing (DSS) and adaptive beamforming. Thanks to these advancements in radio technologies, the 5G RAN is far more capable in terms of performance and capacity and uses spectrum and power much more efficiently than the previous generation of mobile technology.
Without 5G, the 5G Autonomous Edge does not have the critical element of URLLC which enables mobile access networks to deliver carrier-grade reliability with 99.999% of packets delivered within 1 millisecond or better. This kind of performance is not easily realized with prior generations of mobile wireless networks. Connectivity capable of 99.9999% reliability was only feasible on deterministic wired networks found in industrial settings such as an auto manufacturing plant.

With 5G, not only is URLLC possible, it can be delivered with the flexibility and the agility of wireless connectivity out to the far edge. This is a game-changer that opens up possibilities for new industrial applications, such as mobile robotic assembly and autonomous drone delivery, that can benefit from low-latency distributed computing for decentralized analytics, orchestration and control. The continuing evolution of 5G technology will bring about enhancements to URLLC. These improvements will push life-critical applications that require sub-millisecond tactile control and immersive video monitoring such as remote robotic surgery into the realm of the possible.
Transformative Technology - Artificial Intelligence

Artificial intelligence (AI) provides the eyes, ears, brains and hands for the next generation of business applications as well as the digital infrastructure on which they will run. It is a broad area of technology that includes computer vision, natural language processing (NLP) and predictive analytics. While AI may seem to be a new technology topic, much of the math and science of AI have been with us for decades. AI has recently reached a tipping point thanks to three trends that are accelerating the evolution and the adoption of the technology by businesses and consumers.

Advances in semiconductor technologies, such as a new generation of neural processors as well as SoC (System on a Chip) and SIP (System in Package) architectures with coprocessors dedicated to handling AI operations, are making it possible to design and build AI-capable computing devices that are smaller and more energy efficient. As a result, AI computing is proliferating outside of data centers and beyond HPC (High-Performance Computing) platforms and is now showing up on our personal devices and a wide range of IoT endpoints.

New computing architectures that use a fast-growing range of hardware acceleration techniques are enabling AI workloads to run on the pervasive x86 platform, which is the mainstay of data center, network and personal computing. The support for AI computing on general purpose computing devices dramatically expands the footprint for where AI operations such as model training and inference can be deployed given the massive installed base of standards-based servers.

Open sourcing of AI technologies and development is making AI much more accessible and approachable. Widely available and open AI frameworks, libraries and tools such as TensorFlow and Caffe are accelerating developer adoption and making it easier for enterprises to build and deploy AI-enabled applications. Community sourced algorithms and models are also serving as accelerators of AI application development and the business innovations they enable.

From a 5G Autonomous Edge perspective, these AI trends are having a democratizing effect by making the power of AI accessible to a growing citizenry of developers and hardware designers that will fast-track the implementation of AI in new business applications and digital infrastructures. Furthermore, the advancements in AI are changing the nature of software and application design. We are rapidly moving away from codified programs toward intelligent algorithms that will enable us to realize the autonomous applications and systems that will populate the 5G Autonomous Edge.

Transformative Technology - Cloud Computing

For the past decade, cloud computing has profoundly transformed the way applications are architected, developed, deployed and managed. It has also changed how consumers and enterprises consume computing infrastructure and platform services as well as software. While we are familiar with how cloud has changed the IT world, it is also influencing the evolution of communications technology (CT) while beginning its foray into transforming the operational technology (OT) domain of the Industrial Internet.

The network is becoming cloud native transcending the notion of software-defined networks (SDN), virtualized network infrastructure and functions. Telecommunications technology vendors are quickly adopting and advocating cloud-based architectures for their RAN and network core technologies, which promise to make the 5G network more flexible, extensible, and a platform for accelerated service innovation.
Containers and microservice-based architectures are changing the cloud as we know it. The cloud is rapidly moving past its infrastructure virtualization roots toward becoming a platform for distributed software development and application service delivery. Kubernetes allows applications and application functions to run across a cluster of virtualized servers providing a level of elasticity and deployment flexibility that has taken cloud computing to the next stage of its evolution. The advent of serverless computing is poised to transcend the concepts of containers and microservices to the function level. This will open up new possibilities for how cloud-based software can be designed and the speed with which applications and their functions can be deployed across the cloud.

