Technology, Media, and Telecommunications Predictions 2021
Deloitte's Technology, Media, and Telecommunications (TMT) group brings together one of the world's largest pools of industry experts—respected for helping companies of all shapes and sizes thrive in a digital world. Deloitte's TMT specialists can help companies take advantage of the ever-changing industry through a broad array of services designed to meet companies wherever they are, across the value chain and around the globe. Contact the authors for more information or read more on www.deloitte.com.
# Contents

<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foreword</td>
<td>2</td>
</tr>
<tr>
<td>Gaining an intelligent edge</td>
<td>4</td>
</tr>
<tr>
<td>The cloud migration forecast</td>
<td>14</td>
</tr>
<tr>
<td>5G is not hazardous to your health</td>
<td>24</td>
</tr>
<tr>
<td>The next-generation radio access network</td>
<td>36</td>
</tr>
<tr>
<td>Women’s sports gets down to business</td>
<td>50</td>
</tr>
<tr>
<td>The hyperquantified athlete</td>
<td>66</td>
</tr>
<tr>
<td>TV’s New Year’s resolution</td>
<td>78</td>
</tr>
<tr>
<td>From virtual to reality</td>
<td>96</td>
</tr>
<tr>
<td>Video visits go viral</td>
<td>104</td>
</tr>
</tbody>
</table>
Foreword

TMT Predictions 2021: The COVID-19 catalyst

We can’t talk about 2021 without mentioning COVID-19. All nine of our TMT Predictions topics for this year were shaped to some extent by the pandemic. We hope that during 2021, we will be closer to the end of the pandemic than to the beginning, focusing on thriving in the future rather than responding to the present.

Across multiple parts of the technology, media, and telecom (TMT) ecosystem, observers have been repeating some version of the comment, “There have been five years of change in five months due to the pandemic.” COVID-19 has been a catalyst—an unwelcome one, but still a catalyst—for needed changes across the TMT landscape.

A catalyst is a substance that causes a chemical reaction to occur more rapidly than it would have without. Sometimes, only a tiny amount of catalyst can trigger a significant change. In the same way, the SARS-CoV-2 virus, which is only 100 nanometers across and weighs one trillionth of a milligram, has catalyzed considerable changes in many facets of TMT.

Some of these changes happened extremely quickly. Movements to the cloud, to video visits for medicine, and to the intelligent edge were already underway in 2019, but they have been accelerated by years or even decades because of the pandemic. We expect rapid evolution in these three areas to continue through 2021.

Other parts of the TMT industry were also shaped by the pandemic, although not to the same extent. The future of women’s sports may be imperiled by restrictions on gatherings. Upgrades to 8K TVs may be accelerated by lockdowns and the rise in time spent watching video. Digital reality headsets helped enterprises train workers and school educate students virtually when COVID-19 made meeting in person difficult. Misinformation flourishes against a backdrop of uncertainty, and myths about 5G’s health risks surged. And the hyperquantification of athletes may make sports viewing more compelling in a time when stadiums are less accessible.

Faster change isn’t always positive change, of course. But many of the trends accelerated by the pandemic look likely to make the world a better place. More gender balance in sport is a good
thing—end of story. Medical video visit technology that works in the developed world during a lockdown will likely lead to more and better medical access in the developing world and remote areas. Greater use of the cloud, and of open or virtual radio access network (RAN) solutions, may make software and cellular service more affordable for the economically disadvantaged ... and may also do so more sustainably, helping not just people but the planet too.

In a chemical reaction, when a catalyst is removed, the reaction returns to its slower rate. Will the postpandemic world see change, disruption, and innovation decelerate from current levels? Or will the acceleration induced by COVID-19 persist for the long term—perhaps even permanently?

Welcome to TMT Predictions 2021.
Rising from decades of instrumentation, automation, and connectivity, the intelligent edge is maturing into a revolutionary set of capabilities that are already transforming some of the largest technology and communications companies on the planet. Although market estimates vary considerably, Deloitte predicts that in 2021, the global market for the intelligent edge will expand to US$12 billion, continuing a compound annual growth rate (CAGR) of around 35%. Expansion in 2021 will be driven primarily by telecoms deploying the intelligent edge for 5G networks, and by hyperscale cloud providers optimizing their infrastructure and service offerings. These highly capitalized leaders are establishing the use cases and best practices that may make it easier for companies across multiple industries to attain the capabilities of the intelligent edge. By 2023, 70% of enterprises may likely run some amount of data processing at the edge. As one leading graphics processing unit (GPU) manufacturer has stated, “We’re about to enter a phase where we’re going to create an internet that is thousands of times bigger than the internet that we enjoy today.”

Though challenges and headwinds exist, we believe that the intelligent edge is poised to transform the computing landscape, propelling the world’s largest technology companies toward the next generation of connectivity and operational efficiency. By bringing powerful computing capabilities closer to where data originates and needs to be consumed, the intelligent edge unlocks the potential for faster, less expensive, and more secure operations in everything from autonomous vehicles to virtual reality to the Internet of Things (IoT)—helping to accelerate the Fourth Industrial Revolution.

What is the intelligent edge?

The intelligent edge is the combination of advanced connectivity, compact processing power, and artificial intelligence (AI) located near devices that use and generate data. It represents an evolution and convergence of trends in industrial monitoring, automated manufacturing, utility management, and telecommunications, amplified by cloud computing, data analytics, and AI. The intelligent edge puts these latter capabilities physically near where data needs rapid analysis and response, enabling that data to be acted on directly or filtered to push only the most important bits to the core. In particular, the intelligent edge’s ability to bring cloud capabilities to remote operations could greatly amplify their performance.

The monumental rise of AI and the evolution of computation to support it are critical enablers, driving tectonic shifts in the semiconductor industry. Graphics processing units (GPUs) have been moving into data centers to support AI workloads and dedicated AI chips are reaching out to the edges, including devices, to operate on incoming data immediately. Advanced connectivity ties them all together, while virtualization enables services to seamlessly run...
across a web of diverse and dynamic components, from the cloud to the edge. Indeed, effective planning and implementation of an intelligent edge strategy may require coordination and orchestration of multiple ecosystem providers.

The monumental rise of AI and the evolution of computation to support it are critical enablers, driving tectonic shifts in the semiconductor industry.

This rise of the intelligent edge will likely drive the evolution of service architectures to become more location-driven, decentralized, and distributed. The intelligent edge does not replace the cloud or data centers but, rather, is an element within a holistic cloud-to-edge architecture. Some components of a service will run in a centralized cloud, others at the data center, and more yet at the edges on sensor arrays, autonomous vehicles, and potentially billions of machine endpoints. The ways that computation operates on different parts of the data journey, where it operates, and the differing requirements those operations place on connectivity and speed may reshape how services are architected by distributing components based on their needs.

There are challenges to overcome, however. Standards and best practices have yet to cohere, and issues with interoperability and security will likely become more visible. The intelligent edge today combines solutions from telecoms, hyperscalers, and technology providers, and effective implementation requires coordination and integration across multiple sectors. Who owns which pieces? Who makes the most of their capabilities? Who will deliver the best end-to-end solutions for the rest of the market? The answers to these questions could shape the landscape for years.

Why is the intelligent edge important?

For businesses with data-driven use cases, the intelligent edge can offer the following key capabilities:

• More efficient use of bandwidth and greater network visibility, which can lower costs

• Resilience against poor, unreliable, and lost connectivity due to lower dependency on wide area networks (WANs)

• More control over data triage, normalization, residency, and privacy through the ability to keep more data local rather than needing to transmit it across the network to the core

• Support for low-latency use cases and fast response times

• Greater automation and autonomy

With these capabilities, the intelligent edge can add greater visibility across operations, support faster data analysis and real-time response, and enable better automation and more dynamic systems. If certain microservices require very low latency and high security, such as facial recognition for facility access, they can execute at the edge rather than in the cloud. This can enable much tighter decision loops, reducing the costs and security risks of network transit: The edge can send the most important bits to the core and the core can manage the edge.

The intelligent edge can support large-scale transformational solutions that could radically evolve manufacturing, logistics, robotics, mobility, and consumer electronics. For instance, an
intelligent edge system can shift supply chains from somewhat fragile, linear systems to more programmable, responsive, and adaptive digital networks, which are able to reshape themselves to address changing demands and disruptions. As another example, utilities and similar organizations could use the intelligent edge to connect AI-enabled drones to address mounting risks posed by aging infrastructure, which could make identifying and addressing those risks much faster. One Norwegian oil rig has already deployed a remote autonomous robotic dog that can patrol the rig and visually inspect for issues such as gas leaks. These devices can be deployed 24/7 to map and monitor assets, flag problems, and alert the rig’s networks and crews of potential risks.

The advantages of using the intelligent edge for such operations can be significant. Consider how an automated drone could inspect a pipeline for defects. With cloud, the drone might fill its local storage with video of a pipeline inspection, and then return to its base station. It would then upload the video to a remote data center, potentially sending gigabytes of information over the network and requiring significant time to do so. The cloud would apply machine learning (ML) to evaluate the data for defects and, upon spotting them, return that information back to the pipeline site to provoke a response: treat the defect and possibly reroute flows.

With edge intelligence, the same AI/ML inference algorithms that the cloud uses to evaluate the video can be run at the edge directly on the drone. Instead of scanning and analyzing the entire video, the drone could operate on a small, near real-time buffer of the video feed to classify defects. When it identifies a defect, the drone can immediately notify nearby crews to treat the problem. Only frames with defects are archived in the cloud to feed models and training sets, which can then update other drones in the field for better spotting.

This minimizes the data load for analysis and transit, greatly reduces the time between inspection and action, and uses the network only for critical information that will add to the toolchain and drive greater insights and learning.

Who needs the intelligent edge?

The intelligent edge can benefit any business that manages infrastructure, networks, clouds, data centers, and connected endpoints such as sensors, actuators, and devices. It can support consumer use cases that require very low latency, such as cloud gaming and augmented and virtual reality. It can enable enterprise uses that require aggregating, securing, and analyzing a great deal of data across operations and customers. And it can improve industrial processes for managing quality, materials, and energy use, such as monitoring factory floors, assembly lines, and logistics.

Not all businesses will be able to implement intelligent edge solutions broadly right away. Many may need to invest in the right infrastructure and partnerships first before seeing a return on investment from narrow use cases. But laying these foundations can give organizations much greater opportunities in the future.
What's driving demand?

In the year ahead, we anticipate that early growth will be led not only by large telecoms but by hyperscale service providers, content delivery network (CDN) providers, and technology companies as they consume and sell intelligent edge solutions. Technology companies, while marketing intelligent edge components, appliances, and software layers to early adopters, may also seek to reinforce their own manufacturing and supply chains with intelligent edge capabilities. Similarly, telecoms, hyperscalers, and CDNs are not only making more capabilities available to their customers, but also expanding their own intelligent edge infrastructures to advance strategic initiatives. In the medium term, the use of intelligent edge beyond these early adopters will likely grow among manufacturing, logistics, and supply chain.

Most spending on edge computing and intelligence today comes from US telecoms and communication service providers (CSPs). With more devices moving on and off networks, and with more diverse bandwidth needs emerging, these network providers are facing mounting management challenges. They are using intelligent edge technologies to transform and reinforce their own infrastructure, such as by expanding central offices to become next-generation data centers and edge hubs, enabling high-density and dynamic connectivity for 5G and multi-access edge computing (MEC), and virtualizing more of their networks with solutions such as open RAN. These steps support their core business of delivering greater quality of service to subscribers and selling networks to enterprise customers.

Hyperscale cloud and service providers are also moving quickly to add intelligent edge capabilities to their data-driven businesses. Demand is growing from use-case—driven solutions such as autonomous vehicles and mobile robotics that require low-latency, high-redundancy capabilities, as well as from manufacturing and supply chains seeking greater transparency and resilience in a post–COVID-19 world. Further demand may come from the intelligent edge’s ability to address emerging regulations for data sovereignty and compliance. With an intelligent edge, data can be secured and held locally, keeping it within the region it was collected instead of sending it to foreign clouds. This could potentially draw more investment from social media platforms looking to the intelligent edge to help them comply with regulatory regimes, such as the General Data Protection Regulation (GDPR), that may require personal information to be processed locally and anonymously.

Finally, rising demand from industries like manufacturing and mobility (such as automakers and ride-hailing services) could lead to the development of more packaged and managed offerings. This could make it easier for more businesses to attain intelligent edge capabilities. As the COVID-19 crisis accelerates migration to the cloud, businesses undergoing their first wave of cloud transformation could design cloud-to-edge solutions that best meets the needs of their use cases.

Who plays in the intelligent edge ecosystem?

No single provider seems able to build an effective intelligent edge solution by itself. Telecoms, hyperscalers, CDNs, and tech companies all play a role in enabling the intelligent edge, with each providing a part of the solution. Coordinating these various components is not easy: With each company pursuing its own strategic goals, cooperation is often not far from competition. Understanding the role of these players, what they offer, and how they fit in the competitive landscape can better arm organizations looking to attain intelligent edge capabilities.
TELECOMS AND THE EXPANDING INTERNET

For many intelligent edge deployments, telecoms are key partners, with the largest telecoms increasingly selling their own edge computing and IoT solutions as well as on-premise private enterprise networks. As providers within the edge ecosystem, telecoms can offer their enterprise customers and partners end-to-end connectivity for wired and wireless networks, from fiber, cable, 4G/LTE, and Wi-Fi to advanced solutions with 5G and Wi-Fi 6. They also lease portions of their network that enable direct access to backhaul and offer real estate to host edge appliances and data centers in central offices. Cell tower companies can also lease space for edge appliances at their towers, which typically connect directly to fiber.

One key challenge for telecoms is that they no longer have a near-monopoly on connectivity and communication, despite the billions of dollars they spend each year to modernize their networks. Many CSPs deliver network services on top of IP networks, and more hyperscalers and CDNs are building their own network infrastructure. However, telecoms’ ability to provide the combination of backhaul, 5G, Wi-Fi 6, and the intelligent edge could unleash stronger capabilities for managed connectivity, giving them a competitive advantage in delivering next-generation networks with greater quality of service.

HYPERSONS ENCIRCLING THE GLOBE

Some of the world’s most valuable companies—the so-called hyperscalers—offer end-to-end technology solutions, from cloud services to enterprise productivity and consumer lifestyle experiences. These businesses are highly data-driven, both for optimizing and extending their internal operations, and to deliver better services and drive continued engagement by understanding their customers and users. They are implementing the intelligent edge to advance both efforts by extending their hyperscale capabilities closer to the things they measure, and the enterprises they serve.

Hyperscalers’ biggest challenge will likely be to maintain control over data while partnering with connectivity providers. Hyperscale networks could indirectly encroach on CSPs, especially if they were to offer connectivity as a part of their platforms. Nevertheless, building high-quality networks is difficult and costly, especially at scale, which makes partnerships with connectivity providers a more feasible route.

As hyperscalers pursue intelligent edge deployments, they will likely establish new proofs of concept and use cases that will further drive adoption across the market. For example, some hyperscalers are positioning voice AI as a key interface modality across all their consumer services. Being able to run natural language processing on the device—a smartphone, smart speaker, or car—without having to go back to the cloud can reduce latency, guard against connectivity failures and security breaches, and deliver greater quality of service while retaining data within regulatory jurisdictions.

For businesses deploying intelligent edge capabilities, hyperscalers offer public clouds and the IT services and management solutions that run on top of them, as well as AI capabilities. Some hyperscalers already control their own CDNs as well. Their development and marketing of these intelligent edge solutions aim to support their expanding business offerings across cloud, payments, health care, transportation, and media and entertainment.

CDNS ASPIRING TO BECOME THE OTHER HYPERSONS

Arguably, CDNs constitute a proto-edge in that they were developed to shift content closer to the consumer to accelerate the early Web. Some CDN providers are now actively pursuing and marketing intelligent edge solutions, putting them in both cooperation and competition with telecoms and hyperscalers.
CDNs have strong relationships with many businesses, and they boast some of the industry’s highest-quality, redundancy, security, and delivery capabilities. The intelligent edge may reinforce their delivery networks and enable greater quality of service at a time when more companies need strong delivery services. However, because some hyperscalers and telecoms already control their own CDNs, the sectors currently driving the intelligent edge may not need external CDN providers. Additionally, CDN companies may not have strong relationships with the industries likely to adopt edge next, such as manufacturing and automotive.

TECH PROVIDERS OFFERING HARDWARE AND MANAGED SERVICES
Underneath the cloud-to-edge value chain reside innumerable hardware components and software layers that support it. Tech providers have a vast growth opportunity in selling their products and services into data centers, networks, on-premise facilities, and endpoints from factory floors and smart buildings to computers and vehicles.

Data center equipment and Wi-Fi connectivity will likely be in especially high demand. One forecast expects the edge data center market to approach US$16 billion by 2025. Tech providers can help telecoms transform their central offices into modern data centers, help manufacturing facilities build stronger on-premise capabilities, and provide intelligent edge appliances and micro data centers to enterprise customers. They can also offer the middleware and management layers that tie these hybrid clouds together into seamless fabrics. In this sense, the edge is another location for the hybrid cloud.

Also, not all intelligent edge solutions require 5G, and more solutions may develop hybrid networks that include Wi-Fi, LTE, and 5G, depending on the use case. This means that some tech businesses may see more opportunities as Wi-Fi providers. The latest generation of Wi-Fi, Wi-Fi 6, offers bandwidth slicing, better power management, and support for an increased number of devices, all of which can enable much more robust and dynamic local networks. Private local networks could also help some businesses reduce their dependence on cloud providers and telecoms.

All this being said, tech providers’ position in the intelligent edge marketplace may be challenged by some of their largest customers—the hyperscalers and telecoms—building more of their own components, as well as the ongoing trend to dematerialize hardware into software wherever possible. However, as the market matures, tech providers may also serve more second-tier adopters less able to build their own solutions. This shift may be the next milestone for the realization of Industry 4.0, in which next-generation cloud-to-edge architectures become more standardized and commoditized.

DON’T FORGET SEMICONDUCTORS
Running AI on lightweight edge devices can require more specialized compute solutions such as custom field programmable gate arrays (FPGAs) and application-specific integrated circuits (ASICs), as well as GPUs designed specifically to accelerate AI algorithms on devices, in edge appliances, and in micro data centers. Demand for edge AI chips has grown substantially, and the market will likely expand considerably in the next few years. To meet this demand, more semiconductor companies are accelerating production of edge AI chips. Some are pursuing monumental M&A deals to shore up their position in the next wave of machine intelligence and computation. Hyperscalers are also designing more of their own specialized chips to support their largest operations. And foundries and chip design firms are responding to meet the demand driven by data centers, AI, and the ever-expanding digitization of industrial systems.
THE BOTTOM LINE

Each party in the intelligent edge ecosystem—telecoms, hyperscalers, CDNs, tech providers, and semiconductor makers—has a stake in the success of the intelligent edge and the next generation of cloud-to-edge architectures and services. The landscape is young and dynamic, but it also has decades of momentum behind it due to ongoing computing and network trends toward instrumentation, monitoring, and automation of facilities. With digitization, connectivity, and data analysis now maturing rapidly, the intelligent edge is beginning to transform some of the largest physical systems on the planet.

Like previous large-scale infrastructure upgrades, innumerable unexpected innovations will likely emerge as intelligent edge adoption expands. Executives should wade in and develop pilots with a clear path to results and strategic value. They should move beyond buzzwords and terminology to focus on use cases, metrics, and outcomes. Most importantly, they should understand that “edge” and “intelligence” are just components of a more holistic solution for faster data handling, greater autonomy and transparency across operations, and a more flexible and adaptive enterprise.

What are the potential headwinds?

As it matures, the intelligent edge market faces some challenges. The COVID-19 crisis has disrupted demand, challenged supply chains, dragged down earnings, raised costs for many businesses, and injected uncertainty into outlooks. Additionally, the ongoing trade war continues to drive supply uncertainty, adding complexity, cost, and time to navigating supplier networks. In this environment, demand is difficult to auger and supply may be unreliable. Such conditions may engender a more conservative approach to capex on emerging technologies. Businesses may feel they can reinforce their existing cloud capabilities with less risk than implementing an intelligent edge, especially during a crisis that is demanding more remote and connected services. Building the fundamentals of cloud migration may occupy many businesses in the near term, leaving edge development to the largest companies.

Additionally, while edge opportunities are becoming clearer, many companies may still regard them as a forward-looking strategic investment rather than an obvious way to drive their current business. Implementation can be challenging and costly, often requiring orchestration between multiple providers. Standards are still forming, best practices are not yet clear, and security across an abundance of diverse edge endpoints cannot yet be guaranteed. In times of greater economic restraint, investments in the future may be easiest for the largest and most durable businesses—and even they may need to orchestrate services across providers.

The year 2021 may thus see the intelligent edge colonized primarily by already-dominant tech sector and telecom leaders, further reinforcing their competitive advantage in the coming wave of transformation. The efforts of these early adopters over the next year can help the intelligent edge prove its value. In the next two years, the market will likely sort out best practices, establish standards and interoperability, and potentially lift early leaders while making it easier for smaller businesses to adopt intelligent edge capabilities.
Endnotes


2. This estimate is based on 3rd party market research and Deloitte’s own evaluations. For this evaluation, the market includes hardware, edge networks, applications, and services that directly serve demand for edge computing and intelligence. Given the broad nature of the market, these estimates may be more directional than precise.


7. Patrick Moorehead, “It’s official – NVIDIA acquires Arm for $40 billion to create what could be a computing juggernaut.”

8. Mark Casey, Jeff Loucks, Duncan Stewart, and Craig Wigginton, Bringing AI to the device: Edge AI chips come into their own, Deloitte Insights, December 9, 2019.


10. Ibid.

11. Chris Arkenberg, Naima Hoque Essing, Sanket S. Nesargi, and Jeff Loucks, Unbundling the cloud with the intelligent edge: How edge computing, AI, and advanced connectivity are enabling enterprises to become more responsive to a fast-moving world, Deloitte Insights, September 8, 2020.


17. Pablo Valerio, “To comply with GDPR, most data should remain at the edge,” IoT Times, October 31, 2018.


28. Casey, Loucks, Stewart, and Wigginton, Bringing AI to the device: Edge AI chips come into their own.


GROWTH IN CLOUD computing has been a megatrend over the last decade, with the market experiencing triple-digit annual growth as recently as 2015. Even though growth among the largest hyperscale public cloud providers had declined to “only” 31% annually by the end of 2019, and this rate had been projected to (slowly) decline further in 2020 and 2021 as the industry matures, growth in cloud continued to outpace that in many other sectors.