Everything is becoming a cloud host thanks to container-based and serverless computing. Application functions and entire applications can find a virtual runtime environment on computing devices from a cluster of servers in the cloud to a smart camera monitoring a production line on a factory floor. In the telco world, we are seeing the deployment of cloud-based network applications on vCPE (virtual Customer Premise Equipment) platforms that are increasingly using containers and microservice-based software architectures running on standard x86 network equipment. Serverless computing is now bringing the cloud to personal computing and IoT endpoint equipment furthering the reach of the cloud to the edge of the network and beyond.

The New Platform for Business Innovation and Reinvention

While 5G, cloud computing and AI are exciting technologies individually, the compelling 5G Autonomous Edge opportunity is forming where these transformative technologies intersect. It is at these intersections where the new frontier of innovation resides for business automation and autonomy. It is also where the three transformative technologies will converge and influence the formation of the 5G Autonomous Edge defining its character as a game-changing digital platform architecture that will bring about:

- Ubiquitous Intelligence
- Edge Cloud Computing
- Autonomous Infrastructure

Ubiquitous Intelligence (AI + Cloud)

The confluence of AI and cloud computing technologies is making intelligence ubiquitous across a cloud continuum that extends from central data centers all the way to devices that reside and connect to the edge of the network. These AI clouds are hosted across a growing range of devices that are able to support AI workloads and collectively form an AI cloud fabric or continuum.

As AI infrastructure continues to become increasingly virtualized, we are seeing a transition from hardware-centric AI implementations and architectures toward AI platforms and applications being delivered as a service. Artificial Intelligence as a Service (AIaaS) providers are making AI platforms and tools more accessible and more economical as a pay-as-you-go OpEx spend rather than a significant CapEx investment. Consequently, AI is becoming democratized thus fueling the broad adoption of the technology among enterprises for internal purposes as well as for product and service augmentation.

The AI cloud continuum also offers developers and architects a multitude of new options for designing, building and deploying AI-based applications. Model training and inferencing workloads can be
disaggregated and distributed across the AI cloud continuum making new AI application architectures possible that can take advantage of the elasticity and scalability characteristic of cloud computing.

### The AI Cloud Continuum

<table>
<thead>
<tr>
<th>Number Placement</th>
<th>Cloud Deployment</th>
<th>AI Workload</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tens of Billions</td>
<td>On-Device</td>
<td>• On-device real-time AI inference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Real-time data stream analytics</td>
</tr>
<tr>
<td>Hundreds of Millions</td>
<td>On-Prem Edge Cloud</td>
<td>• Local real-time AI inference and data analytics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Real-time inference compute offload/augmentation</td>
</tr>
<tr>
<td>Millions</td>
<td>Far Edge Cloud</td>
<td>• Decentralized mesh ML &amp; DL model training and analytics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Real-time inference compute offload/augmentation</td>
</tr>
<tr>
<td>Tens of Thousands</td>
<td>Near Edge Cloud</td>
<td>• Large-scale distributed ML &amp; DL model training &amp; analytics</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Near-real-time inference compute capacity offload</td>
</tr>
<tr>
<td>Hundreds</td>
<td>Public &amp; Enterprise Private Cloud</td>
<td>• Massive-scale ML &amp; DL model training on Big Data</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• AI Inference for large-scale batch analytics applications</td>
</tr>
</tbody>
</table>

With the rapid adoption of container-based software and the advent of serverless computing, AI applications and functions are becoming increasingly portable and can be deployed across the 5G Autonomous Edge. The deployment of these applications will transverse public and private clouds as well as carrier-hosted and private networks. This means that AI workloads can run almost anywhere including on device. We are already seeing deep learning (DL) and machine learning (ML) inference executed on video cameras, intelligent IoT gateways and controllers, drones, servers, base stations and more thanks to ubiquitous intelligence.

**Edge Cloud Computing (5G + Cloud)**

5G is much more than just a new radio technology and improved performance and capacity. 5G enables the cloud-native transformation of the mobile network. This means that the entire 5G network from RAN to core can be implemented on a cloud infrastructure. 5G is also part of the broader evolution of telecom networks into the telco cloud that is occurring as the central offices of network operators are converted into cloud-hosting data centers.

With the comingling of 5G and cloud computing, we are seeing the emergence of edge cloud computing which is extending the cloud out to the network edge. It is being prompted by the fast-evolving standard for MEC (Multi-Access Edge Computing) jointly established by 3GPP (The 3rd Generation Partnership Project) and ETSI (The European Telecommunications Standards Institute), which specifies the implementation for cloud-based compute and network resources located at base stations or cloudlets (mini data centers at the edge of the network). Base stations and cloudlets will play important roles in the 5G Autonomous Edge as the hosts for edge clouds. A network of edge cloud nodes will provide the
proximity of compute infrastructure and platform services essential for low-latency mobile applications and URLLC.

MEC is also driving the convergence of IT with mobile network services and functions forming a hyper-converged edge cloud. This means that the container-based applications and serverless functions that run across the edge clouds hosted at near (telco central offices) and far edge (base stations) nodes will be tightly fused with 5G network functions and services delivered as network slices. Network slicing is an important 5G feature that allows for discrete communications services to be composed from fine-grained, virtualized network resources and functions.

As part of MEC implementations, applications deployed on edge clouds will be able to tap into the vast trove of information of the mobile network to determine the best node to service an edge device or subscriber. The real-time state of network and compute resources, precise location data, and the policies that apply to a subscriber or device can also be used in administering an edge application or function. Armed with this rich contextual and location data, serverless functions will be deployable across the 5G network close to where they are needed and within milliseconds of when they are needed.

Lastly, the 5G RAN itself is also becoming a part of the edge cloud. As the RAN is virtualized, initiatives such as the OpenRAN and O-RAN projects are promoting the adoption of cloud computing technologies and principles in the design and deployment of mobile radio access networks, effectively making them telco cloudlets. Network functions can now be abstracted and disaggregated from the physical infrastructure opening up new possibilities for how RANs can be designed and deployed. These OpenRAN and O-RAN projects are also encouraging the use of open source software and standardized x86-based general purpose hardware, which will improve the interoperability of technologies and services across mobile networks as well as accelerate the growth of the 5G Autonomous Edge footprint.

**Autonomous Infrastructure (AI + 5G)**

The 5G Autonomous Edge will not only foster a class of intelligent business applications, it will also bring about an infrastructure that will help the 5G Autonomous Edge economically and operationally scale. AI is widely considered vital in addressing the immense complexities and dynamics of the emerging 5G network. It will also need to address the equally daunting challenges of managing the infrastructure resources and orchestrating software applications and services across a network of edge clouds.

Fully intelligent automation will be essential in the operation of the 5G Autonomous Edge. While this might sound like science fiction, it is not. Closed-loop automated systems and operations exist among us today in the form of hyperscale data centers that are operated by the major cloud service providers. While these data centers are highly automated, AI is increasingly being used to optimize and fine tune operations to save energy, adaptively improve service quality, secure environments, and maximize profits.

For the 5G Autonomous Edge, AI-based algorithms and analytics will provide the predictive insight needed to make intent-based orchestration and management of compute and network resources possible. Data will be predictively loaded and AI-compute workloads proactively processed at the edge cloud node best suited to deliver application services within service level commitments and proximity to a subscriber or device. This use of predictive algorithms will significantly improve the overall performance and responsiveness of a real-time autonomous application. They will also enable the economical scaling and securing of the 5G network and edge cloud infrastructures and operations.
An Autonomous Infrastructure for Autonomous Applications

The 5G Autonomous Edge is exciting because of its promise of a new class of autonomous applications. The idea of autonomy is not just about the processing and analysis of data in support of decision-making. It is also a matter of control; the ability to institute action in an intelligent response to events and changes in a business environment. The autonomous application at the 5G Autonomous Edge will help enterprises make the transition from automation to truly autonomous operations which represents the next phase in the evolution of digital business.