It would have not been surprising to see cloud spending go down a few points in 2020, given the spending reduction in multiple areas driven by the COVID-19 pandemic and the associated global recession. Instead, the cloud market has been remarkably resilient. By some metrics, growth was more or less flat in 2020; by some other ways of measuring growth, it increased faster than in 2019, even in the face of the steepest economic contraction in modern history. The likely reason: COVID-19, lockdowns, and work from anywhere (WFA) have increased demand, and we predict that revenue growth will remain at or above 2019 levels (that is, greater than 30%) for 2021 through 2025 as companies move to cloud to save money, become more agile, and drive innovation.

“We’ve seen two years of digital transformation in two months.”
— Microsoft CEO Satya Nadella

How do we know that cloud is growing?

Of course, cloud is not the only solution in play. When viewed at the total company level, very few systems will be only on-premise, only public cloud, or only private cloud. Most deployments will likely use a combination of a public cloud and a private environment that remain distinct entities but are bound together, an approach known as hybrid cloud. Hybrid cloud can take many forms, such as a combination of private with public cloud or public cloud with on-premise resources, but all offer the benefits of multiple deployment models.

Clearly, hybrid cloud is the new normal. According to a March 2020 report, more than 90% of global enterprises will rely on hybrid cloud by 2022. Another survey from the same month found that 97% of IT managers planned to distribute workloads across two or more clouds in order to maximize resilience, meet regulatory and compliance requirements, and leverage best-of-breed services from different providers.

But even though many organizations will retain at least some on-premise resources, and even in current economic conditions, cloud providers have much to look forward to as migration to the cloud accelerates. Various tangible and measurable indicators highlight the magnitude of this acceleration, including workload, revenues among public cloud providers, revenues among the semiconductor companies whose chips power the cloud, and growth in cloud traffic across global telecom networks.
As a note, many cloud forecasts are black boxes, based on proprietary information that cannot be replicated. However, metrics such as hyperscale cloud revenues, data center chip revenues, and cloud traffic are all publicly available, and anyone can reproduce our work and see the same trends.

ON-PREMISE VERSUS CLOUD WORKLOAD
An April 2020 survey of 50 CIOs found that respondents expected to see the proportion of total workload done on-premise drop from 59% in 2019 to 35% in 2021, a reduction of 41%. Moreover, they expected public cloud’s proportion of total workload to grow from 23% to 38% in the same timeframe, with private and hybrid cloud reaching 20% and 7% of workload, respectively. Sixty-eight percent of the CIOs ranked “migrating to the public cloud and/or expanding private cloud” as the top IT spending driver in 2020, up 20 points from a similar survey only six months earlier.

HYPERSCALE CLOUD REVENUES
The five largest hyperscale public cloud providers that disclose segmented revenues saw their combined revenues grow by 31% in 2019 to US$94 billion. Despite widespread tech spending weakness in calendar Q1 2020, revenues grew by 31% over the same period in the previous year. In calendar Q2, growth showed a 25% year-over-year increase, which rose slightly to 27% in calendar Q3, resulting in a 28% growth rate over the first nine months of the year. Revenues for 2020 are likely to be over US$125 billion, increasing to more than US$160 billion in 2021. And although percentage growth was down three points in the first nine months of 2020 compared to all of 2019, it is worth noting that in absolute dollar terms, the total hyperscale market (not just the five largest) grew more in 2020 than in 2019. The market in the first half of 2019 was about US$40 billion larger than in the first half of 2018, while the market in the first half of 2020 was about US$50 billion larger than in the first half of 2019.

**FIGURE 1**
Cloud spend has done better than overall IT spend
YoY growth in spend, percent

<table>
<thead>
<tr>
<th></th>
<th>2019</th>
<th>Q1 2020</th>
<th>Q2 2020</th>
<th>Q3 2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global noncloud IT infrastructure</td>
<td>31%</td>
<td>-16%</td>
<td>-9%</td>
<td>-10%</td>
</tr>
<tr>
<td>Hyperscale cloud revenues</td>
<td>3%</td>
<td>31%</td>
<td>25%</td>
<td>27%</td>
</tr>
<tr>
<td>Data center chip revenues</td>
<td>1%</td>
<td>42%</td>
<td>51%</td>
<td>18%</td>
</tr>
</tbody>
</table>

Source: Deloitte analysis of quarterly and annual financial statements.
DATA CENTER SEMICONDUCTOR REVENUES
As of mid-2020, there were 541 hyperscale data centers globally, with 26 added in the first half of 2020 and another 176 planned over the next few years. All of these data centers need chips. Though chip spending and cloud revenues are not perfectly correlated, they are connected in the long run, with growth in chip revenues usually being a leading indicator: The chips need to be bought and installed in the data centers before the cloud revenues start flowing.9

The three largest semiconductor companies that disclose segmented data center sales saw their combined revenues grow by only 3% in 2019 to just under US$30 billion. In calendar Q1 2020, they saw growth explode by 42% compared to the previous year. In calendar Q2, their revenues grew further to 51% year over year. Although growth declined to 18% in calendar Q3, this still works out to a 36% increase in the first nine months of 2020. Total 2020 revenues are likely to exceed US$35 billion, and could top US$40 billion in 2021.

GLOBAL CLOUD NETWORK TRAFFIC
According to a deep-packet inspection report on network traffic during the period from February 1 to April 19, 2020, global cloud traffic as a percentage of total internet traffic rose from 1.26% to 1.83%, up by 45%. Over the same timeframe, overall internet traffic grew by 38%, meaning that cloud traffic, measured by the absolute number of bits per day, rose by almost exactly 100%.

MORE EVIDENCE OF CLOUD GROWTH
Only about 10 large public hyperscale cloud providers and chip companies break out their cloud revenues on a quarterly basis in detail. However, many other companies sell chips, storage, and connectivity solutions into the cloud space. Although these companies are not necessarily providing detailed quarterly numbers, their commentary has been in line with those that do release exact numbers. As just one example, Micron, a supplier of storage for multiple markets, said in Q2 2020 that “Our cloud DRAM sales grew significantly quarter over quarter, with strong demand due to the work-from-home and e-learning economy and significant increases in e-commerce activity around the world.”

Additionally, investors have been pumping funds into the cloud sector, with total assets in the three largest cloud exchange traded funds (ETFs) reaching US$6 billion as of mid-August 2020, US$2 billion more than at the start of the year. Not only were assets up, but so was performance, with the three cloud ETFs yielding an average year-to-date return of 47% as of October 30, compared to only 22% for the NASDAQ and 1% for the S&P 500.

Additionally, investors have been pumping funds into the cloud sector, with total assets in the three largest cloud exchange traded funds (ETFs) reaching US$6 billion as of mid-August 2020, US$2 billion more than at the start of the year.

All the COVID-19–driven interest in cloud is driving mergers and acquisitions as well. In just the first four months of 2020, the value of completed data center acquisitions, at US$7.5 billion in only 28 deals, was greater than in all of 2019.

THE CAVEAT
Although cloud is growing rapidly overall, it serves multiple industries, many of which have cut spending sharply. This means that while overall growth is strong, it is not uniform. As stated by the
industry publication *SiliconANGLE*: “Because the big cloud players ... are so large, they are exposed to industries that have been hard-hit by the pandemic. As a result, we see pockets of spending deceleration even at these companies.”

**What happens next?**

Although the growth in cloud in the first nine months of 2020 was very high, many forecasts expected it to slow to some extent in subsequent quarters. However, two factors suggest that this decline in growth could be less than expected. First, although lockdowns are unlikely to be as uniform as they were earlier in 2020, flareups in the pandemic and more-localized lockdowns are still driving WFA and cloud growth. Second, in the longer term, the WFA “forced experiment” is being seen as a success by many workers and employers. As an example, Siemens is allowing employees to WFA, where feasible and reasonable, two to three days per week going forward; this policy applies to more than 140,000 employees at about 125 locations in 43 countries. Continuing or growing WFA arrangements such as this could strengthen ongoing demand for cloud.

As far as the industry landscape goes, many technology markets see significant concentration with one or two large companies accounting for almost all of the market, and cloud is no exception. The two largest hyperscale providers accounted for 78% of all revenues among the top five hyperscale providers in 2019, and the largest chip company accounted for 82% of total data center semiconductor revenue in the same year. At least so far, the effect of COVID-19 has not led to increasing concentration; indeed, the leading hyperscale providers’ market share declined slightly (two to seven percentage points) during the pandemic-related surge in cloud growth. Longer term, as growth returns to historical rates, it seems likely, based on the history of technology, that market concentration will increase again. Economies of scale usually matter, and while it isn’t necessarily “winner take all,” it may be “winner take most.”

The market for hyperscale cloud services might be shifting from a global market to a decoupled market split between China, served mainly by China-based providers, and the rest of the world (ROW), served primarily by US-owned hyperscale companies. Based on limited data, it appears that the Chinese cloud market grew faster than the ROW hyperscale market through September 2020, and we would predict that decoupling would continue, if not increase.

Finally, it is worth noting that in addition to the pandemic, the move to cloud has a long-running tailwind in terms of demand. Flexible consumption models, also known as “everything (or anything) as a service” or XaaS, have become an increasingly important strategic shift for enterprises across all industries. This market draws on more than just cloud, but cloud is a critical enabler. As of 2018, the XaaS market was nearly US$94 billion, and a pre-COVID-19 forecast predicted a five-year annual growth rate of 24%, resulting in a market of over US$340 billion by 2024. COVID-19 likely will accelerate the growth in flexible consumption models, but even postpandemic, those making this cloud-driven shift can see greater financial predictability, lower unit costs from aggregation, and enhanced customer relationships. Companies that have shifted their offerings to an XaaS model have already experienced considerable success with both consumers and investors, challenging conventional valuations and placing pressure on industry players that are retaining traditional business models such as perpetual licensing and long-term contracts.
THE BOTTOM LINE

Cloud providers can take several steps to support their continued growth.

First, to paraphrase Spider-Man, “With great growth comes great capital expenditures.” Higher-than-expected growth is good news, but to keep up with it, cloud providers will likely need to spend more on capex. In 2019, total hyperscale spending on capex (which includes both IT infrastructure and physical infrastructure spend) was over US$120 billion. Given the continued growth in revenues, it seems likely that hyperscale capex will continue to grow at double digits, reaching US$150 billion by 2022. Additionally, investment isn’t needed just for capex. For cloud providers, artificial intelligence (AI) apps and dev tools are critical to attracting and maintaining enterprise customers and require investment or acquisition.

Also, as cloud moves from roughly one-third of enterprise workflow to roughly two-thirds, and that more quickly than expected, concerns around privacy and security should urgently be addressed. As just one example, the health care industry, which has been among the fastest to shift to cloud during the pandemic, will likely increase its exposure to new vulnerabilities, especially if the migration is not done properly. As articulated by Healthcare IT News: “While cloud computing better optimizes the use of resources in health care, it also creates significant risks. This is especially true when cloud adoption happens faster than proper due diligence can be applied by information security personnel. This trend will persist well after the pandemic.”

One emerging development for hyperscale cloud providers is the intelligent edge. The intelligent edge places computing power, specifically AI computing power, not in centralized data centers but closer to the end user, typically less than 50 kilometers. The intelligent edge is not a replacement for enterprise and hyperscale cloud data centers, but a way to distribute tasks across the network to increase timeliness, connectivity, and security. In the intelligent edge model, much of the data that used to always go to the data center doesn’t go there anymore, and hyperscale providers should make sure this data finds its way back to centralized clouds for analysis and AI training—and ensure that they’re not cut out of access to this data. Another goal for cloud providers is to develop vertical-specific apps that must reside at the edge due to latency requirements and other factors. One way that hyperscalers can deal with both data transfer and app development is through partnerships.

For their part, cloud users should consider the following factors as they continue to migrate to cloud:

The cloud migration strategy. Cloud migration isn’t just about moving to the cloud; it entails a state of continuous reinvention if cloud is to reduce costs and create new opportunities. Prepandemic, cloud migration was already often complex. Even a single application could be tied to multiple business processes, affecting vendors, balance sheets, and regulatory compliance, and different stakeholders could have different motives and expectations from the migration. A simple process could often turn into a fog of conflicting goals, broken dependencies, and cost overruns. Postpandemic, all of these factors will likely be even more challenging. It is critical to “disrupt your market without interrupting your business” during the migration.

Cloud, security, and COVID-19. As noted above, increases in cloud usage mean increases in the cyberattack surface, making security more important than ever—especially given the growth in usage driven by COVID-19. In an April 2020 survey of security professionals, 94% believed that the pandemic increases the level of cyberthreat. Almost a quarter said that the increased threat is “critical and imminent.” Only 15% believed that the cyberthreat will return to previous levels postpandemic, while five out of six believed that the new threat level is permanent.
**Cloud costs and benefits.** As multiple enterprises shifted rapidly to cloud during the pandemic, some saw costs balloon. Some companies saw costs rise by 20% to 50% just from the increase in usage, even without adding in the cost of new applications or data.\(^{28}\) As organizations migrate, there is also a cost of duplication, with organizations paying for both cloud and legacy systems at the same time as well as the cost of synchronizing data between them.\(^{29}\) Going forward, companies should think about cost planning (for instance, to take advantage of reserving instances at a discount), which can reduce expensive fixes due to rushed deployments. Cost governance systems can also help maintain control over expenses. To conduct a cost-benefit analysis, companies can use a cloud value calculator to evaluate the gaps between the current state and potential future opportunities. This can help optimize infrastructure, increase staff productivity, and enhance business value.\(^{30}\)

**New opportunities for value.** Moving to the cloud is not only allowing organizations to recover but positioning them to thrive postpandemic, increasing resilience and supporting business continuity at first, and then allowing to them to do new things and offer new services. Going forward, cloud can support benefits including collaboration, automation, scale, innovation, and agility.\(^{31}\) For example, with regard to innovation, two-thirds of respondents in a 2018 Deloitte survey said that cloud fully allowed them access to the newest technologies.\(^{32}\) Another study showed that 93% of companies surveyed used the cloud for some or all of their AI needs, requiring less investment in infrastructure and expertise.\(^{33}\)

Thanks to COVID-19 driving enterprises toward cloud, the cloud market will likely emerge from the pandemic stronger than ever. Cloud providers and others in the ecosystem have the opportunity to capitalize on increased usage to grow and flourish, while cloud users can seek to explore new ways for cloud to create value. Already, cloud has become much more than an alternative computing approach; in the near future, it is poised to become standard operating procedure for all types of businesses.

---

**SUSTAINABLY POWERING THE CLOUD**

Many pundits as well as organizations, such as the World Economic Forum\(^{34}\) and the OECD,\(^{35}\) have called for a “green reset” following COVID-19. In that regard, cloud computing can help, not hinder, progress toward a more sustainable future.

As recently as 2018, cloud computing companies were attracting criticism for their energy consumption, with some assigning them the epithet “energy hogs.” More recent articles, however, have pointed out that these concerns are unjust. Cloud computing output rose by 600% between 2010 and 2018, while its energy use rose only 6%.\(^{36}\)

Further, not only is cloud’s energy use growing more slowly than its computing output, but cloud computing often uses less energy than the noncloud alternative. On average, using the public cloud for office productivity applications emits about 6–7 kg of CO\(_2\) per employee per year, while a nonvirtualized on-premise solution would be responsible for nearly 30 kg of CO\(_2\) per employee per year.\(^{37}\) In the consumer world, as of 2019, viewing a two-hour Netflix movie consumed just over eight watt-hours of energy from Netflix and its distribution partners.\(^{38}\) This is a fraction of the energy needed to drive to a movie theater or even of taking public transit.
The key measure for measuring the environmental intensiveness of cloud computing is power usage effectiveness (PUE), in which great strides have been made. The average annual PUE for large data centers has fallen by 36% over 13 years, from 2.5 in 2007 to 1.59 in 2020 (lower numbers are better). Hyperscale data centers do even better on the PUE scale, at around 1.1–1.2. That said, this improvement has slowed in recent years, and both hyperscale cloud companies and chip companies are working hand in hand to keep power consumption as low as possible.

Two megatrends account for much of cloud computing’s energy demands. The first is obvious: if overall use of the cloud doubles, then, all other things being equal, energy use doubles as well. The second is that AI is making up a higher and higher percentage of cloud computing. Between mid-2012 and 2018, the cloud computing resources used by state-of-the-art AI grew by 300,000 times. Much of this can be attributed to the increasing use of a kind of AI called machine learning—more specifically, a subset of machine learning called deep learning. In the past decade, deep learning has become more widely used in the cloud, offering tremendous new functionality. But the way that deep learning has become more powerful is by using much larger data sets for training, which means more computation, which in turn results in greater energy use. According to one Canadian AI researcher, “The concern is that machine-learning algorithms in general are consuming more and more energy, using more data, [and] training for longer and longer.”

Several avenues exist to keep cloud’s rising energy needs under control:

**Improving overall transistor efficiency.** Over time, transistors keep getting smaller. Today, a state-of-the-art 10-nanometer (nm) semiconductor device fabrication node is about 20–30 times more power-efficient than a 90-nm node from 15 years ago. Over time, we expect data center chips to keep becoming still more efficient.

**Moving to special AI chips.** Initially, all AI computing took place on general-purpose central processing unit (CPU) chips. Over time, more specialized AI-specific kinds of chips were put into data centers: graphics processing units (GPUs, used for both training and inference), field programmable gate arrays (FPGAs, mainly used for inference), and application-specific integrated circuits (ASICs, used for both training and inference.) Although power efficiency varies by the exact chips and tasks, at a high level, GPUs are 10–100 times more efficient than CPUs at training and 1–10 times as efficient at inference; FPGAs are 10–100 times more efficient at inference; and ASICs are 100–1000 times more efficient at both training and inference. As more of these AI-specific chips are deployed in data centers, data center efficiency should keep improving.

**Improving special AI chips.** Moving from CPUs to specialized AI chips improves AI power efficiency, but AI chip technology is not standing still either. As one example, the next-generation AI GPU Tesla T4 from Nvidia doubled the energy efficiency for inference AI compared to its predecessor, Tesla P4.

**Moving some AI processing to the edge.** At one time, almost all AI was performed on the cloud, as edge devices such as smartphones, cameras, sensors, and robots lacked the onboard processing capacity for complex AI. By 2024, in contrast, more than 1.5 billion devices will have dedicated edge AI chips (called neural processing units, or NPUs) in them by 2023. These NPUs, mainly used for inference, will likely reduce the demand for energy over time: They tend to be more efficient than the chips in data centers, and they eliminate the need to send large data files up to the cloud, which consumes more energy. And not only are edge AI chips relatively efficient already, but they are progressing rapidly, with one company announcing that its current edge AI solution is 25 times more efficient than last year’s model.
Endnotes

1. Jared Spataro, “2 years of digital transformation in 2 months,” Microsoft, April 30, 2020. The cloud growth figures in the first two paragraphs are from Deloitte’s analysis of publicly reported quarterly financial statements.


6. Eric Jhonsa, “Public clouds are bright spot as IT outlays slow due to virus,” The Street, April 3, 2020.

7. Deloitte analysis of publicly reported quarterly financial statements.


9. Deloitte analysis of publicly reported quarterly financial statements.

10. Sandvine, COVID internet phenomena spotlight report, accessed November 4, 2020. In the previous paragraph, the semiconductor growth figures are from Deloitte’s analysis of publicly reported quarterly financial statements.


17. Deloitte analysis of publicly reported quarterly financial statements.

18. Ibid.


24. Chris Arkenberg et al., *Unbundling the cloud with the intelligent edge: How edge computing, AI, and advanced connectivity are enabling enterprises to become more responsive to a fast-moving world*, Deloitte Insights, September 8, 2020.


32. Gillian Crossan et al., *Accelerating agility with XaaS: Many companies are using IT as-a-service to steer their way to competitive advantage*, Deloitte Insights, 2018.


36. Steve Lohr, “Cloud computing is not the energy hog that had been feared,” *New York Times*, February 27, 2020.


38. Netflix used 451,000 megawatt hours to deliver video in 2019, and Netflix viewers consumed about 112 billion hours during the year. See: Netflix, *Environmental social governance*, 2019; Elsa Keslassy, “Netflix’s Cindy Holland says subscribers watch an average of two hours a day,” *Variety*, March 11, 2019.


41. Will Knight, “AI can do great things—if it does not burn the planet,” Wired, January 21, 2020.


44. Khan and Mann, *AI chips*, p. 23.


5G is not hazardous to your health
Busting the radiation risk myth
Paul Lee, Kevin Westcott, and Cornelia Calugar-Pop

As 5G becomes more widespread, some have sounded alarm bells about its supposed health hazards. Two main concerns have been voiced, both related to the radiation associated with the technology. The most common perception is that 5G causes cancer. The second fear is that 5G-emitted radiation weakens the immune system, enabling COVID-19 to spread.

Both of these fears, in our view, are grossly overblown. We predict that in 2021, it is very unlikely that the radiation from 5G mobile networks and 5G phones will affect the health of any single individual, be it a 5G user, a user of any other generation of mobile phones, or any individual in the vicinity of a mobile network but not actually using a mobile device. There is no link between the growth in COVID-19 infections and the roll-out of 5G networks.

Unfortunately, while extensive scientific evidence proves that mobile phone technologies have no adverse health impacts—not just for 5G but also earlier generations—we also predict that between 10% and 20% adults in many advanced economies will mistakenly equate 5G with possible harm to their health. A Deloitte consumer poll in May 2020 found a fifth or more adults in six out of 14 countries surveyed agreed with the statement “I believe there are health risks associated with 5G” (figure 1).

Radiation ≠ radioactivity

What has led to this high level of concern about 5G’s health impacts? It may have its roots in a simple confusion of terminology.

Radiation is commonly assumed to involve radioactivity, but this is not necessarily the case. The confusion arises because the words “radiation” and “radioactivity” have been conflated, a confusion that has persisted since the dropping of the Hiroshima and Nagasaki nuclear bombs in 1945. As a result of these events, as well as incidents at several nuclear power plants in the 75 years since, the term “radiation” has often mistakenly become associated with destruction on a massive scale. In a similar vein, “radiation”—in reality, radioactivity—is also associated with cancer, reinforcing the word’s connotation of mortal peril.