Autonomous Control Loop

The basis for any 5G Autonomous Edge application is the autonomous control loop. It is, in essence, a learning system that is able to adapt to changing environmental conditions and events by identifying the optimal response. In other words, businesses can now transcend rules and static policy-based automation and enter the realm of autonomous applications that are able to operate in a zero-touch fashion.

Regardless of cycle time, scope or scale, an autonomous control system will execute a control loop of one form or another comprised of three key operations: cognition, intelligence and control as represented by the closed loop of autonomy.

- **Cognition** – Contextual and situational awareness are prerequisites for an autonomous control system. Billions of edge devices will enable an enterprise to construct a real-time cyber-physical view of the state of their business – what is commonly referred to as a digital twin. Today, IoT devices are generating much of the data deluge at the edge. Cognitive AI functions such as image recognition play a critical role in translating sensory data, in the form of video streams and images, into contextual awareness for an autonomous control system.

- **Intelligence** – The intelligence operation of an autonomous control system is the brain of such a system. Data analytics are required to convert contextual awareness acquired through a system’s cognitive operations into business insights. Decision models that use machine learning and deep learning methods such as anomaly detection and predictive analytics support optimal decision-making based on contextual and situational understanding.

- **Control** - The ultimate purpose of an autonomous control system is to provide intelligent automation and control of processes and functions that deliver business results. This is realized through the application of AI-enabled robotic automation that involves the execution of actions in the digital and physical domains, thus closing the loop of autonomous control.

As an application continues to execute through its autonomous control loop, it learns and gets smarter. It gains experience and is able to adapt its logic and policies to yield optimal outcomes.
5G Autonomous Applications

5G autonomous applications are about real time. They will be based on a closed-loop autonomous system architecture, but they will be unique in that they will capitalize on the novel capabilities of ubiquitous intelligence, edge cloud computing and autonomous infrastructure that come with the 5G Autonomous Edge.

When we say real time, we mean milliseconds in system response time. This entails a new level of performance that makes many of the aspirational 5G use cases, such as self-driving cars and remote robotic surgery, real possibilities, thus opening up a vast new frontier of autonomous applications.

As AI becomes increasingly ubiquitous and economical, the 5G Autonomous Edge will become a critical enabler of massive distributed control systems that are able to support synchronous coordination and control of a large number of physical and digital things and services connected to the mobile wireless network. Applications based on artificial swarm intelligence concepts become viable lending to the emergence of self-organizing and self-administered network of things that are able to autonomously address a business problem or need.

Architecturally, massive distributed control systems are an integrated chain of autonomous control loops. In aggregate and in concert, they deliver real-time coordination and control for highly distributed autonomous applications. Thanks to ubiquitous intelligence, AI functions can be strategically deployed across devices and infrastructure thereby localizing the processing and minimizing the volume of data that is transmitted between devices and services in a system. This approach also helps to minimize the bandwidth impact on communications latency and reduces the volume of requests needed to institute coordination and control across a swarm of connected things and digital services.

AI will also play an important role in predictive control, which will be vital for highly distributed autonomous applications in addressing the latencies inherent in the 5G network. ML models and algorithms can provide the clairvoyance needed to proactively queue up compute and network resources across the 5G Autonomous Edge. This will enable an autonomous application to preemptively execute network and compute operations which will reduce the processing delays that introduce latency to an autonomous control system’s overall performance.

5G Autonomous Edge Infrastructure

Autonomous applications need an infrastructure that can provide the low latency, industrial reliability and the massive scalability that are the promises of the 5G Autonomous Edge. Not only will ubiquitous intelligence and edge cloud computing enable autonomous applications, they will also enable the 5G Autonomous Edge infrastructure.
One of the key features of the 5G Autonomous Edge as an infrastructure is web-scale edge cloud computing. Clusters of edge clouds will bring the general compute and AI processing power of the large centralized cloud data center closer to the edge of the network using mesh computing mechanisms and federated learning models. By pooling compute resources of local cluster of cloudlets, a network of edge clouds will be able to scale to support the big data analytics requirements as well as the massive parallel processing required for AI model training.