The point here is that radiation is not inherently radioactive. The literal definition of radiation is merely any process by which an object (from a human to a radiator to a star) emits energy (whether heat, light, or radioactive particles), which then travels through a medium (such as the air or hot water) and is absorbed by another object (a human body, a frying pan, a steam engine, or anything else). Radiation thus spans multiple everyday processes to which most people do not give a moment’s consideration. In our daily lives,
FIGURE 1

A substantial proportion of consumers in advanced economies believe that 5G can be harmful to their health

Agreement with the statement "I believe there are health risks associated with 5G," by country

<table>
<thead>
<tr>
<th>Country</th>
<th>Strongly agree</th>
<th>Tend to agree</th>
<th>Neither agree nor disagree</th>
<th>Tend to disagree</th>
<th>Strongly disagree</th>
<th>Don't know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>17%</td>
<td>19%</td>
<td>18%</td>
<td>16%</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>Belgium</td>
<td>18%</td>
<td>16%</td>
<td>25%</td>
<td>8%</td>
<td>10%</td>
<td>23%</td>
</tr>
<tr>
<td>Italy</td>
<td>13%</td>
<td>15%</td>
<td>30%</td>
<td>15%</td>
<td>13%</td>
<td>14%</td>
</tr>
<tr>
<td>Poland</td>
<td>11%</td>
<td>14%</td>
<td>25%</td>
<td>20%</td>
<td>17%</td>
<td>14%</td>
</tr>
<tr>
<td>Ireland</td>
<td>9%</td>
<td>11%</td>
<td>25%</td>
<td>19%</td>
<td>21%</td>
<td>16%</td>
</tr>
<tr>
<td>Germany</td>
<td>8%</td>
<td>12%</td>
<td>23%</td>
<td>16%</td>
<td>20%</td>
<td>20%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>6%</td>
<td>11%</td>
<td>22%</td>
<td>15%</td>
<td>28%</td>
<td>18%</td>
</tr>
<tr>
<td>Sweden</td>
<td>7%</td>
<td>10%</td>
<td>17%</td>
<td>9%</td>
<td>32%</td>
<td>25%</td>
</tr>
<tr>
<td>Australia</td>
<td>6%</td>
<td>10%</td>
<td>25%</td>
<td>17%</td>
<td>25%</td>
<td>18%</td>
</tr>
<tr>
<td>Norway</td>
<td>5%</td>
<td>10%</td>
<td>17%</td>
<td>11%</td>
<td>39%</td>
<td>17%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>6%</td>
<td>8%</td>
<td>25%</td>
<td>17%</td>
<td>26%</td>
<td>18%</td>
</tr>
<tr>
<td>Japan</td>
<td>4%</td>
<td>9%</td>
<td>39%</td>
<td>16%</td>
<td>9%</td>
<td>24%</td>
</tr>
<tr>
<td>China</td>
<td>4%</td>
<td>9%</td>
<td>32%</td>
<td>30%</td>
<td>21%</td>
<td>4%</td>
</tr>
<tr>
<td>Finland</td>
<td>5%</td>
<td>8%</td>
<td>15%</td>
<td>18%</td>
<td>36%</td>
<td>17%</td>
</tr>
<tr>
<td>Denmark</td>
<td>4%</td>
<td>9%</td>
<td>18%</td>
<td>16%</td>
<td>27%</td>
<td>27%</td>
</tr>
</tbody>
</table>

Note: Weighted base: Respondents who own or have access to a mobile phone or smartphone in Australia (1,915), Austria (952), Belgium (1,909), China (1,880), Denmark (518), Finland (520), Germany (1,868), Italy (1,902), Ireland (948), Japan (1,791), Netherlands (1,953), Norway (475), Poland (1,909), Sweden (903), United Kingdom (3,841).

most of us are regularly exposed to multiple types of radiation, most commonly in the form of sunlight.

By this definition, 5G does generate radiation, but at very safe levels, and none of it is radioactive radiation. 5G base stations and phones, and the frequency ranges within which 5G operates, are very likely to be operating well within safe parameters in 2021 and throughout 5G’s lifetime, which may extend to two decades. Radiation within these parameters does not significantly raise the risk of cancer. It also does not weaken the immune system, and thus has not contributed to the spread of COVID-19.

Radio waves, 5G, and health: A new variation on a familiar theme

Understanding in more detail why 5G and other cellular mobile technologies are safe requires an elementary knowledge of how mobile networks and similar radio transmission-based services work.

Mobile phone networks are a variation on a conventional radio setup. A central transmitter relays sound via radio waves to a device equipped with an antenna (indeed, what is commonly referred to as “radio” today is a reference to the use of radio waves to distribute radio stations’ content). Radio waves, also known as radiofrequency electromagnetic fields (EMF), are a form of radiation, but a harmless one. Traditional television also uses the same distribution technique: A central transmitter relays sounds and images that are then received by antennas.

Today, billions of people receive TV and radio content via a network of transmitters that connect with receivers in homes, offices, and vehicles around the world. Although the proportion of video content delivered on demand has steadily risen over the past decade, most of the hours of video consumed globally are still likely to be transmitted wirelessly via national networks of transmission towers.

All generations of mobile phone technology, including 5G, take the same familiar underlying approach. A mobile network consists of a nationwide grid of cell sites. These sites feature transmitters that generate radio waves, which are received by a mobile phone’s antenna.

The radio waves generated by mobile networks, TV stations, and radio stations are innocuous. On the very broad spectrum (known as the electromagnetic spectrum) in which radiation exists, radio waves fall on the low-frequency, very low-energy end. Such radiation is sometimes referred to as nonionizing radiation. This contrasts with radiation such as X-rays, gamma rays, and some types of ultraviolet light, which fall into the high-frequency, very high-energy end (figure 2). These types of radiation are referred to as ionizing radiation, so called because it has sufficient energy to damage DNA by removing electrons from atoms, potentially leading to cancer.

Admittedly, one difference between broadcast radio and TV and mobile telephony is that the receiving device can also transmit. But even this capability is not wholly new. Walkie-talkies were first used in the 1940s; mobile phone networks, though designed to be scalable to whole countries and used for person-to-person calls, work on the same principle.

Another minor variation is the reach of each transmitter. For television, most transmitters in use today have a range of 65 to 90 kilometers. For FM radio, the range is likely to be up to 45 kilometers. Mobile phone tower transmitters, in contrast, typically have a range of between 50 meters and 20 kilometers, with the majority being low-range transmitters of less than 500 meters. The average range per base station has generally decreased over time as the number of base stations has increased, with the majority of additional base...
stations covering much less territory. With the rollout of 5G networks, cell size may be as little as 10 meters in radius (known as small cells), with a transmission power of 100 milliwatts.3 (The rationale for reducing cell size is to enable higher performance with regard to download speeds or the number of users per square kilometer.) With small cells, the base stations are small enough to be wall-mountable or attached to lampposts.4 As cell size decreases, transmission power required declines.

The reason for mentioning these similarities is to point out that mobile telephony, including the latest 5G standard, relies on the same underlying transmission methods that have been used for decades. Content is created, relayed over radio waves, and received—a technique that has been delivering content wirelessly for more than 100 years.

Like the technology itself, concerns about the health effects of wireless transmission are not new. Some individuals were concerned about the impacts of earlier mobile network generations as well as of other types of wireless networks, principally Wi-Fi and TETRA (a type of private radio communications network often used by emergency services). Looking further back still, some people worried about the health impacts of emissions from television transmitters too.5

A common concern going back decades has been the risk of brain and skin cancer from mobile phones. However, this concern has been demonstrated to be unproven. A 2019 study of mobile phone use and the incidence of brain tumors in Australia found no increase in the incidence of brain tumors since the 1980s. The researchers looked at the periods 1982–1992, 1993–2002, and 2003–2013, which covered the introduction of analog cellular (1G), 2G, 3G, and the beginning of 4G. Their conclusion: “[There have been] no increases in any brain tumor types, including glioma and glioblastoma, during the period of substantial mobile phone use from 2003 to 2013.”6 As for skin cancer, a 2018 review of medical studies undertaken between 1995 and 2017 found that “overall evaluations showed that

---

FIGURE 2
The radio waves generated by 5G fall on the low end of the electromagnetic spectrum

Source: Adapted from SCAMP/Imperial College London/EBU.
the effects of mobile phone radiation on skin diseases are weak and have no statistical significance.” These skin cancer studies alone comprised data from 392,119 individuals—a very large sample size.

What has been different about 5G is that the means of amplifying misunderstandings about its health impacts, either deliberately or ignorantly, are greater than ever, because the capability to share true or false information is greater than ever before. Many individuals who believe 5G may cause them harm have had these ideas suggested to them, most typically via social networks, in sensational but plausible language.

**Extremely low power, extremely low risk**

Of course, nonionizing radiation is not always completely harmless. The most common form of nonionizing radiation is visible light, which has a higher energy level than radio waves. An excess of visible light—or even radio waves—can produce heat, and in extreme cases cause burns and body tissue damage. However, the power behind the radiofrequency radiation generated by mobile networks is controlled and poses virtually no risk to consumers.

Power transmission from mobile telephony, including 5G, is far lower than that from light bulbs, TV, radio towers, or even sunlight on an overcast day. The quantity of this power is measured in watts, and a single watt is a tiny quantity of energy. The power transmitted by the mobile phones used in 2021 and into the foreseeable future can reach up to two watts, depending on the age of the phone; it can be as low as 0.001 watt, with the vast majority of devices in use this year peaking at 1 watt. By comparison, the power transmitted by CB radios, which have been in use for decades, reaches up to four watts. As with a car journey, the shorter the distance, the less the power required. A phone held next to the head or kept in a pocket would have the greatest impact. The radiation level from a phone or speaker placed on a table near the user would be lower. A smartphone will transmit more power when base stations are relatively distant, but most smartphones are used predominantly indoors, and tend to be connected to Wi-Fi routers (which are effectively miniature base stations), which are often mere meters away. In all of these cases, the amount of power transmitted is minimal—certainly much lower than required to be harmful. Further, a smartphone transmits power only when sending or receiving data, a mechanism designed to prolong battery life.

The power generated by mobile network base stations is similarly low. A base station’s transmissions range in power from a quarter of a watt for a small cell (which would often be indoors and cover a small range) to 200 watts for a minority of 5G base stations. More typically, an outdoor base station with the greatest range would have a power output of between 10 and 100 watts. The output of indoor base stations, which usually have a range of hundreds of meters or less, is much lower.

As with a phone, a base station’s power level declines with distance from its transmitter. An individual 100 meters away from a 5G macrocell antenna located at 30 meters’ height would absorb less than one microwatt (one-thousandth of a watt) of power. When one is directly next to a base station supporting any generation of mobile standard (not just 5G), exposure limits may be exceeded. But these areas are inaccessible to the public, sometimes because of their height (20 meters or higher for larger sites), their location (often at the top of buildings), or their design (because the units are enclosed). In the case of indoor base stations, excessive exposure would only happen within a few centimeters of the transmitter.
Average broadcast transmission power has declined as the number of base stations deployed has increased, resulting in a smaller distance between base stations and users. Transmitter power levels for 1G and 2G networks were far more powerful, on average, than those used for 4G or 5G, since 1G and 2G transmitters covered a far greater range, often tens of kilometers in each direction. In contrast, 4G and 5G masts in city centers and other traditionally congested areas may cover just 100 meters.

It is worth reiterating how minuscule a watt is. An incandescent bulb, which radiates light via a wire filament that is heated until it glows, is rated between 25 to 200 watts. In domestic settings, people may be less than half a meter away from a light bulb. A person this distance from a 25-watt bulb would be exposed to thousands of times more radiation than an individual who was 10 meters (unusually close) to a relatively high-powered 5G base station.14 This is not just the case in 2021—it should hold true always. Similarly, people absorb five times more radio frequency exposure from FM radio and television broadcasts than from mobile network base stations.15 The broadcaster transmitter power levels used for TV and FM radio can reach up to 100,000 watts.16 For AM radio, the transmission power may reach 500,000 watts.17

Humans have coexisted with incandescent light bulbs, and their radiation, since the 1880s with no known malign effects (except, of course, from being burnt from touching a lit bulb). As for broadcast power, the first television station went on the air in 1928,18 and the first commercial radio station launched in 192019—yet no reliable account of people being harmed by the radiation these stations generate has ever been reported.

5G is even safer than previous mobile network generations

In 2021, consumers who are concerned about the health impacts of mobile networks are likely to be most worried about 5G, the latest generation of mobile technology. However, in some ways, 5G is likely to have even lower potential health impacts than earlier generations of mobile telephony.

5G has been designed to use less power than previous generations to reduce operational costs; as a result, it emits less power as well. This is accomplished via the new, advanced radio and core architecture used in the 5G standard, with 5G networks assisting 5G devices in minimizing power transmit levels.20 5G base stations also can be put into sleep mode when there are no active users (for example, at night). This capability is not available with 4G networks, which transmit control signals even when there are no users in range.21

5G also incorporates a technique known as beamforming, an approach that involves directing a narrow beam of radio waves to the user device (such as a smartphone). This method is equivalent to directing a narrow beam of light from a pocket flashlight at a target, focusing the radio waves on the device. This method not only enables higher connection speeds, but also leads to lower radio wave exposure than prior network generations,22 which would often spread radio waves across a wide arc, similar to a car’s headlight.

Some people may conflate the risks associated with beamforming with industrial-grade laser beams. A manufacturing-grade laser beam, which is 100 million times as powerful as a typical laser pointer, is capable of melting steel.23 But beamforming in 5G networks involves innocuous levels of power.
As a final note, tests of 5G sites in 2020 by regulators such as Ofcom in the United Kingdom have found that their EMF levels are well within International Commission on Non-Ionizing Radiation Protection (ICNIRP) guidelines. ICNIRP is an independent scientific commission based in Germany that works with the World Health Organization (WHO), the International Labour Organization (ILO), and the European Commission. The highest EMF level recorded among the 22 locations tested was 1.5% of the acceptable level—in other words, 98.5% below the acceptable level. Most of the sites tested supported four generations of mobile technology; that is, a combination of 2G, 3G, 4G, and 5G (in many markets, 5G-only base stations remain relatively rare). At all of these sites, 5G contributed the least to the EMF fields measured. In 19 of the 22 locations, the highest 5G band value was less than 0.01% of the acceptable ICNIRP level.

5G and the spread of COVID-19

One myth about 5G’s impact on health that has been widely spread in 2020 is the fictional association between the roll-out of 5G and the spread of COVID-19. Put plainly, the idea that 5G transmits COVID-19 is as bogus as it is impossible. COVID-19 is a virus spread through respiratory droplets from other people. A virus does not travel via radio waves.

A variant of 5G misinformation related to COVID-19 is that 5G emits radiation that weakens people’s immune systems, making them more susceptible to illness. This is similarly false.

It is likely that misinformation about 5G’s relationship to COVID-19 will be as pervasive in 2021 as it was in 2020. An Ofcom survey at the end of June 2020 found that 29% of respondents had come across false or misleading information about COVID-19 in the prior week. The most common topic, seen by 21% of respondents, was “theories linking the origins or causes of COVID-19 to 5G technology.” Misinformation about 5G was even more prevalent earlier in the year: An Ofcom survey undertaken from April 10–12, 2020, found that 50% of respondents had seen false or misleading statements about 5G. The good news is that these people recognized the misinformation as such. The bad news is that the majority (57%) of those who saw what they regarded as misinformation did nothing about it.
THE BOTTOM LINE

Our research shows that understanding of 5G's benefits is low in multiple markets, with up to two-thirds of adults stating that they do not know enough about 5G in general as of mid-2020. Among women, the proportion is even higher, at three-quarters. The lack of understanding also peaks among older users.

In the vast majority of cases, we would hope that the widespread provision of accessible but comprehensive information about how 5G and other wireless technologies work would put consumers' minds at rest. Mobile operators, mobile handset providers, telecommunications regulators, government communications bodies, and science programs on broadcast and on-demand platforms could, in combination, counter the vast tide of misinformation about 5G. Individual companies and regulators could also work together to constrain the ability to share misinformation, despite the commercial impact that this may have.

What would make an information campaign about 5G successful? Among other things, it would need to be both proactive and reactive. It would need to be designed for all types of users, not just those with a science background. And misinformation would need to be met with information of a similar caliber. Headlines need to be met with headlines. Rebuttals to fake claims should be expressed using similar channels and similar language. Misinformation spread by well-known individuals should be countered with information from well-known individuals. Celebrities with marketing relationships on the basis of significant social media presence could be asked to become the figureheads of information campaigns. Soap operas with mass market reach could include storylines that feature discussions and explanations about 5G. Without such widespread and accessible efforts, the facts may be ignored. It may not be enough only to post a link to ICNIRP.

Information campaigns also should do more than explain why 5G is safe. They should also educate people about its positive applications—for example, making everyday mobile applications, such as browsing and maps, notably faster. Carriers could also talk about how 5G could make other mainstream applications, such as driving, easier and safer. (A car could provide regular status updates, including video footage, to manufacturers, which could enable them to identify flaws faster.) The telecom industry could also showcase the many applications that 4G has enabled and that consumers have come to depend on.

There should also be debates on effective ways of preventing the proliferation of misinformation on social media. As social media becomes a far more common source of news, checks and balances that control for accuracy while still permitting freedom of speech are becoming increasingly important. Offering the ability to fact-check is one way to help social media users filter facts from fiction, but all readers cannot be relied on to have the discipline to verify.

It may not be possible to persuade everyone that 5G is safe. There is likely to be a niche—perhaps less than 1% of the population—that will remain convinced not just that wireless technologies are harmful, but that their deployment is deliberate and that the intent is to cause harm. Unfortunately, while such niche views have in the past lacked widespread amplification, social media has often provided the mechanism for conspiracy theories to flourish and proliferate. If education is to be effective in curbing popular fears, it should be compelling, consistent, and pervasive, and it should begin now.
### FIGURE 3

**Understanding of 5G is low in multiple markets**

Agreement with the statement “I do not know enough about 5G,” by country

<table>
<thead>
<tr>
<th>Country</th>
<th>Strongly agree</th>
<th>Tend to agree</th>
<th>Neither agree nor disagree</th>
<th>Tend to disagree</th>
<th>Strongly disagree</th>
<th>Don't know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>34%</td>
<td>32%</td>
<td>18%</td>
<td>8%</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>Sweden</td>
<td>33%</td>
<td>32%</td>
<td>18%</td>
<td>8%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>27%</td>
<td>37%</td>
<td>20%</td>
<td>10%</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>Ireland</td>
<td>25%</td>
<td>38%</td>
<td>18%</td>
<td>10%</td>
<td>7%</td>
<td>2%</td>
</tr>
<tr>
<td>Australia</td>
<td>25%</td>
<td>37%</td>
<td>19%</td>
<td>12%</td>
<td>6%</td>
<td>2%</td>
</tr>
<tr>
<td>Norway</td>
<td>28%</td>
<td>32%</td>
<td>23%</td>
<td>10%</td>
<td>6%</td>
<td>1%</td>
</tr>
<tr>
<td>Denmark</td>
<td>33%</td>
<td>27%</td>
<td>22%</td>
<td>9%</td>
<td>6%</td>
<td>4%</td>
</tr>
<tr>
<td>Netherlands</td>
<td>27%</td>
<td>32%</td>
<td>22%</td>
<td>10%</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>Japan</td>
<td>24%</td>
<td>35%</td>
<td>22%</td>
<td>11%</td>
<td>3%</td>
<td>5%</td>
</tr>
<tr>
<td>Austria</td>
<td>26%</td>
<td>32%</td>
<td>19%</td>
<td>13%</td>
<td>8%</td>
<td>3%</td>
</tr>
<tr>
<td>Germany</td>
<td>23%</td>
<td>34%</td>
<td>22%</td>
<td>12%</td>
<td>5%</td>
<td>3%</td>
</tr>
<tr>
<td>Poland</td>
<td>20%</td>
<td>36%</td>
<td>22%</td>
<td>12%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>Finland</td>
<td>20%</td>
<td>34%</td>
<td>19%</td>
<td>15%</td>
<td>6%</td>
<td>6%</td>
</tr>
<tr>
<td>Italy</td>
<td>18%</td>
<td>35%</td>
<td>27%</td>
<td>12%</td>
<td>6%</td>
<td>3%</td>
</tr>
<tr>
<td>China</td>
<td>7%</td>
<td>38%</td>
<td>30%</td>
<td>19%</td>
<td>5%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Note: Weighted base: Respondents who own or have access to a mobile phone or smartphone in Australia (1,915), Austria (952), Belgium (1,909), China (1,880), Denmark (518), Finland (520), Germany (1,868), Italy (1,902), Ireland (948), Japan (1,791), Netherlands (1,953), Norway (475), Poland (1,909), Sweden (903), United Kingdom (3,841).

Endnotes


4. Very small base stations are also known as femtocells. Larger base stations are picocells, microcells, and macrocells. With each generation of cellular technology, base station equipment has become progressively smaller.


7. For more information on the methodology, and for a more detailed review of the conclusions, see A. Keykhosravi et al., “Radiation effects of mobile phones and tablets on the skin: A systematic review,” Advances in Medicine, April 2018.


12. Behnke, “Is this anything to worry about?”


14. Behnke, “Is this anything to worry about?”

15. According to the WHO, this is because the frequencies used for FM transmission and TV are typically lower than for mobile telephony, that is at around 100 MHz, 300 MHz–400 MHz, and 900–1800 MHz and up respectively. See World Health Organization, “Electromagnetic fields and public health,” accessed November 3, 2020.


17. Ibid.


The next-generation radio access network

Open and virtualized RANs are the future of mobile networks

Naima Essing, Kevin Westcott, Sanket Nesargi, and Jeff Loucks

MOBILE NETWORK OPERATORS (MNOs) are known for their ability to build and operate massive, high-performance wireless networks. They rely on highly specialized radio access and networking equipment with tightly integrated proprietary software to deliver the cellular services that connect our cell phones, tablets, computers, and other devices. But high costs, limited flexibility, and constrained vendor choice are prompting MNOs to shift away from such systems toward more open, standards-based, software-centric virtual platforms.

Many MNOs are well on the journey toward opening and virtualizing their core networks, achieving significant operational gains. They now have their sights on their distributed mobile edge networks: the radio access network (RAN). And because MNOs must replace or augment existing RAN equipment to deliver 5G service, they have the opportunity to adopt open and virtualized RAN architectures—which we will refer to simply as “open RAN”—as part of these deployments.

The open RAN market is still in its early days. We estimate that there are currently 35 active open RAN deployments across the globe,1 many of which involve MNOs testing open RAN in greenfield, rural, and emerging markets. Although deployments are starting slowly, they could easily double in 2021. While it may take anywhere from three to five years for the technology to fully mature, open RAN adoption should accelerate rapidly thanks to the logic of its network design and its strategic alignment with carrier needs. Economic and competitive forces are also converging to drive the market forward. If this trend continues, the open RAN market has the potential to grow substantially, with some estimating double-digit growth rates2 that will push open RAN to approach 10% of the total RAN market by 20253 from less than 1% today.4 Moreover, if governments force MNOs to replace installed 5G RAN equipment from restricted vendors, the growth rate may be even higher.

Why open RAN?

At its most basic level, the RAN architecture at the mobile network edge comprises a remote radio unit (RRU or RU) at the top of a cell tower that communicates with a baseband unit (BBU) located at the tower’s bottom. The RAN uses proprietary hardware and vendor-defined communication interfaces, and its software-driven functionality is tightly integrated inside the hardware.

While these traditional systems have worked well for MNOs, they have many drawbacks. Making any upgrade or change to the wireless network, even seemingly minor ones, requires replacing physical hardware throughout the network—a costly, manual, and time-consuming process. Moreover, the proprietary nature of the equipment and
interfaces that connect the hardware locks MNOs into existing relationships with the vendor that originally supplied them.

Virtualizing the RAN and replacing proprietary interfaces with standards-based interfaces enables equipment interoperability and multivendor RAN deployments. This gives network operators more flexibility to pick and choose among best-of-breed solution providers. By opening the market, currently dominated by a handful of vendors, to new suppliers, open RAN can not only lower costs but also prompt greater innovation through competition, as well as allow MNOs to avoid restricted vendors.\(^5\) Additionally, because they allow operators to use software to push out network functions and intelligent automation, virtual architectures can speed the roll-out of new services that can help carriers better manage their networks, improving network performance.

Open RAN is not an entirely new idea; MNOs have discussed the concept of an open RAN architecture for decades. But despite open RAN’s appeal, adoption has hitherto been slow and met with skepticism due to technical engineering and integration challenges. Substantial confusion over the terminology and available technology options has also hindered adoption.

Now, however, open RAN’s momentum is growing as the ecosystem develops, partnerships form, suppliers ramp up investments, and operators commit to experimentation, trials, and deployments. Over the past several years, aggressive experimentation through both lab trials and live deployments are closing performance gaps between open and proprietary RAN solutions, steadily tearing down perceived barriers. Rising capital costs and national security concerns that further limit financial flexibility, as well as the rise of government policies to support vendor choice, are also accelerating the movement toward virtual and open RAN architectures. Finally, open RAN is riding the wave of several technology trends, including 5G, cloud virtualization, distributed edge computing, and artificial intelligence (AI)–driven automation. All of these factors can help push open RAN from just a cool idea toward reality.

### The taxonomy of virtual and open RANs

As is common with emerging technologies, the open RAN taxonomy is fluid. Evolving developments in engineering, configurations, and standards have led to conflicting terminology. Here, we will attempt to untangle the language needed to better discuss and understand the technology.

Open RAN encompasses two underlying concepts: virtualization and openness.

**Virtual RAN** decouples software-driven functionality from the underlying hardware, replacing purpose-built hardware with a programmable RAN infrastructure built with...
low-cost, general-purpose hardware. This allows operators to use a single virtualized BBU to support multiple radios instead of needing a proprietary physical BBU with fixed functionality at every cell site. These virtual architectures facilitate the dynamic introduction and administration of software-based services at RAN edge networks without having to replace the underlying hardware.

**Open RAN** takes virtualization a step further. It not only disaggregates software applications from the underlying hardware infrastructure, but also replaces the proprietary communication interfaces between baseband components—the centralized unit (CU), the distributed unit (DU), and the radio units (RUs)—with open, standards-based interfaces. Open and standards-based interfaces enable operators to source the radio, baseband, and software from different vendors with plug-and-play interoperability.

A RAN can be virtualized but not open. Virtualizing and opening RAN edge networks are separate decisions. An operator can virtualize the RAN by disaggregating software functionality from the underlying RAN hardware and migrating to a cloud-native architecture **with or without** opening and standardizing the communication interfaces. For purposes of this article, however, we use the term “open RAN” to refer to a virtualized and open RAN architecture that gives MNOs both the flexibility to virtually manage RAN baseband components and the choice to source baseband hardware and software components from different vendors (figure 2).

Many would argue that virtualization is only a first but necessary step in the journey toward an eventual full opening. As long as the interfaces remain closed and controlled by the vendor, new entrants cannot participate.

Besides different degrees of openness, there are also multiple “flavors” of open RAN. For example, the O-RAN Alliance (with O-RAN) and the Telecom Infrastructure Project (with OpenRAN, not to be confused with our use of “open RAN”) promote specific standards-based open RAN architectures.

---

**FIGURE 2**

**A RAN can be virtualized but not open**

- Virtualized baseband on commodity hardware
- Interfaces within the RAN are not open
- Virtualized baseband split into DU and CU
- Open interface between CU and DU
- Open interfaces between DU, CU, and RU
- Compliant with specifications from the O-RAN Alliance
- Open interfaces between DU, CU, and RU
- Compliant with specifications from the TIP OpenRAN

Note: CU = centralized unit, DU = distributed unit, RU = radio unit
Source: Deloitte analysis.
reference architectures, standards, and protocols that seek to foster vendor interoperability.

Multiple converging factors are driving open RAN adoption

Several main drivers—each, however, associated with a set of hurdles—are helping to accelerate open RAN adoption.

VIRTUAL RAN ARCHITECTURES LOWER TOTAL COST OF OWNERSHIP

One of the most compelling value propositions of virtual RAN architectures, open or closed, is in their potential to lower the total cost of ownership of RAN networks. These cost reductions can derive from sources such as:

Lower upfront capital deployment costs. With rising capital intensity and slowing subscriber and revenue growth, MNOs view virtualized RAN as an important lever for fundamentally changing network economics. For instance, lowering capital costs is key to the optimal deployment of next-generation 5G wireless networks. 5G will require the addition of approximately three to four times more cell sites, albeit generally smaller cells on rooftops, lamp posts, and utility poles. Greater network density is needed to achieve 5G’s promise for enhanced coverage, capacity, speed, and low latency, as well as to overcome the limited propagation characteristics of high-band spectrum. Estimates call for the addition of more than 2 million 5G cell sites in the United States by 2021, up from roughly 200,000 today. By allowing operators to aggregate baseband functionality using a single virtualized BBU to support multiple radios, open RAN reduces overall hardware cost and enables a smaller, simpler, and more energy-efficient installation footprint.

Virtual architectures can also “future-proof” investments in the physical network. Operators can use software to upgrade RAN features and functionality on the same physical infrastructure to keep pace with changing market conditions instead of having to rip out and replace whole physical systems.

Opening RAN architecture interfaces introduces vendor competition that can further reduce hardware costs. Open RAN allows operators to pick and choose among vendors—not just traditional telecom vendors, but big tech companies such as hardware manufacturers, webscalers, original design manufacturers, and others looking to enter the market. The ability to change out individual RAN components with off-the-shelf hardware from any vendor can improve flexibility as well as reduce costs and downtime for system scaling and maintenance. That said, these savings from vendor choice are theoretical. Many industry pundits argue that the additional cost, time, and effort to test and integrate multivendor systems could well offset, if not completely negate, any benefits from vendor diversity. Additionally, the performance of processing intensive RAN capabilities on general-purpose hardware may not match what can be achieved on bespoke optimized hardware platforms.

Lower operating expenses through automation. Open RAN has the potential to reduce ongoing network operating and maintenance expenses while simultaneously addressing the conflicting challenge of rising data traffic growth and customer expectations. Software-mediated RAN architectures empower operators with new levels of operational flexibility and intelligent automation that fundamentally shift how they manage networks. Programmable RAN infrastructure also makes it simpler and more cost-effective to roll out new features and functions at distributed RAN locations at a mobile network’s edge. Moreover, open interfaces enable these new network features and functions to operate on any vendor’s hardware without having to send out
engineers and technicians for vendor-specific integration, as is common practice today. Thus, open RAN could replace much of the time-consuming and manual work of maintaining, upgrading, and optimizing networks with light-touch, centrally managed, automated computing processes.

This type of automation, however, comes with its own challenges. To achieve these types of operational efficiencies, operators should become more adept in IT-style systems management and software engineering, imitating cloud service providers whose velocity in innovation showed the world how to drive value from network platforms. Operators should also adopt service-oriented practices using rapid-fire DevOps and continuous innovation and delivery (CI/CD) practices to deploy new applications with speed and precision—something operators did not have to do with more traditional RAN deployments.12

Not all operators are eager to take a more hands-on approach toward operations. For instance, anything that could interfere with network reliability is likely to give pause to MNO adoption.13 To reduce this risk, some carriers prefer the ease and simplicity of traditional systems in which they rely on a few trusted vendors to provide new, fully tested, carrier-grade solutions with turnkey deployment, maintenance, and integration support. Should something go wrong, these operators can lean on their one vendor for remediation instead of having to identify and isolate the issue, then chase down the culprit from a pack of small, unproven vendors who may point the finger at anyone but themselves.

OPEN RAN ENCOURAGES INNOVATION
Beyond the capex and opex savings discussed above, open RAN also drives faster innovation. Instead of having to replace network gear to introduce new features and functions, MNOs using open RAN can use software updates on white-box gear to affect change, materially shortening upgrade and innovation cycles. Moreover, vendor interoperability negates the need to send out technicians for custom onsite integrations, further reducing the time, effort, and cost of launching new products and services.

Suppliers also benefit from open RAN because it opens up market participation and lowers barriers to entry. Because of interoperability, vendors can develop products and solutions for use by multiple operators instead of having to create unique one-offs for a specific operator. Interoperability also fosters best-of-breed solutions because vendors can focus on what they do best—whether hardware, software, or silicon—instead of having to develop an entire integrated end-to-end system.

Perhaps most importantly, open RAN systems enable MNOs to leverage insights from the traffic flowing across their networks to develop solutions that improve network performance. Open interfaces encourage third-party development of AI/machine learning-driven solutions, which help operators deal with an increasing array of bandwidth-intensive applications and the explosion of data flowing over ubiquitous networks and devices. Applications such as RAN intelligent controllers (RICs) and self-optimizing networks (SONs) are emerging as essential and cost-effective methods to manage future network complexity. The ultimate objective is to lower costs and achieve super-lean operations with zero-touch, fully automated end-to-end network management and service orchestration using AI-optimized closed-loop automation.14

Open RAN innovation offers operators additional opportunities to evolve from merely providing commoditized “dumb pipe” connectivity to delivering differentiated customer experiences. For instance, MNOs might offer enterprise customers networks optimized for specific use cases.
Examples of these could include ultra-reliable networks with near–real-time response for factory robotics or pervasive low-power wide-area networks for widespread monitoring of assets such as gas pipelines or oil rigs. While it remains to be seen whether enterprises would take up, let alone pay for, these types of services, they do have the potential to open the enterprise market to operators—a market in which they have not been historically active.

While innovations made possible by open RAN could generate new revenue, it also introduces the risk of competition from new entrants. Although “coopetition” currently characterizes the competitive landscape, open RAN makes it easier for alternative players (webscalers, equipment vendors, system integrators, and so on) to capitalize on the greater availability of new low-cost wireless equipment to disrupt the mobile communications market. Rakuten and DISH Networks, for example, plan to use open architectures to build low-cost alternative networks in direct competition with incumbent MNOs. And better access to lower-cost wireless equipment was the initial inspiration for Facebook to start the Telecom Infrastructure Project.

OPEN RAN DRIVES GREATER VENDOR DIVERSITY AND SUPPLY CHAIN SECURITY

Consolidation over the years has concentrated the RAN vendor market to five major players: Huawei, Ericsson, Nokia, ZTE, and Samsung. Together, they account for more than 95% of the market, with the top three vendors, Huawei, Ericsson, and Nokia, controlling 80% of the market.15

Market concentration has come to the forefront of political debate in the United States, where the government advisory and prohibition against using federal funds to purchase communications equipment or services from companies that pose a national security risk has restricted business with Chinese equipment manufacturers.16 This effectively eliminates two of the five top vendors mentioned above (Huawei and ZTE) and highlights the degree of market concentration. Additionally, in April 2020, the US Department of State announced the 5G Clean Path initiative, which restricts the use of untrusted vendors in the transmission, control, computing, and storage equipment of all 5G mobile traffic entering or exiting American diplomatic systems at home or abroad.

The United States is not alone in making such moves. Multiple government initiatives around the globe are aiming to restrict the use of untrusted vendors. In July 2020, the United Kingdom announced that it would ban new purchases and require the complete removal of restricted kit from UK networks. And Australia, New Zealand, and Japan all effectively ban the use of untrusted vendors from their 5G deployments.

With more countries restricting vendors, the urgency for a new approach is driving greater worldwide interest in open RAN. To allow for alternatives, US policymakers increasingly seem to favor open RAN initiatives. They prefer the market development of alternative vendors to expand the supplier ecosystem that can give MNOs greater flexibility and choice.

For US MNOs, the lack of a US “national champion” equivalent to the largest players could become problematic if trade tensions escalate and national security is tied to homegrown network providers. However, even though it lacks a major integrated wireless equipment manufacturer, the United States is home to some of the most prominent emerging open RAN startups, including Altiostar, Mavenir, and Parallel Networks. Most of these suppliers offer open RAN networks compliant with the O-RAN architecture. Moreover, the United States boasts many of the most significant players in the hardware, silicon, and software supply chain ecosystem. These companies can partner with one another to cobble together an end-to-end carrier-grade open RAN solution.
The industry is uniting around open RAN

By opening the market and introducing competition, open RAN sets up a rift between traditional incumbents and forward-looking new entrants. Most of these new entrants, however, still need to establish themselves in the market; for now, the balance of power rests firmly in the camp of a few traditional vendors, which may, in fact, emerge stronger. Nonetheless, open RAN will likely force incumbent vendors to shift their business models away from a hardware to a more software-centric approach, introducing new business and competitive risks through the transition.

This tension is creating momentum for several industry-led open RAN initiatives that seek to unite an ecosystem of supply chain partners and advance open RAN through the definition, development, and testing of standards and reference architectures. Beyond the standards defined by the 3rd Generation Partnership Project (3GPP), multiple industry groups are leading the open RAN movement, each with a different purpose. Prominent industry-led open RAN initiatives include:

**O-RAN Alliance.** This alliance, formed in early 2018, is a worldwide carrier-led effort that seeks to define new radio architectures. Its main objective is to open designs and interfaces between the RRU and BBU. It also focuses on vendor interoperability.

**Telecom Infrastructure Project (TIP).** Launched in early 2016 by Facebook, the TIP has more than 500 members and 12 project groups. Its OpenRAN project group focuses on building white-box baseband and radio unit designs based on O-RAN Alliance architecture and interfaces. TIP’s primary goals are to develop an ecosystem to spur innovation, enable supplier diversity, and reduce deployment and maintenance costs across access, transport, and core networks.

**Open RAN Policy Coalition.** The Open RAN Policy Coalition, launched in mid-2020, advocates for government policies to help drive open RAN adoption. Its growing membership spans operators, equipment manufacturers, software developers, and silicon chip makers.

**Open Networking Foundation (ONF).** In August 2020, the ONF announced several new initiatives in the open RAN domain. This group is looking to deliver open-source implementations of functionality included in open RAN components such as CU, DU, and RICs.

From a technical perspective, the O-RAN Alliance’s work is the most foundational, prompting partnerships with many other organizations. The TIP announced a liaison agreement with the O-RAN Alliance in early 2020 that will allow the two groups to share information and hopefully prevent duplication of effort. In mid-2020, the telecom industry lobby group GSMA announced that it would partner with the O-RAN Alliance to accelerate the adoption of open RAN technologies. The ONF has also indicated it will work with the O-RAN Alliance to develop its solutions.

Like other aspects of open RAN, the dependencies between and interactions across these groups can be confusing. And while well-intentioned, the plethora of initiatives has the potential to further fragment the industry, with each offering slightly different flavors of open RAN. To deliver a solution that universally appeals to multiple stakeholders, it will be critical for these various organizations to harmonize their efforts and provide a simple way for operators to learn about and embrace open RAN.
A SAMPLING OF OPEN RAN VENDOR ECOSYSTEMS

Companies big and small across the telecommunications ecosystem, including tier one and tier two operators, webscalers, traditional vendors, and startups from all corners of the industry, are beginning to coalesce around the open RAN model. Some of these players include:

**Traditional RAN equipment vendors.** These companies differ widely in the degree to which they embrace open RAN.

- Examples: Huawei, Ericsson, Nokia, Samsung, ZTE

**New open RAN vendors.** Several well-financed new companies are focusing on software-driven RAN architectures using plug-and-play, hardware-agnostic infrastructure to radically improve mobile network economics. These companies are slowly strengthening their reputation and positioning themselves to become key beneficiaries of the shift to open RAN.

- Examples: Altiostar, Mavenir, Parallel Network

**Network hardware, software, and component providers.** Many traditionally enterprise-focused hardware and component vendors are participating in open RAN industry initiatives to position their product portfolios to meet carriers’ emerging needs. While most of these vendors do not offer radio technology, they are seeking partnerships in which they contribute software, hardware equipment, or components in efforts to engineer a complete integrated solution.

- Examples: Cisco, Fujitsu, IBM, NEC, Samsung, Hewlett Packard Enterprise (HPE), Dell Technologies, Lenovo

**Chipset vendors.** Each chipset vendor offers distinct solutions, with several providing accelerator kits to advance open RAN adoption. Sentiment is growing for the need for significant investment in 5G RAN silicon solutions to close the performance gap between restricted and unrestricted vendors.

- Examples: Intel, Qualcomm, Nvidia, Marvell, Xilinx

**Cloud service providers.** Cloud providers initially sought to offer MNOs cloud-based virtual environments to house and run internal applications and, eventually, external software-defined solutions. Now, webscalers and MNOs are increasingly partnering with each other to provide joint enterprise-oriented solutions for specific use cases, such as the need for low latency. With the emergence of new types of wholesale wireless arrangements from new entrants like Rakuten or DISH, webscalers and over-the-top media providers may emerge as formidable competitors to traditional mobile operators. They may even one day use open architectures to develop competing networks in their quest to connect the next billion consumers. It is this vision, coupled with frustration with the high cost of telecom equipment, that led Facebook to spearhead the TIP.

- Examples: Amazon web services, Microsoft, Google, IBM, Facebook
Barriers and challenges

Many challenges to open RAN adoption exist, many of which involve highly technical engineering issues that are beyond this chapter’s scope. The following are some of the more commonly cited concerns that are slowing open RAN’s adoption among MNOs:

Carrier-grade scalability. Experimentation with open RAN has thus far been largely limited to local and regional deployments. At a small scale, open RAN’s complexity of integration and its load on RAN network functions is readily managed. Several operators are experimenting with open RAN architectures in underserved areas where the potential for stranded investment and pressure for high performance is low since there is little or no existing infrastructure. In Turkey, for example, Vodafone is working with vendor partners using agile methods to make rapid iterative updates to software and equipment configurations, tracking key performance indicators to provide evidence of and confidence in achieving performance thresholds. Whether this architecture is scalable to larger networks with greater traffic loads and higher performance requirements is still unproven. Still, some evidence of scalability comes from one of open RAN’s few live deployments, undertaken by the Japanese e-commerce giant Rakuten. Rakuten is on track to deploy 7,000 open RAN sites in Japan by the end of 2021—the equivalent of a medium-sized European country such as Austria or Portugal. But as the number of subscribers on this network is relatively low, the technology’s scalability to support tens of millions of subscribers is still in question.

Sunk costs. A traditional RAN’s total cost of ownership, including the underlying equipment, site rental, support, maintenance, and energy costs, can be the most expensive part of a mobile network, representing 65%–70% of its total cost. Given that, open RAN’s appeal from a cost perspective is easy to understand. Several studies have concluded that open RAN can reduce capex by 40%–50% and operating expenses by 30%–40% relative to a traditional cellular setup. Most of these studies cite Rakuten, which is striving to build the first and largest end-to-end virtualized cloud-native network using open RAN architectures.

The caveat is that while the total cost of ownership claims may be valid in greenfield environments such as in Rakuten’s effort, this magnitude of cost savings seems highly improbable in “brownfield” environments where significant investments have already been made. One of the main reasons for this is that 5G deployments build on and require interoperability with existing 4G infrastructure—and 4G’s closed vendor implementations lock operators into using the same vendor. Accordingly, operators seeking to adopt open RAN in existing infrastructures would need to replace legacy equipment, which would significantly raise an open RAN deployment’s overall cost.

Vendor interoperability concerns. New solutions must compete against proven, tightly integrated legacy RAN systems designed and optimized for high performance. While open RAN provides greater vendor choice and flexibility in implementation, it also increases opportunities for incompatible configurations from multiple possible combinations of software and hardware. Each combination of multivendor end-to-end solutions must undergo extensive testing in a controlled environment, which would require significant additional time, effort, and cost relative to traditional setups.

To explore ways to alleviate this problem, several industry-led initiatives, including the O-RAN Alliance, are hosting “plugfest” events that bring together diverse ecosystems of component vendors to test, validate, and harden end-to-end operator solutions that can also interoperate with existing
legacy architectures. Several leading vendors and consortia are also launching communal labs to test and validate interoperability in a controlled and managed environment.22

**System integration.** Integration complexity also presents a significant obstacle to open RAN adoption, as one of the key benefits of remaining with the traditional model is that operators can turn complete responsibility for implementation, upgrades, and maintenance over to the vendor. Should something go wrong—and it always does—accountability in the current environment is clear. If operators are to evolve away from reliance on integrated turnkey systems from a single vendor, they should incubate new capabilities to orchestrate and manage complex multivendor RAN deployments, which will necessitate the use of in-house, vendor-supplied, or third-party systems integrators. And since open RAN is a relatively new area, there are not many integrators with RAN experience that also have the operator’s best interest in mind.