Autonomous applications will also be able to scale AI inference workloads running on premise or on a device connected to the mobile network. This will be done by dynamically offloading inference workloads to an edge cloud hosted at a local base station that has available capacity. Applications will also scale up to edge clouds hosted at POPs (Points of Presence) or COs (Central Offices) if more capacity is needed for workloads that do not have stringent requirements for ultra-reliable low-latency connectivity.

Hybrid cloud and multi-cloud approaches to edge cloud computing will allow enterprises to marry private and public edge cloud services and capacity for more scaling options. By capitalizing on the portability of containers that run AI functions and an edge cloud infrastructure based on standardized computing hardware and open source software, AI workloads (and their data) can be automatically shifted or scaled where they are needed and where they are able to deliver against service level commitments for an autonomous application.

In order to support 5G autonomous applications, the 5G network and edge clouds need to acquire and share context, location and policy data with the applications that run on them. This constant stream of information about the state and condition of the 5G Autonomous Edge infrastructure will enable the intelligent and intent-based deployment of AI compute functions and data that will make 5G autonomous applications possible.

**Reinventing Businesses and Industries**

While the technical aspects of 5G Autonomous Edge are transformational, the innovations that it will foster will reinvent businesses and revolutionize industries. The 5G Autonomous Edge not only provides the core capabilities that will change the way we think of computing, it will change how business applications and services can be designed and delivered. It is also where the more aspirational and demanding 5G use cases, such as immersive media, industrial internet, and remote robotic surgery,
will become realities. In a sense, we will be moving away from a decade of centralized cloud computing toward cloud computing that is truly ubiquitous.

As a cutting-edge technology, 5G Autonomous Edge will play an outsized role in progressing enterprises to the next phase of their digital reinvention in becoming autonomous enterprises. Pioneering organizations will transcend process automation and make the leap toward fully autonomous operations as well as products and services. Business leaders can consider four areas to innovate on top of the 5G Autonomous Edge: operations, services, infrastructure, and media and communications.

**Autonomous Operations**

**Smart Factories and Warehouses** - Industrial use cases such as smart factories or smart distribution centers that are populated with autonomous or semi-autonomous robots are ideal scenarios for the 5G Autonomous Edge. Current and imminent 5G technologies can provide the deterministic, near-zero latency communications that is critical for wireless production control and management. In brown field environments where assets and devices may not have onboard intelligence to handle AI functions and workloads, 5G Autonomous Edge architectures will make it possible to bring AI inference and autonomous control as close to the endpoint devices and sensors as possible.

**Omnichannel Retail Experiences** – Real-time personalization of the customer experience is becoming increasingly important in industries like retail which are under extreme competitive pressure to be one step ahead of a customer’s intent. Retailers are also driven to provide compelling brand experiences that are increasingly dependent on social and digital marketing. In essence, digital retail is now about gathering contextual insights of a customer’s offline and online presence as well as sentiment in shaping a tailored cyber-physical customer experience and journey. The 5G Autonomous Edge provides retailers with a platform that can deliver a portable omnichannel experience that follows the customer and intelligently delivers offers and branded content in real time to improve sales.

**Autonomous Services**

**Robot Fleet Management** - The orchestration and control of autonomous robot fleets, which includes cars, trucks and drones, can benefit from the massive distributed control that the 5G Autonomous Edge can provide. Real-time route optimization and capacity management will be important capabilities especially for ride-sharing services as well as logistics and transportation use cases. These applications will be able to use highly granular real-time contextual data sets including traffic and road conditions, weather, and customer supply chain data to fine tune optimizations and improve asset utilization and yield.

**Autonomous Infrastructure**

**Smart Cities & Infrastructure** – Smart cities and smart infrastructure involve a large number of parties that contribute to the management of various resources, infrastructures and assets that comprise a city. These resources and infrastructures include water, electricity, emergency services, traffic control, lighting, and sanitation services among others. The 5G Autonomous Edge can support the real-time orchestration and management of resources across multiple suppliers and agencies, thereby reducing response times and enhancing the quality of service. The massive distributed control that the 5G Autonomous Edge enables can help cities manage hundreds of thousands, if not millions, of assets and devices that deliver municipal services and manage critical civil infrastructure.
Interactive Media & Communications

Massive Interactive Media – The 5G promises of URLLC and enhanced mobile broadband can be instrumental in enabling new interactive media experiences for live events where media production requires real-time AI functions to augment content as it is captured. Features such as the real-time overlay of a football player’s stats or embedded advertisements over a volumetric live-action broadcast feed become possible and economical. Immersive content can be gamified to create new interactive media experiences that can be personalized in real time based on online engagement as well as context and location information captured at a live event. Time-sensitive applications such as micro-betting can become a more trustworthy and granular gaming experience.