While anecdotal, Rakuten’s experience again provides grounds for optimism. In assembling its greenfield mobile network, the company undertook a significant role as system integrator to orchestrate at least 10 different vendors, achieving many industry firsts in the process. Rakuten’s CEO originally thought that the RAN integration would be the most complex part of coordinating these vendors. But he later acknowledged that although the RAN integration was extremely challenging, only 10% of the challenge came from the RAN and 90% came from “everything else.”23

**GETTING STARTED WITH OPEN RAN**

Many operators are anxious to move forward with open RAN, and the market is developing rapidly. It may take some time, but many are confident the industry will eventually overcome the technical and engineering challenges that stand in the way of reaching a truly scalable commercial-grade solution. A mid-2020 survey of operators found that a majority believed that practical applications of 5G open RAN BBUs would emerge within two years.24 The same study found that operators are not necessarily waiting for full feature and performance parity. Many said that they would be willing to accept open radio units if they showed 80% of the performance capabilities of a traditional integrated system, particularly for service in underserved areas.25

Even if open RAN is still not quite ready for widescale commercial deployment, this is no time to stand still. Since the telecom industry works on long multiyear planning cycles that can span decades, operators would benefit from taking action today. Now is a good time to assess the current state of their business, understand where the company needs to go, and determine how it needs to change.

Operators should start educating themselves on the opportunities and challenges presented by open RAN. A good place to begin is by separating hype from reality by participating in industry consortia, learning from those already testing open RAN in labs and field trials. Operators can also engage with vendors and other experts to understand global operator deployment trends, assess technology and ecosystem maturity, and evaluate the total cost of ownership of alternate deployments based on their own unique starting point.

While engineers tackle the technological issues, operators can take decisive action in developing an operationalization plan, building an organization and culture of innovation and continuous improvement to support new software-centric business and operating models enabled by open cloud-native architectures. If they want to successfully transition to open RAN, operators should acquire new capabilities, hire and develop internal talent, and adopt new ways of working.
SIGNIFICANT LIVE OPEN RAN IMPLEMENTATIONS AND TRIALS

**Rakuten**, with its 7,000-site deployment planned for launch by the end of 2020, is leading the telecom industry’s transformation toward open RAN architectures. Rakuten, which has a reputation for disruptive innovation, is leveraging the strengths of different vendors for various parts of the network, which has never been done before. The lack of legacy infrastructure helps reduce the risk and cost of deployment.

**Dish Network** in the United States intends to build the first fully open RAN-compliant stand-alone nationwide 5G wireless network. The company is in the process of selecting the vendors that will help build the new, greenfield network. Dish plans to cover 70% of the US population with 5G by June 2023.

**Vodafone** is conducting extensive open RAN trials in the United Kingdom, South Africa, Mozambique, Turkey, Ireland, and the Democratic Republic of Congo with clear timelines for commercial deployments across multiple wireless technologies (2G, 3G, 4G, and 5G). The company is seeking to deploy open RAN technology across its considerable European footprint, which covers 100,000 cell sites and 400 million people across 14 countries.

**Telefónica** announced it will conduct 4G and 5G open RAN technology trials in Germany, Spain, the United Kingdom, and Brazil in 2020. The Spanish operator will collaborate with several vendors to develop and deploy O-RAN trials across its footprint. Telefónica also has several open RAN projects in Peru.

**Deutsche Telekom** is collaborating with two vendors on developing a programmable open RAN platform based on a disaggregated open RAN architecture. The collaboration is part of Deutsche Telekom’s European Open Test and Integration Center to test O-RAN-compliant solutions.

**Etisalat**, a telecom services provider based in the United Arab Emirates, is launching the first Middle Eastern virtual RAN in collaboration with leading RAN technology vendors. Its solution seeks to decouple programmable RAN software elements from the hardware. This would allow generally available hardware and server platforms to process the RAN software, improving deployment flexibility, scalability, agility, and energy efficiency.

THE BOTTOM LINE

The traditional RAN represents one of the last bastions of closed proprietary systems. If history repeats itself, the adoption of open RAN may mimic the time it took the industry to transition to open and virtual core networks—the seven years between 2013, when the tenets underlying core network virtualization were introduced, and 2020, when more than half of the industry’s core wireless shipments migrated from purpose-built to virtual network solutions. The expectation is that more than 80% of core wireless network deployments will be virtualized by 2023. Though open RAN is still in its infancy, the clear growing interest in the technology could be the start of a large and significant trend with the potential to revolutionize the telecom industry.
Endnotes

1. Deloitte analysis of publicly available information (e.g., press releases, company websites, and industry newsletters); analysis may combine multiple deployments in the same country and may not be exhaustive.


6. Open RAN enables the relocation of the BBU from the cell site to a virtualized central (CU) or distributed (DU) data center.


10. Higher integration costs are discussed later in this article.


12. To push much of the traditional function of the core network out to the RAN edge network, Rakuten had to spend considerable time and energy in creating an entirely new automated operations support system (OSS), working closely with software developers to translate and enrich IT workload management techniques to meet carrier-grade requirements.

13. A recent survey on Connectivity and Mobile Trends highlights that consumers care most about network reliability and availability from their provider. See Kevin Westcott et al., Build it and they will embrace it, Deloitte Insights, 2019.


17. The 3rd Generation Partnership Project (3GPP) is a standards organization which develops protocols for mobile telephony.


20. Ibid.

21. Rakuten convinced Nokia to open its optical networking and radio equipment to enable open RAN in exchange for revenue associated with managing and maintaining the Phase 1 build out.

The next-generation radio access network: Open and virtualized RANs are the future of mobile networks

25. Ibid.
Women’s sports gets down to business
On track for rising monetization
Paul Lee, Kevin Westcott, Izzy Wray, and Suhas Raviprakash

DELOITE’S TMT PREDICTIONS reports have historically required a base level of a billion dollars in revenue before an emerging industry would be considered for inclusion. On this basis, the global women’s sports industry (excluding mixed events), measured by the aggregate of TV rights, sponsorship, and matchday (live-event) revenues, is unlikely to qualify in 2021.

TV rights and sponsorship deals for most women’s sports, where they exist, are worth at most millions of dollars, with the majority below this value. In 2021, we predict women’s sports revenues will be well under a billion dollars—a fraction of the global value of all sports (men’s, women’s, and mixed), which in 2018 reached US$471 billion, an increase of 45% over 2011.1

But in spite of that, we are including this topic, as we believe we should, because we predict that women’s sports will grow to be worth a great deal more than a billion dollars in the years ahead. Its ability to generate substantial TV audiences, deliver value to sponsors, and draw tens of thousands of fans per event has been demonstrated on multiple occasions over the past decade. The fan interest is there: A recent multicountry study found that 66% of people were interested in at least one women’s sport, and among sports fans (of whom 49% are female), that figure rises to 84%.2 And the COVID-19 pandemic has catalyzed fundamental reappraisals of many aspects of society, one of which is how women’s sports should be perceived, promoted, and commercialized.

In short, women’s sports is ripe for greater monetization—if certain key elements fall into place. A challenge in 2021 and beyond will be for women’s sports to pull in substantial TV and stadium (as permitted) audiences consistently across multiple sports. Then, the value to sponsors will be self-evident, which in turn should raise marketing spend and awareness. But for this to happen, the entire sports industry—spanning federations, leagues, teams, sponsors, and regulators—should invest on a sustained basis in creating more opportunities for women’s sports to prove its commercial worth.

Women’s sports has been gaining ground despite roadblocks

Women’s sports events have demonstrated their mass-market appeal, and thus monetary potential, on multiple occasions.

Historically, these occasions have been infrequent. In some cases, the ability for women’s sports to thrive on a sustained basis has been artificially constrained. For example, in the case of women’s
football in England, 53,000 people watched Dick Kerr’s Ladies beat St Helen’s Ladies in 1920. In the following year, the national Football Association (FA) banned women from playing on Football League grounds on the basis that “the game of football is quite unsuitable for females and ought not to be encouraged.” This ban was not lifted until 1971.

Nonetheless, women have been steadily gaining ground in a number of sports. For instance, the 1900 Olympics was the first with female participation, featuring 22 women out of 997 athletes; by 2012, both the United States and Canada Olympic teams included more women than men. It was not until 1967 that a woman first ran the Boston Marathon, 70 years after its establishment. Within a dozen years, the first woman to run a marathon in under 2 hours 30 minutes did so at the New York City Marathon.

In sports where men’s and women’s games have relatively equal marketing support, their commercial impact has been roughly equivalent. Tennis, for which the prize money at Grand Slam events is the same for women as men, is arguably the best example. In the United States, TV ratings for tennis grand slams—a major driver of pay TV subscriptions and advertising revenues—have been slightly higher for women than men. Indeed, tennis is the only sport in which female athletes were among the top 100 best-remunerated sports stars in 2019.

Over the last decade, women’s sports has demonstrated time and time again its ability to drive large audiences. This has catalyzed increased interest in bidding for TV rights and sponsorship deals. The value of these deals is modest by comparison to men’s sports, but it is rising. It is worth recalling in this regard that revenues for men’s sports have grown substantially, particularly over the last two decades.

Our expectation is that women’s sports has similar potential for growth, especially as we believe that there is significant untapped interest in watching women’s sports. Realizing this potential should drive rising investment in women’s teams and sponsorship deals, and this in turn should inspire more girls and women to aspire to compete at the highest levels.

Where does the revenue come from?

Elite sports revenues are founded on three main pillars: TV rights, event-day attendance, and sponsorships.

TV RIGHTS: VIEWERSHIP OF WOMEN’S SPORTS IS GROWING
TV rights are the biggest source of revenue for major sports rights holders. Because the value of rights, either for generating advertising or subscriptions revenue, depends on audience size, televised women’s sports would have to attract a substantial number of viewers for it to generate significant revenue. Fortunately, the mass-market appeal of televised women’s sports across a variety of sports is being repeatedly demonstrated, though coverage remains rare relative to that for men’s sports.

To date, women’s football (“soccer” in the United States) has enjoyed the biggest TV audiences. The 2019 FIFA Women’s World Cup in France generated record viewership: A total of 993 million people watched it on TV, with a further 482 million accessing it via digital platforms. The final alone was watched live by 260 million viewers, including 14.3 million in the United States, a country noted on the men’s side for lagging behind most of the rest of the world in terms of soccer’s popularity. Indeed, the 2019 Women’s World Cup final was more popular among US viewers than the men’s final in 2018, with the women’s game drawing 22% more audience.
Throughout the 2019 tournament, too, individual playoff games generated significant audiences in terms of both absolute size and audience share. The average live audience per match, at 17.3 million viewers, was more than double the average 8.4 million viewers per game for the 2015 FIFA Women’s World Cup in Canada. Perhaps surprisingly, the majority (61%) of the viewers were men, illustrating women’s soccer’s broad appeal.

The 2019 FIFA Women’s World Cup USA-England semifinal match reinforces these statistics. In the United Kingdom, 11.7 million people watched the United States beat England in this game, representing just over half of the total TV audience at that time. This number set a viewership record for a female football match and made the game the United Kingdom’s most-watched program in 2019 at that point in the year. For its part, the United States contributed 7.4 million more viewers to the same game despite it being broadcast live on a weekday afternoon in US time zones. Adding in online streaming viewers, the game’s total US audience peaked at 20 million, making it the most-watched women’s soccer game in the United States since the prior Women’s World Cup final in 2015, which attracted 25.4 million viewers (boosted by being shown in the evening slot). To put this into perspective, neither the men’s National Basketball Association (NBA) nor the National Hockey League (NHL) finals in 2015 had that many viewers in the United States market.

One might argue that FIFA World Cups always generate strong viewership, but other women’s football tournaments featuring national teams have also enjoyed rising ratings. The United States’ three games in the 2019 SheBelieves Cup (an invitational round robin between the United States, England, Japan, and Brazil) recorded an average viewership of 439,667 in the United States. This number was nearly 50% higher than the average viewership of the US men’s Major League Soccer (MLS) games in the season’s opening weekend. And even niche audiences for women’s soccer are growing. In the United States, the first National Women’s Soccer League (NWSL) game in 2020 had 572,000 viewers, a record audience for the NWSL by a large margin.

To put this into perspective, neither the men’s National Basketball Association (NBA) nor the National Hockey League (NHL) finals in 2015 had that many viewers in the United States market.

The story is roughly the same for women’s tennis. In fact, in the United States, viewing figures for the women’s US Open tournament have been greater than for the men’s. In 2019, the women’s US Open finals (featuring an American player) attracted an average viewership per game of 3.1 million, considerably more than the 2.8 million viewers who watched the men’s final (without an American), which was the most-watched men’s final since 2015. Across all Grand Slam tournaments in 2018, two of the women’s finals had higher ratings in the United States than the men’s.

Nor are audiences lacking for other women’s sports. From cricket to netball, many women’s teams are showing strong viewership, especially when the national team is playing (see sidebar, “Women’s sports is attracting more and more viewers”).
The caveat here is that women’s sports’ ability to deliver strong TV ratings is not matched by its representation in media coverage. For instance, an analysis of 250,000 news articles in more than 80 languages found that women’s tennis grand slam events received 41% less coverage than the men’s events. This may slow women’s tennis’s momentum, as greater awareness would likely increase viewing yet further. In addition, women’s matches are sometimes not shown on TV, or are relegated to secondary or online channels. The Women’s Twenty20 World Cup Cricket final, played in Melbourne on March 20, 2020, had a stadium audience of 86,174 fans, but it was televised on a secondary channel to avoid clashing with a news bulletin. And when Japan’s team won the football Women’s World Cup in 2011, coverage of the game was obscured by a large number 7, a reminder to viewers of the number of days left to digital switchover.

Yet even this disparity in coverage appears to be changing. The successes of women’s sports events are prompting a growth in coverage, especially on secondary or online channels, albeit from a low base (see sidebar, “TV coverage of women’s sports is growing”).
Besides game coverage, more sports content with female protagonists or subjects is being created. The United Kingdom’s BT Sport and Insight TV have partnered to create a reality TV show, *Ultimate Goal*, that follows the journey of 28 women competing for the chance to play in a one-off match watched by talent scouts from leading women’s football clubs. According to BT, this demonstrates its commitment to “inspire a new generation of girls and women to participate in football, on and off the pitch.” Additionally, Sky is creating original content focused on women’s sports, including interviews with women sports stars. Its lineup includes a new Extraordinary Women series featuring sports stars such as rower Victoria Evans. Other programs include “Training with MMA Star Leah McCourt” and “My Life in Lyon,” a behind-the-scenes look at the life of footballer Lucy Bronze.

With growing audiences and expanding coverage, the market for women’s sports rights is starting to develop. In the last few years, TV rights deals have risen in value and grown in scope. In some cases, they have been negotiated for the very first time, sometimes with little or no money changing hands, with the trade being the guarantee of TV coverage.

Admittedly, the rights values for women’s sports are still low. In the United States, ESPN pays US$25 million for its TV deal with the WNBA. In comparison, the value of the rights for US men’s basketball was US$2.6 billion as of 2019. But the values for women’s sports are on a rising trajectory, with recent years seeing some notable gains (see sidebar, “TV rights for women’s sports are rising in value worldwide”).
EVENT-DAY ATTENDANCE: WOMEN’S SPORTS GAMES ARE FILLING MAJOR STADIUMS

As well as strong TV ratings, women’s sports events are also generating strong event-day attendance levels, with records being broken only to be broken again a few months later.

The draw of women’s sports has been inconsistent in the past. Attendance at women’s matches has sometimes been in the low thousands or even hundreds, especially as some contests have taken place at training grounds with limited seating, potentially deterring fans from attending due to lack of capacity and facilities. For instance, the Mexican women’s football league, the Liga MX Femenil, played some of its first (2017–2018) season’s initial games at training grounds with minimal room for fans. This also meant that the matches could not be televised.44

In recent years, however, audiences for women’s sports in the tens of thousands have become far more common, possibly because games have been played in larger stadiums. To return to the Liga MX Femenil, the league enjoyed rising attendance throughout its first season; the final game, played within a year of the start, drew by 51,211 spectators.45

Other women’s events have drawn more spectators still. The all-time record for a women’s football match was for the Women’s World Cup final in 1999 when the United States played China, with a live audience of 90,185.46

TV RIGHTS FOR WOMEN’S SPORTS ARE RISING IN VALUE WORLDWIDE

• In the United Kingdom, the BBC reportedly paid €10–12 million for the rights to the Women’s Euro football competition in 2021 (now 2022) being hosted in England, up from the €1 million that Channel 4 reportedly paid for the 2017 edition.34

• Also in the United Kingdom, where BT Sport and the BBC have a three-year deal with the Women’s Super League (WSL) football franchise from 2018–2019 to 2020–2021, rights have been awarded on the basis of guaranteed coverage. BT Sport has committed to show 30 live matches per season, while the BBC is showing one live match per week via online or on-demand channels.35

• England’s Football Association (FA) has appointed a company to manage the next round of rights sales from 2021–2022 onwards, with the expectation that coverage will be in exchange for fees.36 The FA has also appointed an agency for international sales of the WSL, and announced a three-year rights deal with Sky Mexico and the Scandinavian broadcaster NENT in September 2019.37

• In France, Canal Plus and TF1 jointly obtained rights for Euro 2021 at a reported deal value of €13million,38 more than double the €5 million paid for the prior tournament in 2017.29

• In Spain, the Women’s Association of Football Clubs (ACFF) announced a three-year deal worth €9 million for the rights to Liga Iberdrola, the first Spanish women’s football division.40 Four of the clubs in ACFF have additional rights to show selected games on their own video platforms.41

• In the United States, the National Women’s Hockey League (NWHL) agreed to a three-year deal with Twitch for its games to be live-streamed.42 The NWHL will be receiving a media rights fee for the first time in its history.

• Rising rights values are enabling record transfer fees, with Danish footballer Pernille Harder joining England’s Chelsea F.C. Women in September 2020 for a record fee.43
National games are among the most popular for matchday attendance. In November 2019, a record 77,768 fans watched England’s women’s football team lose to Germany at London’s Wembley Stadium. This was slightly more than the 77,277 fans who attended the England men’s team match against Montenegro in the same month at the same venue. National women’s cricket teams have enjoyed similarly massive one-off attendances. 86,174 fans watched the T20 World Cup final between Australia and India at the Melbourne Cricket Ground, the largest crowd ever for a women’s sports event in Australia. The Women’s Six Nations rugby tournament has also been popular among fans. A (then) record 10,545 people attended the England-Italy match in March 2019. This record was then surpassed a year later, when 10,974 fans watched England vs. Wales. In Ireland in 2019, 6,047 women’s rugby fans watched the Irish team play the French, a record for a stand-alone Ireland Women’s home game.

Club games also draw significant crowds. In March 2019, 60,739 football fans watched FC Barcelona Femení beat Atletico Madrid at the Wanda Metropolitan stadium in Madrid, beating the prior attendance record for a women’s fixture of 48,121 at the same stadium. In November 2019, 38,262 football fans attended the women’s “Super Derby” at Tottenham Hotspur between London rivals Spurs and Arsenal. Anfield, Liverpool FC’s stadium, drew 23,500 fans to its local derby against Everton, the first-ever WSL match it had hosted. The United States’ WNBA recorded attendance of 1.33 million for the 2019 season, with the Los Angeles Sparks having the highest home-game attendance—a total of 192,224 across all home games with an average of 11,307 fans per game.

SPONSORSHIP: IN ITS INFANCY, BUT ON A STRONG TRAJECTORY
The global value of sports sponsorship has been sized at US$44.9 billion per year, of which women’s sports is a fraction of this. But strong TV and matchday audiences are encouraging more sponsors to consider women’s sports. The nascency of the market may make women’s sports better value for money than the men’s equivalents. Other appealing factors are likely the relative ease of concluding a deal as well a steeper potential growth trajectory.

As sponsorship interest grows, rights for women’s teams are increasingly being sold individually rather than being bundled with the men’s team. Indeed, by the time of the next FIFA Women’s World Cup in 2023, we expect all women’s teams to have at least one sponsorship agreement distinct from the men’s teams.

Several recent agreements exemplify the growing role of sponsorships in women’s sports. In 2018, Visa signed a seven-year deal to become the first-ever sponsor of Union of European Football Associations (UEFA) women’s football, becoming the main partner of flagship events such as the UEFA Women’s Champions League and the UEFA Women’s European Championship. Also in 2018, financial advisory company Vitality was named sponsor of the 2019 Netball World Cup as part of an three-year extension to an existing sponsorship deal. In July 2018, Stanley Black & Decker became the Catalan football team FC Barcelona Femení’s first shirt sponsor. And more sponsorships have been announced in the two years since (see sidebar, “A sampling of sponsorships”).
A SAMPLING OF SPONSORSHIPS

2019 and 2020 have seen the announcement of a number of major (particularly relative to previous women's sports deals) sponsorships. In 2019, these included:

• Barclays agreed to become the title sponsor of the FA WSL. The multi-million-pound sponsorship deal was the largest-ever investment in UK women's sports by a brand.59

• Budweiser became an official partner to the England Women's football team in 2019, claiming the title of the team's official beer.60 The company had already been a sponsor of the men's team.

• Boots signed a three-year deal to sponsor the women's national football teams of England, Wales, Scotland, Northern Ireland, and Ireland.61

• Westfield extended its sponsorship of the Matildas, the Australian women's national football team, and the W-League, the premier league for women's football in Australia, by two additional years.62

• Iberdrola, a Spanish energy company, extended its sponsorship of Spain's women's football league for a further six years.63

Still more major sponsorships have been announced in 2020. In the United States, the WNBA announced Changemaker sponsors, each chosen for their commitment to “driving positive change for the WNBA, women's sports, and women in society.”64 These inaugural Changemakers included AT&T, official marquee partner, and Deloitte, official professional services provider.65 The sponsorships entail marketing amplification and strategic collaboration as well as financial contribution.66

In Europe, PepsiCo signed a five-year deal to sponsor UEFA Women's Football. The deal will run alongside PepsiCo's sponsorship of the UEFA Champions League.67 PepsiCo will be a main partner of the UEFA Women's Champions League, the UEFA Women's EURO, the UEFA European Women's Under-19 and Under-17 Championships, and the UEFA Women's Futsal EURO.68

We also expect a growing number of women’s teams to have multiple sponsors. For example, in January 2020, FC Barcelona Femení added a second sponsor with the announcement of the club’s first official street clothing partner, Naulover.69

Sponsors are more likely to commit to spending equal amounts on women’s teams as men’s teams if they are sponsoring both. Adidas, which sponsored six women’s teams at the Women's FIFA World Cup in 2019, announced it would offer equal performance bonuses to men’s and women’s teams.70

Investment is increasing as women’s sports gains momentum

Already, we are seeing increased investment in women’s teams around the world, often via acquisitions. Valuations are still a fraction of those of men’s teams, but the low sums involved may make investments in women’s teams more attractive. In 2020, France’s Olympique Lyonnais Groupe paid US$3.15 million to acquire an 89.5% stake in NWSL’s Seattle-based team.71 Real Madrid launched its women’s team in July 2019 by acquiring the existing Madrid-based team CD Tacon for €500,000. The team, which had been promoted to La Liga Iberdrola, the country’s top women’s league, in May 2020,72 will train and play at Real Madrid City.
CONDITIONS FOR ELITE WOMEN ATHLETES ARE STEADILY IMPROVING

The growing funding of women’s sports has been mirrored by improvements in conditions for female athletes. For example, in January 2020, the WNBA agreed to new terms that included a 53% pay raise. In 2019, the average WNBA player earned US$116,000 per season; the new agreement raised the base salary to US$130,000 and created additional bonus and prize pools. Under the new structure, top players could earn more than US$500,000 per year. Additional benefits under the new terms include paid maternity leave, dedicated areas for nursing mothers, fertility and adoption services, and improved travel conditions (flying premium economy rather than economy).

Similar changes are taking place in Europe. In January 2019, 28 full-time contracts were introduced for England’s women’s rugby team. This move was intended to help accelerate the team’s development ahead of the 2021 Rugby World Cup in New Zealand by enabling players to focus on rugby full time. The French women’s national football Under-16 to Under-19 teams have been based at the National Football Centre—part of INSEP, a national team training center for elite athletes in Clairefontaine-en-Yvelines—since 2014. The facility provides similar conditions for women and men’s teams (parity is relative progress), as well as athletic and school education.

Acquisitions are not the only source of money flowing into women’s teams. Manchester United in the United Kingdom reportedly invested £5 million to re-form its women’s football team in 2018 after a 13-year absence. The team subsequently ranked as the fourth-best team in England after just two seasons.

Prize money in women’s sports is also going up, increasing the flow of cash into national teams. The 2019 FIFA Women’s World Cup awarded a total of US$30 million in prizes, with US$4 million going to the winning team; the 2023 tournament is expected to award more than twice as much, with over US$60 million on offer. While these amounts are small relative to the winnings garnered by men’s teams, growth in women’s teams’ awards are outpacing those of their male counterparts.

While these amounts are small relative to the winnings garnered by men’s teams, growth in women’s teams’ awards are outpacing those of their male counterparts.

Investments are happening at the youth level as well, where football programs for boys have historically abounded but been lacking in most markets for girls. For example, Real Madrid’s women’s squad will include youth teams from the Under-16 to Under-19 age groups. At the federation level, FIFA is supporting doubling the number of FIFA members’ associations that organize youth leagues by 2026. FIFA is also encouraging the playing of football in schools by girls, which should create a pipeline of future women players. Its goal is to increase the number of women playing football to 60 million by 2026.
COVID-19 AND WOMEN’S SPORTS

Women’s sports has massive potential if its momentum can be maintained. However, COVID-19 has introduced some headwinds. The pandemic put the brakes on most sporting activities in 2020, and it was men’s sports and its top leagues that resumed activity first.

The resumption of elite sports will ideally embrace women’s as well as men’s sports. One encouraging sign was that the UEFA Champions League’s resumption for the 2019–2020 season included the women’s as well as the men’s competition, with both tournaments concluding in August 2020. The creation of protective bubbles, which were initially used to enable elite men’s teams to resume competition (though in empty stadiums), are also now being provided for women’s events to enable them to resume.

In schools and at youth levels, it took a long time before team sports could resume, constraining boys’ and girls’ teams’ ability to practice. This period could still be used, however, to improve skills in areas such as fitness and flexibility.

THE BOTTOM LINE

The most important lesson from all this is that women’s sports has immense potential value, not just in monetary terms, but also in terms of what it signals for gender parity. For women’s sports to fulfill its potential, however, requires action by all interested parties:

• **Broadcasters should continue to invest in women’s sports.** Women’s events may be particularly attractive to public service broadcasters in markets where they are no longer able to compete successfully for rights to men’s elite sports. There is a vast difference in the cost: Men’s events in major countries may cost over a billion dollars per season, while women’s events are still often willing to trade coverage for rights.

• **Women’s teams should slipstream men’s teams, but also keep their distance.** Creating a women’s team within an existing successful franchise best known for its men’s team provides instant brand recognition and can also provide immediate access to facilities. Women’s teams can also replicate existing rivalries, offering up local derbies (such as Atletico Madrid versus Real Madrid or Liverpool versus Everton), historical rivalries (such as the Los Angeles Lakers versus Boston Celtics), regional tournaments (such as the Ryder Cup), or national rivalries, which are innumerable. But women’s teams should also make sure that they stand apart when it comes to negotiating sponsorships and TV rights so that that women’s sports becomes valued in its own right, and becomes more investable as a result. Sports franchises should hire specialists to sell sponsorship rights—they will have the experience and savvy to maximize value. Further, women’s teams should capitalize on the additional flexibility they have on building digital presence and relationships with fans. They are in a position to negotiate contracts that are not as restrictive as for many of the men’s teams. Women’s teams can build a 1:1 relationship with fans via social media and streaming platforms that may be harder for men’s teams, whose primary fan interaction is via broadcast TV.

• **Video content creators should consider the value of female athletes’ stories.** There are many epic stories to be told of women who triumph despite adversity. The first woman to run the Boston Marathon, the winning Japanese and US Women’s World Cup teams from 2015 and 2019, and Lindsey Vonn’s four skiing World Cup championships are just a few of many fascinating stories, pivoting on a foundation of relentless challenge and striving for excellence.
• **Significant women’s events should take place in the largest stadiums available.** Women’s sports has proven its ability to draw tens of thousands of attendees if there is sufficient awareness and if there is ample tension about the outcome, such as major trophies being at stake. (In 2021, of course, this recommendation applies only in markets where matchday gathering is possible.)

• **Sponsors should capitalize on their amazing opportunity in women’s sports.** Not only are women’s sports sponsorships a great value relative to those for men’s events, but they are likely to appeal to female customers and staff members as a signal of commitment to positive changes in society. Women have significant buying power, and sponsorships can help develop their fondness for brands. Sponsors should also consider that both men and women watch women’s sports (as well as the converse); thus, it makes as much sense for male-oriented brands to sponsor women’s events as for more female-focused brands. Women’s teams should also consider that they have the ability to define how they would like their relationship with sponsors to work. Men’s teams have already been in long-term contracts with partners that lock them into specific category definitions, and possibly outmoded notions of how the partnership should work.

• **Sports apparel vendors can explore greater involvement in women’s sports.** Sponsoring women’s sports may be of particular interest to sports apparel vendors, which earn revenues of about US$80 billion per year, of which women’s apparel accounts for about US$27 billion. As one retail leader characterized this industry’s historical approach: “I think for a long time, athletic brands said, ‘We can just shrink it and pink it and that will be good enough for the female consumers.’” Interestingly, considering its smaller market, women’s sports apparel has shown considerable innovation in recent years with the growth of athleisure, which has been designed predominantly for women.

• **Sports federations at both the global and national level should set targets for female representation on boards,** just as has been the case in the general business world. For example, FIFA’s goal is for all its member associations to have at least one woman on their executive committee by 2026, and for one-third of FIFA committee members to be women by 2022. Change takes time, and it may take a decade, or even a generation, for women’s sports to attain its full potential. But its promise of delivering value to sponsors, investors, fans, and athletes and teams themselves is becoming more and more clear. We look forward to a world in which women’s sports has a fully equal status with men’s, in all respects.
Endnotes

1. Mixed events, such as the Olympic Games and Grand Slam tennis, are not included in the revenue calculation.
2. Stefan Hall, “This is how COVID-19 is affecting the world of sports,” World Economic Forum, April 9, 2020.
9. FIFA, “FIFA Women's World Cup 2019™ watched by more than 1 billion,” October 18, 2019.
12. Ibid.
13. Hess, “US viewership of the 2019 Women's World Cup final was 22% higher than the 2018 men's final.”
15. CBS Sports PR, “Challenge Cup opener on CBS is most-watched @NWSL match ever,” Twitter, June 30, 2020.
18. Ibid.
19. RUCK, “England are a TV ratings hit as record number watch Women's Rugby World Cup final,” August 27, 2017.
Women's sports gets down to business: On track for rising monetization

36. Ibid.
43. Ben Morse, “Pernille Harder joins Chelsea for 'record transfer fee for the women's game','” CNN, September 2, 2020.
44. Salvador Perez, “Chivas TV will broadcast the women's team; will play in Verde Valle,” ESPN, August 7, 2017.
47. ITV, “England women set attendance record of 77,768 at Wembley during 2-1 defeat to Germany,” November 9, 2019.
52. Baber, “Attendance record broken in Women's Six Nations.”
64. WNBA, “WNBA announces progressive WNBA Changemakers partnership platform to elevate the WNBA and women's sports,” January 14, 2020.
65. Ibid.
66. Ibid.
67. UEFA, “PepsiCo signs 5-year deal to sponsor UEFA Women's football,” August 26, 2020.
68. Ibid.
75. Young, “WNBA agrees to 53% pay raise, maternity benefits for players in new collective bargaining agreement.”
78. INSEP stands for Institut national du sport, de l'expertise et de la performance.


83. Ibid.


86. FIFA, *Women’s football strategy*. 
FROM CRICKET TO hockey, baseball to basketball, the digital transformation of sports is in full swing. Clubs, teams, leagues, broadcasters, venue operators, and athletes increasingly see the value in analytics and are working to realize that value. Technologies such as computer vision, machine learning, advanced wireless connectivity, and wearable sensors are transforming how athletes train, compete, and manage their careers. This explosion of data, however, is raising new questions about how best to use it—and how to do so ethically. To begin to address such concerns, we predict that by the end of 2021, multiple professional sports leagues will establish new formal policies around the collection, use, and commercialization of player data.

Data is fundamentally changing sports

If measuring something in sports is conceivable, chances are that someone, somewhere is already measuring it. Hundreds of different metrics can be analyzed today through video analytics and wearables such as harnesses, sleeves, bands, straps, rings, and smart fabrics. Over the past decade, the use of analytics in sports has slowly transformed everything from how talent is identified and assessed to how athletes are trained and managed to how games are played on the field, court, and pitch. The data and analytics revolution has begun to blur the boundaries between many disparate areas of sports, including esports, virtual sports, gaming, broadcasting, fantasy sports, betting, and the live-venue experience. As the use of data and analytics in sports matures, the industry will likely have to address issues about not just the enabling technology, but also increasingly about data rights management, privacy, regulations, monetization, and new ways to experience sports.

Almost all of the world’s major professional sports teams have one or more analytics experts on staff with the mandate to find any advantage that can improve the team’s chances of winning. In basketball, video capture and analysis have yielded insights that have led to more three-point shots and a greater emphasis on player-load management. In baseball, intensive statistical analysis of what works and what doesn’t has dramatically changed pitcher management, increased the use of the “shift,” and replaced small ball with swinging for the fences. Similar analyses
now inform how American football teams approach fourth downs and the way soccer teams select shots.

The difference today is that data collection and analysis are becoming increasingly real-time, and it is happening not just on the field and in the gym but around the clock. Further, it is now possible to measure indicators inside as well as outside the body; new layers of positional, biometric, and biomechanical data are creating hundreds of new metrics to feed into decision-making. Finally, advancements in computing power, cloud technology, machine learning algorithms, and high-speed video capabilities are enabling ever more powerful ways to collect and crunch the numbers.

In this era of the hyperquantified athlete, the increasingly urgent question is how to get from data collection (which is easy) to actionable insight (which is hard) to potential monetization (which is really hard)—all the while protecting athletes’ rights, ensuring fair play and competitiveness, and meeting the financial needs of leagues, players, and owners.

PERFORMING IN A PANDEMIC

Sports leagues and teams continue to struggle to define their new normal during the COVID-19 pandemic. Leagues of all kinds have modified when, where, and how they play in their efforts to keep their players, coaches, and staff safe and healthy. Some teams are using bubbles (single-site tournaments) to continue play with no live fans, or relying on limited travel and strict behavioral protocols. But many are also leveraging new technologies to provide an extra level of defense.

In the United States, both the National Basketball Association (NBA) and Women’s National Basketball Association (WNBA) have offered their players the use of Oura rings to wear during their tournaments in Florida to help monitor their body temperature and respiratory and heart rates as a way to determine health risk. A similar technology has been used with promising results by golf’s PGA Tour. One player, using a Whoop fitness tracker strap, noticed significant changes in his respiratory rate while he was sleeping, prompting him to get a COVID-19 test despite being symptom-free. He tested positive and removed himself from competition, potentially protecting other players. Whoop straps have now been made widely available to the Tour’s players. The German Football Bundesliga as well as the United States’ National Football League (NFL) and NBA are using devices to monitor social distancing and enable contact tracing.

These technologies existed prior to the pandemic, but COVID-19 has accelerated their use, giving them the opportunity to prove themselves on the public stage. Athletes are becoming more comfortable with tracking technology as they come to appreciate having more insight into and control over their health and performance. Teams and leagues, meanwhile, have welcomed having additional data to inform health and safety decisions.

Going forward, the COVID-19-driven influx of monitoring technology into sports poses some thought-provoking questions. Will athletes let their organizations increasingly track their health and wellness data in addition to their performance data? Will they accept being continuously monitored while they sleep and relax as well as on the field? After this current pandemic, could monitoring help with outbreaks of other communicable diseases in the locker room? These and similar issues point to the growing need for dialogue and governance over professional sports’ practices around collecting and using athlete data.
The state of play

For purposes of this article, we will focus on two main types of data that are typically collected from athletes:

- **Positional/tracking data.** Positional/tracking indicators measure, in three dimensions, exactly where a player—or ball, puck, or other object—is located on a field or court. This data can include metrics such as position, acceleration, lateral motion, speed, jump height, and other measures. The data is collected either through video analytics or by sensors in combination with global satellite positioning systems and ground-based wireless networks.

- **Biometric data.** Biometric data refers to any kind of biological information from an individual player. These metrics could include everything from pulse rate and blood glucose or oxygen levels to sweat rate and sleep rhythms. Some biometric measurements, such as heart rate, have been used for decades; now, through digital sensors and ubiquitous low-latency communication networks, many more measurements can be made, in more physical locations, at a greater speed.

Note: This list is not exhaustive.

Source: Deloitte analysis.
The market for quantifying athletes is both diverse and fragmented, with a broad range of options for sensor technology, computing power, data storage, and advanced analytics. Purveyors of these technologies range from “sport tech” companies such as Catapult, KINEXON, Stats Perform, and Zebra to tech giants such as Amazon Web Services, IBM, and SAP, as well as a plethora of startups. The broader sports tech market has also seen a great deal of venture capital interest in recent years, with over 3,000 global deals and funding rounds between 2014 and late 2019.⁵

**Purveyors of these technologies range from “sport tech” companies such as Catapult, KINEXON, Stats Perform, and Zebra to tech giants such as Amazon Web Services, IBM, and SAP, as well as a plethora of startups.**

Most, if not all, of the world’s major sports leagues are using these technologies and techniques in various ways. In doing so, they are expanding their view beyond just the technology to how they balance the technology’s use with players’ and coaches’ needs. For example:

- Beginning in 2014, the NFL began using radiofrequency identification (RFID) tags in players’ shoulder pads, balls, and various areas of the field.⁶ Each team is provided their own raw data from that system to analyze and use as they see fit.

- The National Hockey League (NHL) has been experimenting with puck and player tracking for years. After demonstrating the latest iteration of these technologies at its 2020 all-star game, the league is beginning to incorporate them fully.⁷

- Players in the Australian Football League have been using positional sensors and heart rate monitors during games for the last few years.⁸ They even use the data to make in-game adjustments.

- In 2017, Major League Baseball (MLB) approved the use of Whoop straps by players on a voluntary basis to track information that they can use later to better understand their performance.⁹ This is in addition to other approved wearables that allow players to measure factors such as arm stress.¹⁰
POWERING PERFORMANCE AT THE UNITED STATES OLYMPIC & PARALYMPIC COMMITTEE

Scott Riewald is the senior director for high-performance projects at the United States Olympic & Paralympic Committee (USOPC). His job is to build capabilities for and to improve the collaboration between the technology, innovation, and data and analytics teams throughout the USOPC. He has been strengthening connections between these different groups and, in his words, is now “pulling those threads to bring everybody closer together to create a unified approach of how we use data to support Olympic and Paralympic athletes.”

Five years ago, the USOPC realized that it was not where they wanted and needed to be with regard to data and analytics, and made the strategic decision to focus on the opportunity to expand these capabilities. The USOPC works with thousands of elite athletes in scores of different sports, giving the organization access to expansive data sets that create opportunities to discover “hidden insights.” Riewald and the USOPC also work with the various Olympic and Paralympic sports’ National Governing Bodies to establish best practices, provide more consistency around data and analytics, and recommend which technologies to use.

The USOPC utilizes its Athlete 360 performance management platform to collect and analyze athlete data. This system was designed to be device-agnostic so that data from different sports from different sources and different pieces of technology (such as wearable sensors) can be easily aggregated. In the next two to three years, Riewald and the USOPC aim to be able to capture more data in a less invasive manner. As an example, Riewald points to the skin-mounted electronics pioneered by the Rogers Research Group at Northwestern’s McCormick School of Engineering. These devices, when placed on athletes, can noninvasively measure performance-impacting indicators such as sweat rate, electrolyte loss, and blood glucose levels.

As its capabilities mature, the USOPC is looking beyond simple data collection and integration in the quest to uncover meaningful and impactful insights for all athletes. Riewald wants to use data to better understand causality—what performance measures preceded a particular result—to better define the “pathways to success.” His team works hard to make athletes and coaches understand that data collection and analysis are important. They want to make the value proposition clear, convey information so it drives curiosity, and, ultimately, help drive positive behavioral changes. Riewald emphasizes: “Thinking about ways in which you can convey information in order to achieve behavioral change is what underpins what you're trying to get from data-gathering.”

Riewald is acutely sensitive to these developments’ implications for privacy, security, and ethics. Every day, more and more technology becomes available to monitor more and more aspects of an athlete’s life. Trainers and coaches can potentially track everything about an athlete every minute of the day. To balance the benefits against the risks, Riewald says, organizations need a clear reason to monitor athletes and must prove its value to the person being monitored. He warns: “Just because you can measure something doesn't mean that you should. I really believe that we walk a fine line here. We want to provide relevant data that positively impacts athlete health and performance, without doing so much that we risk athletes and coaching going into data overload.”
A competitive edge though hyperquantification

The famous cycling coach Sir Dave Brailsford once said, “It struck me that we should think small, not big, and adopt a philosophy of continuous improvement through the aggregation of marginal gains.” Hyperquantification is helping teams, leagues, and players discover these gains across the life cycle of their sport—talent identification, training, pregame preparation, game play, postgame analysis, and rest and recovery. Of these, three areas that could see further innovation in the near future are:

**Talent identification.** If a team is going to make a significant financial investment in a player, it wants to be confident that the player has the skills, raw physical ability, medical profile, and mental makeup to succeed. This process is becoming more virtual, especially in the wake of COVID-19. In addition to traditional in-person assessments, teams are increasingly using automated video analysis and positional and tracking data in their scouting. These advances have helped to expand the pool of potential professional players to include those that scouts may not be able to travel to see. For example, one pro football hopeful could not attend traditional scouting events due to personal injury and the COVID-19 pandemic, so he submitted his player-tracking data from college games to prove his speed in a real-world environment. In the near future, more biometric data will likely be added to the mix to augment medical data for predicting long-term performance. One day, teams could have access to a player’s biometric data from his or her participation in youth sports through high school and college programs. While data is unlikely to completely replace traditional evaluative methods, it is poised to provide another set of objective measures that will likely become more and more important.

**In-game decision-making.** Today, most biometric data collection and analysis happen during training and practice, or to help with postgame evaluations. Its use is still not widespread in actual games, and even if in-game data is collected, it typically is not used for any immediate coaching decisions. In fact, some leagues prohibit the in-game collection and use of biometrics unless explicitly approved by the league. On the other hand, in-game positional and tracking data is already widely used. Leagues such as the NFL, NHL, and the Mexican football league Liga MX have all done this to some extent, and they have gained better insight into how game strategies are working. As teams gain more experience and data collection becomes easier and analysis faster, real-time analytics that guide how coaches and managers direct their teams may come into wider play, identifying who might be at risk of injury, who is getting close to exhaustion, who might be out of position, and who may be primed to make a big play.

**Injury reduction.** The holy grail for many athletes and teams is being able to predict when conditions may heighten the risk of injury. For teams, it means more wins and more revenue; for athletes, it means having information that can help them extend their careers—and earnings potential—as much as possible. Predicting injuries more effectively requires measures that help balance exertion and strain with the proper amount of recovery time and sleep. The potential benefits are vast: One study estimated that the NFL lost more than US$500 million in 2019 due to injuries. It’s no wonder that the NFL recently partnered with Amazon Web Services to use machine learning and computer vision technologies, powered by multiple different data sets, to gain insight into head and other injuries.

From measuring to monetizing

The hyperquantification of athletes can provide more efficient training, improve competitiveness, and increase the likelihood that star athletes stay
healthy—benefits that, as organizations well know, can potentially lead to higher attendance, more sponsorships, and greater broadcast viewership.

What may be significantly harder is to figure out how this digital revolution can create new revenue streams for teams, leagues, and players themselves. Although innovations are taking place in smaller sports, many major leagues are just starting to test how they can monetize players’ biometric and positional data, with some leagues even banning the commercialization of player data. There is still much to be done to establish clear data-ownership policies, provide equitable revenue distribution, and ensure player protections. But the potential for monetizing player performance data is so great that it will likely encourage the involved parties to begin addressing these concerns in the short to medium term.

We see several areas for experimentation and innovation over the next few years:

**Fan engagement.** Leagues and players’ associations can work with broadcasters and venue operators to use player data to improve the fan experience and create new ways for fans to engage with sports and athletes. This could lead to new direct-revenue streams, generate new marketing and sponsorship opportunities, and/or have a halo effect on already established revenue streams. For example, the Professional Squash Association (PSA), working with Sports Data Labs, tracks and displays players’ heart rates in real time during broadcasts. The goal is to give fans a better understanding of the game’s athleticism as well as create more excitement around the gameplay; any revenue generated through the biometric data is shared among the league, players, and partners. Similarly, the handball league LIQUI MOLY Handball-Bundesliga has teamed up with KINEXON and Content Stadium to stream real-time player performance information in stadiums and across social media platforms.