Immersive Communications & Collaboration – The availability of gigabit connection speeds and URLLC presents pivotal opportunities to reinvent the way we communicate and share. Real-time collaboration on massive CAD files can help a remote engineering team work on a design from anywhere in the world using immersive communications media such as virtual reality (VR) and augmented reality (AR). These ultra-low latency experiences have the potential to foster new levels of productivity and creativity within organizations. Furthermore, remote collaboration and control applications that are safety critical such as remote robotic surgery can benefit from the ultra-reliable connectivity of 5G while capitalizing on the tactile Internet which the 5G Autonomous Edge will make possible.

A rising tide lifts all boats. We can expect that the 5G Autonomous Edge will raise the bar for the art of the possible for digital businesses and push the envelope of industry reinvention, in particular Industry 4.0. It will also have a broader impact of improving the economics of the many business applications and use cases that do and don’t rely on real-time autonomous control. Much like its predecessor, the cloud, the 5G Autonomous Edge will have a definitive influence on the evolution of digital transformation as well as the ICT industry over the decade to come.

Conclusion

The 5G Autonomous Edge is here. Business and technology leaders should start exploring the possibilities of what 5G, AI, and cloud computing in concert can offer to advance their quest to digitally reinvent themselves into autonomous enterprises and achieve new levels of operational and organizational efficiency. It also presents a green field of opportunity to invent the next generation of digital services and experiences for customers and business organizations.

Though 5G is here, we have a way to go before mobile network operators (MNOs) and private network operators are able to realize the full promise of the next generation network. Today, it is clear that the 5G network cannot deliver URLLC on its own. It will require well-architected, finely tuned implementations of 5G technologies and infrastructure. It will also necessitate the thoughtful application of artificial intelligence that will pave the way for the 5G Autonomous Edge.

Like any technology we can expect the 5G Autonomous Edge to evolve as the foundational technologies of 5G, artificial intelligence and cloud computing continue to progress, and as new technological revolutions emerge at the edge. In the meantime, enterprises can realize autonomous applications on top of a fast-expanding field of AI-capable devices that will bring real-time intelligence where it is needed across the network edge.

The next era of business innovation and industry reinvention is beginning at the 5G Autonomous Edge.
About neXt Curve

neXt Curve is a global research advisory firm focused on cross-domain ICT industry and emerging technology research. Our mission is to inspire the digital future of forward-thinking, innovative organizations in the public and private sectors. We provide business insights, thought leadership and leadership coaching to some of the world’s leading companies in the telecommunications, technology and media industries. Our independent research analysts and consultants partner closely with our clients in developing transformative strategies to reinvent their businesses, find new markets for their products and services, and address their most important organizational priorities.

Contact Information

Leonard Lee
Managing Director, neXt Curve
leonard.lee@next-curve.com
+1 (619) 405-8329

For more information, visit neXt Curve at: www.next-curve.com

The information contained in this publication has been obtained from sources believed to be reliable. neXt Curve disclaims all warranties as to the accuracy, completeness or adequacy of such information and shall have no liability for errors, omissions or inadequacies in such information. This publication consists of the opinions of neXt Curve’s research organization and should not be construed as statements of fact. The opinions expressed herein are subject to change without notice. Although neXt Curve research may include a discussion of related legal issues, neXt Curve does not provide legal advice or services and its research should not be construed or used as such.

This material may not be copied, reproduced, or modified in whole or in part for any purpose except with express written permission or license from an authorized representative of neXt Curve. In addition to such written permission or license to copy, reproduce, or modify this document in whole or part, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced. All Rights Reserved.