**Licensing.** Another way for players and leagues to make money is by licensing player data to organizations such as fantasy sports leagues, sportsbook companies, broadcasters, and health and fitness companies. In 2017, the NFL Players Association came to an agreement with Whoop that, among other things, allowed players to sell their personal health data collected by the wearable. Many leagues already have strategic partnerships with sportsbooks, such as the PGA Tour with DraftKings and the NBA with William Hill, laying the groundwork for emerging opportunities. Other types of opportunities also exist: During its short life, the Alliance of American Football, in partnership with MGM, was developing an application that would enable betting during games where odds would be adjusted based on data from player wearables.

One league that sees such opportunities clearly is the WNBA. In its recently enacted collective bargaining agreement with players, they recognized the potential for data collection and use, saying that “the use of wearables and other in-game technology provides a unique opportunity for the WNBA with respect to enhanced broadcasts, differentiated fan experiences, player health, and revenue generation.”

**Key matchups**

Any opportunity to improve athlete performance or to increase the amount of money in sports is clearly extremely appealing. However, no emerging technology or change in approach is without consequences, some expected, some not. We see three critical areas that should be addressed if hyperquantification is to come into its own:

**Quantity vs. quality.** Coaches and their staffs have traditionally relied on gut feel and experience. Even today, some use analytics more widely than others, striking their own balance between data and instinct. With new positional and biometric
data added to the mix, the analytics becomes more complex and the insights harder to communicate. To heighten the data’s utility, much work should be done to determine what measures are most critical—what data will truly make a difference, and what will merely create noise in the system. There is also debate over whether to believe the data over seeing results with one’s own eyes. Are there ways to properly quantify the magic of once-in-a-generation players? Finally, some worry that overanalysis will lead to a less entertaining product for fans, essentially “ruining” sports. For example, in MLB, longer games and less action (more strikeouts and home runs) can be attributed to the extensive use of analytics.  

Performance vs. privacy. Questions abound about who ultimately owns player data, including any sensitive health information that teams could collect. A number of professional sports leagues in the United States have established guidance and rules around this issue in their collective bargaining agreements with players’ associations. These rules specify how player data can be collected, whether or not its collection and use is voluntary, how it will be protected, and how it can be used. Some leagues have also established sensor and wearable committees that approve device and data use. For example, the NFL states that “each individual player owns his personal data collected by sensors,” while the NBA’s collective bargaining agreement says that “a player will have full access to all data collected on him from approved wearables.” But although these agreements afford some protections and benefits, the speed of technology development and the rapid evolution of potential commercial opportunities are not necessarily compatible with the long-term nature of collective bargaining agreements, which last years.

A focal point in this debate is what is best for players. Many athletes are uneasy about using wearables to collect their data, especially biometric data. Some feel like they are part of a laboratory experiment, while others worry that, even though such use is prohibited, biometric data collection could bias contract negotiations. Some also voice concerns that required data collection will increasingly encroach on activities outside of work, such as sleeping. An example of this conflict is “Project Red Card”: A group of hundreds of former and current soccer players in the United Kingdom are suing sportsbooks and data-processing companies, claiming that they have illegally profited from player statistics. The details of this lawsuit are still emerging, but the results could eventually shape how many types of player data can be commercialized, as well as who ultimately profits.

Betting vs. banning. By some estimates, the global sports betting market is predicted to top US$150 billion by 2024. With the 2018 repeal of the Professional and Amateur Sports Protection Act in the United States, and the legalization of sports gambling by many US states, American leagues in particular are hungry to tap into new revenue sources. One of the emerging issues in this space is the potential for using individual athlete data for bets, either through using biometric and positional data to set odds, or by betting on specific measures such as pulse, top speed, and acceleration. Such uses could generate substantial pushback from players, who may see it as a gross invasion of privacy—but it could be a different story if they could significantly profit from it. The state of Illinois has gotten ahead of the curve by banning the use of biometrics in sports betting unless the relevant players’ association gives permission. Even though the desire for this type of wagering exists, the devil is in the details, and players, leagues, unions, legislators, and commercial businesses should all come to the table to decide what is best.
THE BOTTOM LINE
In the near future, it may become commonplace for fans sitting in a stadium or arena to look up at the scoreboard and see players’ top speeds in real time, or a ranked list of players’ level of physical exertion during the game. The same information could also be broadcast to viewers at home and streamed on social media. Fans, both at home and in person, could use their mobile devices to place a bet in real or virtual currency on the outcome of the next play, with the odds influenced by positional and biometric data. Ideally, this would all be backstopped by robust agreements among all involved that guarantee voluntary collection of data, protect players’ rights, provide for proper data security and privacy, and set out detailed licensing and compensation rules.

These issues should be approached with great care, as the emerging hyperquantification landscape has the potential to become much more complex in the near future. As Yogi Berra once famously said, “If you don’t know where you are going, you might wind up someplace else.” Critical open questions include:

• How will advances in computer vision such as automated video analysis, sensor technologies, and machine learning accelerate the state of the art?
• Would players benefit enough financially from sharing their personal and performance data to entice them to do so?
• Will a standard agreement emerge between players and leagues across sports around the collection, use, and monetization of private and sensitive information?
• How will performance data ultimately be used to enhance the in-stadium and remote fan experience?

For the hyperquantification of sports to succeed, the athlete should be at the center of every decision and conversation. Athletes, trainers, coaches, player agents and representatives, and business leaders should become knowledgeable about the enabling technologies and their responsible use. Those collecting and using the data should convincingly demonstrate and effectively communicate its value. Athletes should see that it is in their best interest to share their data and allow for its thoughtful application. Above all, it is critical that trust be built and maintained among all parties. Without it, the potential gains may never be fully realized.
Endnotes


12. Ibid.


14. Riewald interview.

15. Ibid.


18. Joe Lemire, “Wimu Pro wearable grows to all levels of elite Mexican soccer,” SportTechie, June 6, 2019.


23. KINEXON, “KINEXON and Content Stadium transform real-time data to real-time content,” October 14, 2019.
TV’s New Year’s resolution

The start of the 8K wave

Paul Lee, Kevin Westcott, Cornelia Calugar-Pop, and Anil Kumar Tarigoppula

If you have a 4K-resolution TV screen—and many consumers do these days—you’re used to seeing your favorite shows, movies, and videos with few visible individual pixels. But over the next few years, those images may be about to get even sharper. We predict that 8K—an upgrade and complement to 4K resolution—will generate US$3.3 billion in global revenue in 2021, with this amount rising steadily over subsequent years (figure 1).

These revenues will come predominantly from sales of 8K TV sets to consumers (an anticipated 1 million units with an average selling price [ASP] of US$3,300), with the standard becoming increasingly popular for the largest television set sizes. In addition, sales of equipment (such as cameras, monitors, storage, and computers) related to the creation and production of 8K content should generate hundreds of millions of dollars globally for the year.

FIGURE 1
8K sales will likely surge over the next two years
Global revenues from 8K panels, US$ billion


WHAT IS 8K?
The 8K standard is named after its resolution, which is 7,680 horizontal x 4,320 vertical lines on a 16:9 aspect ratio screen, equivalent to about 33 million individual pixels. This compares to 8.3 million pixels for 4K. The 8K standard also includes specifications for:

- Frame refresh rate, which varies from 24 to 120 frames per second
- Color range, which spans every color available in nature as well as additional artificial colors, as well as improved dynamic range related to brightness
- Audio, which is upgraded to 22.2 channels of surround sound
Why some people are skeptical about 8K’s mass-market adoption

At first glance, consumer adoption of 8K seems to face three principal hurdles:

- **Content.** Content is what makes devices useful; without it, a device may be mostly redundant. However, we expect that less than 0.1% of all video content created in 2021 will be in 8K, and that 8K will be an even smaller proportion of the total base of video content (TV programs, movies, user-generated content, and video games).

- **Cost.** 8K TV sets are likely to cost up to tens of thousands of dollars for premium models. The starting price of 8K TVs, at around US$1,500, is likely to be far higher than for 4K TVs, which are available for under US$300.

- **Comparison.** In blind tests, consumers have struggled to distinguish 8K video from 4K.

These hurdles may suggest that 8K is unlikely to flourish in 2021. However, our view is that these challenges can be surmounted. There is already abundant content that can be viewed in 8K resolution, and that quantity will likely grow over the coming years. The cost of entry-level 8K TVs should decline to about US$1,500 by the end of 2021. And the most common comparison of an 8K TV to a 4K one is likely to be based on physical size and peak pixel count, rather than the screen definition used in day-to-day viewing.

Surmounting the content hurdle

With 8K representing less than 0.1% of all video content created in 2021, one might think that viewers would have trouble finding 8K content. But according to our estimates, enough 8K content already exists for an owner of an 8K TV to spend every hour of every day throughout 2021 watching 8K material—without watching a single repeat.

Further, the lack of native 8K content is, in itself, not a major constraint to the purchase of 8K TV sets. Back in 2012, the lack of native 4K content when the standard was launched was expected to deter purchases of 4K TVs. But the relative paucity of 4K content did not faze many consumers. The majority of TV sets sold since Q4 2018 have been 4K despite only a minority of broadcast, satellite, video-on-demand, and disk content being available in native or remastered 4K. If history repeats itself, we would expect that a large proportion of consumers would buy an 8K panel if it were close in price to a 4K set of the same size. And these consumers would be content with their purchase even if they watched 8K content only occasionally.

If history repeats itself, we would expect that a large proportion of consumers would buy an 8K panel if it were close in price to a 4K set of the same size.

It is also important to note that consumers who buy an 8K TV set in 2021 may be buying it for the long term, expecting to keep it for five to seven years (the expected life of a TV set before it is replaced) or more. This allows ample time for more 8K content to become available.

But where will 8K content come from in the near future? We expect that in 2021, at least six sources of content will generate images that can be shown in 8K or that require an 8K screen.
SOURCE 1 OF 6: UPSCALING

The main workaround for the lack of native 8K content (that is, content filmed in 8K) is upscaling, which takes place at the television set. This is the process whereby an input at a lower resolution, such as 4K, is converted into an 8K image using local (on-device) artificial intelligence (AI) algorithms and edge computing. Upscaling is considered successful if the viewer perceives an image converted to 8K as one that was originally captured in that resolution.

Upscaling is considered successful if the viewer perceives an image converted to 8K as one that was originally captured in that resolution.

Ample 4K content already exists that could be “fed” into an 8K TV with upscaling capability. Netflix, as an example, now has one of the largest 4K catalogs in the world, with over 1,000 titles offering a total of tens of thousands of hours of content.8

In upscaling, a TV set analyzes each frame and applies AI to identify the many different objects that may be in it, be it a pair of sunglasses, a burger, or a building. An upscaling engine then converts each 4K-resolution object into an 8K one, restoring edges, repairing any compression artifacts, and coloring each pixel as accurately as possible.9 To inform this conversion, the TV set draws on a database that may contain tens of thousands of image references.10 Upscaling quality for 8K TV sets should improve over time, as more conversions are done and the collective body of knowledge per vendor on how to optimize scaling grows. 8K TV upscaling algorithms can be updated over the air.11

Upscaling is not new to 8K. It was a feature of the first 4K sets in 2012 to address the lack of 4K content then available.12 Back then, upscaling was less sophisticated and convincing. Initial approaches were based on duplicating neighboring pixels: If an empty new pixel was next to a blue one, a blue pixel would fill the space, often resulting in a block of four blue pixels. Later versions of upscaling averaged up to 16 neighboring pixels in each direction to fill an empty new pixel, but again, this can create unsightly artifacts. However, with the development of more sophisticated, AI-driven techniques, these problems are well on their way to being resolved.

Upscaling from HD to 8K is also possible. This requires adding 15 pixels per HD pixel, a far more challenging conversion than for 4K to 8K. But as techniques improve, artificially generated 8K video may become ever more realistic, even from HD.

Upscaling can also be deployed as part of a compression process. Native 8K content is downscaled to 4K, and then restored via upscaling to 8K at the TV set guided by instructions added to the compressed video file. This approach, commercially known as smart downscaling, promises significant reductions in required bit rate for transmission. 8K high-efficiency video coding (HEVC) content can be compressed from 40 megabits per second (Mbit/s) to 25 Mbit/s,13 and AV1 (AO media video 1)-encoded 8K can be reduced from 35 Mbit/s to 17 Mbit/s.14 This approach enables 8K content to be delivered over broadband, satellite, or terrestrial broadcast connections originally designed for 4K.

SOURCE 2 OF 6: REMASTERING

Very high-definition analog and digital content

8K resolution is not completely new. Movies shot on 70mm film have a resolution equivalent to 8K.

A few movies have already been remastered in 8K (ultra-high-definition 4K digital remasters of...
65mm and 70mm movies have already been created). One of the first to undergo this process was *2001: A Space Odyssey*. This remastered version was broadcast by NHK in December 2018 on its BS8K channel. In March 2019, NHK also broadcast an 8K version of *My Fair Lady*. Dozens of other major movies that were shot on 65mm or 70mm could undergo the same treatment. Most iMax movies are shot on 65mm or 75mm as well, so films created for iMax cinemas, or with an iMax version available, could be converted to 8K.

Content shot in 6K, principally to provide more versatility in producing a reframed 4K master, may also be upscaled to 8K. The third season of Netflix’s *House of Cards*, for example, was shot in 6K. The season was released as 4K, but the 6K masters have been archived. Netflix’s *Mindhunter* was also shot in 6K for reasons similar to *House of Cards*.

**SOURCE 3 OF 6: CONTENT SHOT NATIVELY IN 8K**

A few major movies and TV series have already been shot in 8K, although they were released at 4K and lower resolutions. These include:

- *Homecoming*’s second season on Amazon Prime
- *Money Heist*’s fourth season on Netflix
- *Mank*, a black-and-white biopic of *Citizen Kane*’s cowriter, directed by David Fincher
- *The Eddy*, a musical drama by director Damien Chazelle, on Netflix
- *Guardians of the Galaxy 2*

In the future, once a large enough base of 8K TV sets exists, native 8K content could be released as an 8K stream or broadcast.

**SOURCE 4 OF 6: USER-GENERATED CONTENT**

User-generated content has exploded in the last decade. The caliber of photographic and video tools available to consumers at relatively modest prices has steadily grown, thanks to the proliferation of smartphones. One of the most popular applications for smartphones is photography, and in the near term, this may be one of the major sources of content that requires an 8K display.

Most current smartphones have at least one 12-megapixel (MP) camera, which capture images in 12 million pixels (dots of light which, in aggregate, make up each frame). At least a billion smartphones with 12MP capability are likely to be in use in 2021. 12MP images can only be displayed in their full resolution on an 8K TV: A 4K TV, with a mere 8 million pixels, is insufficient, but 8K televisions’ 33 million pixels are ample for 12MP. Mobile devices have collectively captured over a trillion images, of varying artistic caliber but consistent pixel count, which may only ever be seen in their full resolution when displayed on an 8K TV. Smartphones with 64MP and 108MP cameras are also in the market, and their full resolution exceeds that of even an 8K TV.

One of the most popular applications for smartphones is photography, and in the near term, this may be one of the major sources of content that requires an 8K display.

The quality of photos taken on a smartphone, particularly when natural light is available, is very high. These photos may look spectacular when shown off on an 8K television screen, especially if the TV incorporates software that sharpens the image and removes noise.
The next frontier of widespread 8K video is also fast approaching. A few high-end smartphone models offered 8K video capture as of August 2020, namely the Samsung S20, Xiaomi 10, and Red Magic S3. This number is likely to increase over 2021 and beyond. These devices can capture content in 8K at up to 24 frames per second; over the next couple of years, the frame rate is likely to increase. Most owners today may only dabble with 8K video, partly because of its large storage requirements (600MB per minute) and because of the slower frame rate relative to 4K and HD capture. However, consumers who do shoot videos in 8K capture could share this content via online platforms such as YouTube or Vimeo. Over the course of 2021 and the coming years, the volume of 8K videos captured on a smartphone should steadily grow as smartphone memory capacity increases and frame rates go up.

SOURCE 5 OF 6: PROSUMER CONTENT CAPTURED IN 8K OR GREATER THAN 4K

Prosumer photographers, whose subjects span birthdays to corporate videos, may start to experiment with 8K capture, as could aspiring moviemakers with smaller equipment budgets. These content creators now have access to compact 8K cameras whose cost starts at about US$4,000, available from brands including Canon and Sharp. A decade ago, 8K prototype cameras weighed hundreds of kilograms; Canon’s current 8K camera weighs less than one kilogram.

Cameras that can capture content in 6K, which can then be upscaled to 8K are also available. As of August 2020, companies offering cameras with 6K capture included Canon, Sony, Panasonic, and Blackmagic Design.

8K prosumer content is already available online to watch on 8K TVs. YouTube has accepted 8K uploads since 2015, and Vimeo has thousands of videos tagged as 8K.

SOURCE 6 OF 6: GAME CONSOLES WITH 8K SUPPORT

4K has already been a differentiator for prior generations of game consoles, and 8K may prove the same. The year 2021 will be the first full year in which a new generation of 8K-ready game consoles will be sold. There will probably be few titles to play in this resolution in 2021, but over the next five years, more and more 8K games should become available. The latest consoles have been designed to cope with much larger file sizes, with eight core processors that are able to deal with more complex video. That said, 8K screens will likely appeal more to amateur users than to the much smaller population of professional game players, who favor higher refresh rates over resolution and would likely keep on gaming in 4K or HD.

Prosumer photographers, whose subjects span birthdays to corporate videos, may start to experiment with 8K capture, as could aspiring moviemakers with smaller equipment budgets.
8K’s USES BEYOND ENTERTAINMENT

8K can have applications far beyond its use in entertainment. For example, one potential use of 8K could be for remote working. An 8K panel could be used to display multiple pages of content that a person or team is working on more crisply than would be possible at lower resolutions. These pages—perhaps a blend of diagrams, video calls, charts, and data sheets—could all be visible to a remote team scattered in different locations. This would be an upgrade from working from a single screen or having an array of monitors.

Office locations already use large panels for video calls and collaborative working. But for people expecting to work predominantly at home during 2021, large panels equipped with 8K could improve productivity due to their superior resolution. Documents viewed close up (from 20–30 centimeters), which may be pixelated on an HD or even 4K screen, would look clear on an 80-inch screen with 8K screen resolution. Video calls with dozens of colleagues could also be more practical on a large screen, with every individual discernible versus being visible only as a blurred rectangle.

The cost of 8K panels for work purposes may be tax deductible and/or subsidized by an employer. And they could of course switch applications once the working day is done, and be used to watch entertainment programs or play video games instead.

8K screens could also be used to display for online exercise classes, a content genre whose popularity surged during lockdowns while people were not able to go to the gym, go for a run, or ride a bike. The falling price of HD monitors and cameras enabled the first phase of online, interactive exercise classes featuring treadmills and static bikes with integrated screens. Large 8K screens, displayed on a wall, would enable instructors and fellow athletes to be shown larger—even life-size. Screens could also be used to show performance metrics from fitness bands and smart watches.

Still another use for 8K panels is for “digital wallpaper” that decorates part of or all of a large screen. Ever since the advent of digital screens, digital images have been used as screensavers, showing everything from tropical fish to fractal images to personal photos. Vendors of TV streaming peripherals have also included screensavers in their products. 8K TVs, which display a greater range of colors than 4K sets, can expand the range of images that can be effectively shown, including artwork from museums and private collections. A few digital art-as-a-service providers are already in business, and their number is likely to grow as more 8K sets become available.

Surmounting the cost hurdle

One major barrier to consumer adoption of 8K screens is their cost—for now. But in the near future, the cost is likely to drop substantially, especially when one calculates the cost per hour of viewing. Combined with TV screens’ appeal as a status symbol, this may be enough to push many consumers to consider an 8K set good value for their money.

8K PANELS SHOULD STEADILY FALL IN PRICE

8K sets are unlikely to enjoy the exact same pace of adoption as 4K sets, which took only seven years from their launch to become the most popular resolution in 2019, when over 100 million 4K TV sets were sold. 8K adoption will be slower because of their higher costs and larger size. But the dynamic for 8K is likely to be the same: As price falls, demand will ramp up commensurately.
In 2017, the ASP for 8K sets was more than US$8,000. The following year, it had decreased to about US$5,500. By the end of 2021, we expect that entry-level 8K TV sets will be offered for US$1,500 or less, with an ASP of US$3,300 (figure 2). And as prices have fallen, sales have grown. In 2017, about 2,400 8K TVs were sold. In 2018, the first full year in which 8K TV sets were available, sales reached 18,600 units. In 2021, we forecast 1 million unit sales globally, compared to 550,000 the prior year. Though this is still a drop in the ocean relative to the approximately 220 million sets of all resolutions sold in 2018, the upward trend is clear.

8K sales are likely to accelerate as the cost differential between 8K and 4K panels narrows. The smaller the differential, the more likely consumers will opt to future-proof, particularly for larger, higher-priced TV set sizes (65-inch or greater). Moreover, as demand for 8K TV sets rises, supply is likely to ramp up, further driving down their price. In 2018, the global manufacturing capacity for 65-inch displays (of all resolutions) was just 12 million. In 2022, we expect this capacity to increase to support the production of 154 million 55-inch TVs and 37 million 65-inch TVs, with much of the incremental capacity focused on 8K screens. This also parallels what happened with 4K: The number of brands offering 4K TV sets globally more than doubled between 2016 and 2019, from 30 to 70.

While lower prices are good news for consumers, the rate of price decline may, in the medium term (from 2025 and on), largely cancel out any increase in revenues for 8K set manufacturers and retailers. Purveyors of 4K sets have already experienced this phenomenon. In 2020, 4K TV unit sales in the United States were expected to grow by 12% to 25 million, but revenues were projected to increase by a mere 2%.

FIGURE 2
8K panels should fall in price in the short to moderate term
Average selling price of 8K panels, 2017–2022, worldwide

TELEVISIONS, INCLUDING 8K MODELS, CAN COST AS LITTLE AS 25 CENTS PER VIEWING HOUR

Spending US$2,500, or even US$1,500, for an entry-level 8K TV may appear very hard to justify. But with an expected lifetime of seven years and an average usage of three to five hours per day, a US$2,500 set would cost less than a dollar a day and about 25 cents per hour of viewing time, with the cost shared among the entire household.

To put this into context, many consumers are likely to spend more per year on a smartphone. Multiple brands of smartphones cost more than US$1,000, and their ASP across all markets is about US$380.41 The expected lifetime for phones purchased in 2020 is forecast at only 2.7 years.42 Hence, a household with three people may spend around US$3,000 on smartphones over a seven-year period if each person purchases a US$380 smartphone, and around US$7,800 if each person spends an average of US$1,000 per device.

Few consumers are likely to calculate the cost per hour of a new TV set. But the volume of usage of a new 8K TV and the range of applications for which it can be used—as well as the satisfaction of showing a new, large, device to friends and family—are all factors that can help justify its purchase.

MORE TIME AT HOME MEANS MORE TV SALES

TV sets were popular sellers during the lockdown phase of the COVID-19 pandemic in 2020. In the United States, unit sales of TV sets 65 inches or larger went up by 52% in the first half of 2020, and larger sets over 65 inches were up 77% in the second quarter.43 In Germany, TV sets sold strongly even as lockdowns were relaxed, with year-on-year sales up 37% in May 2020 and 21% in June 2020.44 Should a degree of lockdown continue through 2021, at least until a vaccine is widely available and applied, citizens may continue to spend more time at home and less time in indoor venues such as cinemas. To help pass the time, they may continue to choose to upgrade their home entertainment, partly by using savings made from not going out.
NEW BUSINESS MODELS MAY MAKE TELEVISIONS MORE AFFORDABLE

The revenue streams for TV vendors have historically remained largely the same. The manufacturer generates the bulk of revenues at the point of sale, and then waits until the sale of a replacement set to generate further revenues. This model has worked for makers and sellers of television sets, as well as multiple other appliances, for decades.

However, connectivity enables this business model to change in ways that can make TVs more affordable by monetizing information about usage habits. Rather than collect all revenues in one go, a vendor could trade gross margin at the point of sale for revenues from advertisers and content owners throughout the TV set’s life. This could enable TVs to be sold at a lower price, making the TV set more cost-competitive while generating potentially higher revenue over the device’s lifetime ownership. For consumers, this may mean that they are able to acquire an 8K set for the same budget as a 4K set. The only requirement would be to connect the TV and agree to terms and conditions.

Rather than collect all revenues in one go, a vendor could trade gross margin at the point of sale for revenues from advertisers and content owners throughout the TV set’s life.

A connected TV generates information useful to an array of vendors every time it is used. For content vendors, the TV can collect and share information about programs watched, channels chosen, and apps watched and deleted. It could also collect data on what programs are clicked on or, once started, paused or abandoned, which can provide a feedback loop to content creators or content aggregators such as video-on-demand companies.

Additionally, a connected TV can serve as a shop front for content providers’ apps. Space on the TV set can be rented to vendors, or commissions collected with every download. And it could be used to show ads distributed by the TV vendor, which could disintermediate TV broadcasters and other entities that have traditionally sold advertising space.

A connected TV can also collect data on devices in the home to help build a profile of the household based on the quantity and quality of devices owned. It could, for example, be connected to a smart doorbell or to a video baby monitor to be able to show who is at the door or to reassure parents that a baby is sleeping. The ability to collect additional data will depend on each market’s specific data privacy regulations.

This kind of business model is not new, of course. Though relatively novel to TV, other devices already collect or enable the collection of data useful to advertisers. Considering its success, TV is likely to adopt this model over the course of the 2020s.

THE RISING SIZE OF TV SETS IS LIKELY TO MAKE A STRONGER CASE FOR 8K

TV screen size has become steadily and considerably larger ever since TVs were first invented. Between 2004 and 2019, the average screen size of TV sets in the United States increased from 25.4 inches (as measured on the diagonal from the screen’s bottom left to top right corner) to 47 inches. This translates into a steady rise in the average screen size of the installed TV base of one inch per year.

The growth in TV screen size has been enabled by the shift to flat panels, which cost far less per square inch of screen to manufacture than those based on older technologies. Between 2014 and 2019, the
average cost per square inch of screen in the United States fell from US$2.15 to US$0.39, an 82% decline.47 Concurrently, sales of large-screen sales increased. According to one analysis, the most common size of TV set sold in 2019 in the United States was 65 inches. In the same year, the price of 65-inch sets fell by 25%.48 In the first half of 2020, US unit sales of TV sets 65 inches or larger were up by 52% in the first half of 2020, and larger sets over 65 inches were up 77% in the second quarter.49

The larger the TV screen, the more likely a customer may be to opt for 8K so as to have a perfectly smooth, pixel-free image. The number of pixels on the screen at a given resolution is constant regardless of screen size; as screen area increases, the number of pixels per inch (PPI) for each resolution declines (figure 3), resulting in a grainier image. On a 36-inch HD TV, its 2 million pixels are packed so densely that the image appears completely smooth.50 On a larger HD set, say 55 inches (more than double the area of a 36-inch screen), individual pixels may be discernible depending on the viewing distance. 4K video on a 55-inch screen may be completely smooth, but on a 65-inch screen (1.17 m², 40% larger than a 55-inch screen) individual pixels may start to become visible. On an 85-inch (1.99 m²) or 100-inch (2.76 m²) screen, only 8K content, native or upscaled, may guarantee a completely smooth image.

The sheer size of an 80-inch or larger TV set might appear to be a disincentive to purchase, as it would be too large to fit in most people’s cars. But online ordering and delivery are ready solutions for this. In the first half of 2019, one-quarter of all TV sets sold globally were bought online; in China, this figure was 50%.51 Lockdown is likely to have accelerated the move to online TV sales, even if research took place in physical stores.

**FIGURE 3**

8K’s higher pixel density becomes more important as screen size increases

<table>
<thead>
<tr>
<th>Screen size (diagonal, inches)</th>
<th>55</th>
<th>65</th>
<th>75</th>
<th>85</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen area (m²)</td>
<td>0.83</td>
<td>1.17</td>
<td>1.55</td>
<td>1.99</td>
<td>3.34</td>
</tr>
<tr>
<td>Pixels per inch, 8K</td>
<td>160</td>
<td>136</td>
<td>118</td>
<td>104</td>
<td>80</td>
</tr>
<tr>
<td>Pixels per inch, 4K</td>
<td>80</td>
<td>68</td>
<td>59</td>
<td>52</td>
<td>40</td>
</tr>
</tbody>
</table>

Source: Deloitte analysis.
Surmounting the comparison hurdle

Many studies have examined consumers’ ability to tell an 8K image apart from a 4K one. These studies tend to suggest that differentiating between the two is hard, with capability varying by genre of content. Nature documentaries appear to benefit most from 8K, perhaps because of the greater range of natural colors that can be displayed with the standard.

However, the inability to tell 4K from 8K may not matter in the long run. What matters more may be that owners of 8K TV sets enjoy—and derive utility from—the knowledge that their flat panel is capable of showing images in 8K. This phenomenon recognizes that the consumer is not always rational. The rise of the 4K format occurred while most content was available only in HD, and the rise of HD happened while standard-definition content prevailed. Even today, moviegoers may only know that they are watching in 4K if a trailer tells them so: They are not readily able to tell the difference.

The rise of the 4K format occurred while most content was available only in HD, and the rise of HD happened while standard-definition content prevailed.

The popularity of 8K TV sets is likely to be boosted by the appeal of very large panels of 75 inches and larger. The larger the TV set, the more immersive the experience. The television is steadily but inexorably transitioning from being the bulky cube in the corner to the slender panel occupying ever larger expanses of wall in ever more intense and vivid colors. 8K and very large panels are likely to become regarded by consumers as one and the same, regardless of the underlying resolution of the content. This is likely to drive satisfaction among owners and envy among visitors to friends’ 8K-equipped living rooms.

8K’s production end

Though this chapter has focused mostly on the drivers for consumers to adopt 8K, we should also consider the incentives and barriers to creating content in 8K.

8K film and television cameras have existed for more than five years. Red, one of the first companies in the market, launched its first 8K camera in 2015. However, little content has been created in 8K to date. One reason for this is because few 8K TV cameras have been commercially sold: For much of the last decade, only prototypes were available. Additionally, shooting in 8K entails considerable incremental costs, such as for additional storage—one hour of raw 8K footage requires more than seven terabytes, or about 122GB per minute, of storage space—as well as for new monitors and cables.

That said, creating content in 8K also has several advantages that may balance out these current drawbacks. One of these advantages is that shooting in 8K may allow for better quality in postproduction. Downsampling 8K video to 4K often delivers a cleaner overall image than footage originally captured in 4K. This approach also eliminates certain types of artifacts in the final image, and may also help with image stabilization, which requires zooming in up to 20%. 8K capture for 4K masters is likely to be particularly useful for movies with many visual effects (VFX), whose images can be subject to significant postprocessing.

Reframing from higher-resolution images is also easier in 8K. Crops at 4K resolution can be readily...
created from an 8K file without loss of resolution, delivering smaller pixels and more detail. For sports, shooting in 8K enables fast-moving subjects from runners to racing cars to be more easily framed in 4K. Indeed, 8K cameras, whose prices as of summer 2020 started from US$4,000, may even change how sports photography is done. Sports photographers have traditionally relied on ever-faster SLR cameras capable of shooting multiple frames per second, with the hope that one of the hundreds of photos taken will be the money shot. With 8K video, the photographer can film the action at the highest resolution, then pull off a single frame, which will be 33MP in size. Nature documentary makers, whose subjects may move as unpredictably as footballers around a pitch or boxers in a ring, can also benefit from 8K video to generate high-definition stills. The cost of creating in 12K to be able to downsample to 8K now starts at US$10,000, with the launch of Blackmagic’s 12K camera in the summer of 2020.

Sports photographers have traditionally relied on ever-faster SLR cameras capable of shooting multiple frames per second, with the hope that one of the hundreds of photos taken will be the money shot.

Finally, shooting in 8K extends the period in which content could be monetized. Content that is shot in 8K and mastered to 4K for release in 2021 could be rereleased in 8K in subsequent years as more 8K screens become available. This is important because the lifetime of content can be measured in decades. For instance, I Love Lucy, which first aired in 1951, was filmed in 35mm, enabling remastered versions to be created from the original recordings many years later. An HD remaster of I Love Lucy on Blu-ray was released in 2014.

Beyond the consumer to commercial environments

Most 8K content created or remastered in 2021 is likely to be targeted to consumers for consumption at home. But a growing volume of 8K content is also likely to be created for commercial uses such as:

- **Retail displays.** Video displays have long been a major weapon for attracting and selling to customers in physical retail. As retail recovers from lockdown, 8K screens are likely to become an increasingly important differentiator, used both to lure customers into a store and to market products and brands. For retail installations, 8K screens can be built from modular panels, and they can also be attached to curved surfaces. They do not need to be 16:9 rectangles.

- **Advertising panels.** 8K is likely to become increasingly important for outdoor and instore panels due to its higher resolution and greater color range.

- **Trade shows.** Presenters at trade shows need to compete relentlessly for attention. 8K screens, bespoke or based on standard panels, can be one way of attracting visitors to a stand, continuing a decades-long trend of screens being a principal way to draw in potential customers.

- **Live events.** Concerts are designed to be spectacular, and a major element in this are the vast, bespoke screens that serve as a backdrop to musicians. Major concerts featuring 8K walls are likely to become increasingly common once live events are able to resume safely. One of the first bands to use an 8K screen was U2, whose 2017 Joshua Tree tour featured a five-story
Content that is shot in 8K and mastered to 4K for release in 2021 could be rereleased in 8K in subsequent years as more 8K screens become available. This is important because the lifetime of content can be measured in decades.

(14-meter) 8K screen that stretched the 61-meter width of the stage, alternatively showing high-resolution visuals and live images of the band. Other stars to feature 8K screens in their concerts include Taylor Swift and Lady Gaga.62

- **Office space.** Reception areas give visitors their first impression of a company. 8K screens can be used in these areas to showcase a company’s messaging, such as its brand values. Larger companies may be able to readily afford the cost of premium 8K panels costing tens of thousands of dollars.63

8K’s implications for adjacent industries

8K’s spread among consumers can create opportunities for adjacent industries, with the telecommunications industry in particular standing to benefit. Operators should consider 8K TV sets as an opportunity to market higher-speed connections. They could bundle their highest-speed connections—1 gigabit per second (Gbit/s) in many markets—with subscription video on demand (SVOD) services offering 8K content, pairing the highest picture quality with the best broadband connections. However, operators should not expect gains in revenue from increased network traffic just yet. In 2021, 8K is unlikely to have a significant impact on network traffic overall, not only due to the still-niche installed base of 8K sets, but also because most content shown on 8K-capable panels is likely to be in lower resolution.

For their part, SVOD providers offering content in 8K may be able to bundle this content as part of a premium package. They will incur additional costs for storing and delivering higher-resolution content, but some of these should be recoupable by charging a premium for 8K content, as many pay TV (including SVOD) operators already do for

TV’s New Year’s resolution: The start of the 8K wave
THE BOTTOM LINE

Ultimately, 8K’s core attraction to consumers is as an option on the future of video, and it will appeal to buyers for emotional as well as rational reasons. Owning an 8K screen offers the prospect of enjoying movies and television programming in the best possible quality in terms of pixel density, color range, screen brightness, and sound. It opens up the possibility of showing next-generation 8K video games on the best available screen. It also gives consumers the option of using large TV sets in new ways. They will no longer be just for watching video content, but also could be used to display digital wallpaper or, for home workers, productivity tools.

Commentators should not dismiss 8K’s prospects on the grounds of irrationality. Similar arguments were made about 4K: The screens would be too big for rooms, viewers would be seated too close or too far from the screen, or few would be able to discern the difference between HD and 4K. These resoundingly logical objections do not appear to have held sway. The majority of new TV buyers will opt for 4K in 2021, and they are likely to increasingly opt for 8K over the coming years.

For virtually all consumers today, a television set is a necessary fixture of everyday life. A few households will, of course, prioritize bookshelves over TV girth, but these may well be increasingly rare. Besides 8K’s appeal in terms of video and sound quality, the mass market is likely to relish a screen that dominates and defines a room rather than one hidden in a corner, and many will be eager to claim the bragging rights of owning the latest, largest state-of-the art TV set. 8K televisions are eminently suitable for all these purposes—and this will position them to invade the TV market in 2021 and beyond.
Endnotes


2. In the term “8K,” K refers to one thousand, and the naming is based on the approximate number of horizontal lines, which is 7,680. The total number of pixels is derived from multiplying the number of horizontal lines by the vertical lines.

3. 4K is a reference to the number of horizontal lines, rounded up to the nearest thousand. There are 3,840 horizontal lines on a 4K screen and 2,160 vertical lines.


5. The audio specification splits sound into 24 separate channels, spread across 22 speakers and two subwoofers. However, as few homes are likely to have 24 speakers, 8K video is likely also to be offered in 5.1 sound, which is allocated to six speakers. See NHK STR, “22.2 multichannel audio format standardization activity,” 2011.

6. The majority of content is likely to be filmed in HD or SD (high definition, standard definition). A small proportion of content will be filmed in 4K.

7. For more information on one test undertaken, see Scott Wilkinson, “8K vs 4K TVs: Double-blind study by Warner Bros. et al reveals most consumers can’t tell the difference,” TechHive, February 28, 2020.


10. Michael Hicks, “Here’s the secret behind 8K AI upscaling technology,” TechRadar, August 26, 2019.


12. For an explanation of how upscaling was applied on the first 4K TVs, and why this sometimes delivered sub-optimal images, see Hicks, “Here’s the secret behind 8K AI upscaling technology.”


14. Ibid.


16. My Fair Lady had already been scanned into 8K a few years back in order to create a 4K version. The 8K version created for 8K broadcast was based on the existing archived scan. See Carolyn Giardina, “My Fair Lady’ gets 8K restoration,” Hollywood Reporter, March 8, 2019.


20. There are also a few devices with 64-megapixel and 108-megapixel capabilities that may output photos as 12 megapixels by compressing nine pixels into one. For more information, see Sam Byford, “Why Samsung’s 108-megapixel camera isn’t just a gimmick,” The Verge, February 12, 2020; Sam Byford, “Do 64-megapixel phone cameras really make a difference?,” The Verge, October 1, 2019.


27. Ibid.


29. For example, Apple Fitness+ enables data from the Apple Watch wrist wearable device to be displayed on a television set via the Apple TV digital media extender. See Apple, “Apple Fitness+,” accessed September 23, 2020. TMT Predictions 2021 is an independent publication and has not been authorized, sponsored, or otherwise approved by Apple Inc.

30. Amazon Fire TV, “All 182 screensavers on your Amazon Fire TV and their locations (photos),” June 16, 2017; Benjamin Mayo, “Apple TV now features 10 beautiful underwater video screensavers,” 9to5Mac, September 25, 2019. TMT Predictions 2021 is an independent publication and has not been authorized, sponsored, or otherwise approved by Apple Inc.

31. For example, see Loupe, accessed September 23, 2020.


33. The Korea Bizwire, “8K TV war intensifies but consumers still prefer 4K.”

34. Ibid.

35. Ibid.

36. 8K TV sets first went on sale in the last quarter of 2017. See GfK, “4K is becoming the standard for today’s TVs.”

37. Advanced Television, “IHS: Over half of Q4 TV shipments were UHD,” March 7, 2019.


39. GfK, “4K is becoming the standard for today’s TVs.”


45. A further type of business model would be to offer the TV set on a subscription basis, with the cost depending on the specification of the TV, and the quantity of advertising viewed.

46. Advanced Television, “IHS: Over half of Q4 TV shipments were UHD.”


49. Snider, “Despite coronavirus pandemic, consumers still turned on by big-screen TVs.”


51. GfK, “4K is becoming the standard for today’s TVs.”


53. Display Daily, “Production of 8K content for 8K TVs.”


55. Silberg, “8K ecosystem is ready for implementation.”

56. Ibid.


58. Carter and Artaius, “What is 8K? And what does it mean to photographers and videomakers?”


60. Leigh Allen, “Filming the I Love Lucy show,” American Cinematographer, April 1, 2020.


How can a company train workers to unload hazardous materials, configure a wind turbine, or service a jet engine when a pandemic makes it impossible to teach and learn these skills in person? One way to do it is to use virtual reality (VR), augmented reality (AR), and mixed reality (MR) to simulate those environments for workers to practice in. We predict that, led by purchases by corporations and educational institutions, sales for enterprise and educational use of wearable headsets for VR, AR, and MR—collectively known as XR or digital reality—will grow by 100% in 2021 over 2019 levels.

Overall spending on AR and VR headsets, software, and services, including purchases by consumers, rose in 2020 to US$12 billion globally, up 50% from 2019. Although this figure is lower than the prepandemic forecast of almost 80% growth, it was much better than worldwide IT spending, which declined by more than 5% for 2020 year over year. Postpandemic, higher growth is expected to resume for XR, with one group predicting the industry will reach a total of US$73 billion in 2024, or a 54% annual growth rate between 2020 and 2024.

Although the predicted growth rate in headsets specifically is off a low base, with fewer than 100,000 VR, AR, and MR headsets purchased annually by enterprises and schools from 2015 through 2019, the upward trend appears clear. Market growth for these types of headsets has already accelerated in some markets due to the risk of COVID-19 infection driving their use in teaching employees and students virtually rather than in person. With the pandemic accelerating the opportunity to demonstrate their value, digital reality headsets may continue to gain ground after the pandemic ends due to a variety of other benefits, such as lower cost, greater safety, and better learning retention.

XR’s pivot to the enterprise market

Within the total XR industry, enterprise applications such as training and industrial maintenance were predicted to generate US$1.3 billion and US$0.4 billion in 2020, respectively. These numbers are still smaller than consumer sales—but over the next few years, organizational purchases of XR will likely narrow the gap, with all of the fastest-growing digital reality markets expected to be in enterprise or education. Sales of XR for use in public infrastructure maintenance, industrial maintenance, and logistics and package delivery management are predicted to more than double every year from 2019 to 2024. And both post-secondary and K–12 lab and field spending on digital reality is expected to rise more than 120% annually over the same period.

It may be surprising that organizational XR sales are expected to gain ground on those to consumers, especially in a time of pandemic when consumers might have flocked to VR headsets to alleviate the boredom of being locked down at home. (The consumer headset market is almost entirely for VR headsets, with AR and MR being primarily
In tech blogger Ben Evans’ memorable phrase, pandemic lockdowns were a kind of “forced experiment” for various technologies—including work-from-home tools, online grocery delivery, and home entertainment setups for gaming and video streaming—and VR seems a natural technology for consumers to add to the mix. But although consumer VR did not collapse in 2020, neither did it surge. To quote Ben Evans again: “This should have been a [consumer] VR moment, and it isn’t.” Consumers bought about US$2.9 billion worth of VR headsets in 2020, down 12% from US$3.3 billion in 2019, though sales are expected to rebound to US$3.5 billion in 2021. That 2020 decline is better than what happened to cinema attendance during the COVID-19 lockdowns, which were down 66% worldwide for the year. But it is less good than game console sales, which were up 150% annually in March 2020 as billions of people confined to their homes sought ways to entertain themselves.

Given lower-than-hoped for growth in the consumer market, XR headset makers have been shifting to the enterprise. The first AR headset, Google Glass, was originally intended for the consumer market at its release in 2014, but it was relaunched in 2017 in a pivot to the enterprise market. In April of 2020, Magic Leap announced that it too had decided to pursue enterprises as its primary market. Microsoft’s HoloLens has always been targeted at the enterprise, not the consumer. Even consumer VR companies such as Facebook-owned Oculus and HTC started enterprise divisions in 2019. There are rumors that Apple may launch AR and VR products in 2021—likely mainly for the consumer at first, but they may also have enterprise and education applications over time.

Headsets, of course, are only part of a complete XR package. Adding up all the spending on enterprise XR, it is likely that the enterprise digital reality market generated revenues of US$13 billion in 2019, up 19% from the prior year. This number is larger than the total XR market mentioned earlier, as it includes internal R&D, which is large at this stage of the enterprise digital reality market evolution. Determining the unit sales and dollar value of the headset portion only, excluding software, services, and internal R&D, is a difficult task: No XR makers disclose segmented sales in dollars or units on a quarterly or annual basis. But despite this difficulty, it is still possible to glean some information from occasional announcements by headset makers, enterprises, and educational institutions, as well as from media coverage.
Why aren’t enterprise and education headset sales higher?

Between 2015 and 2020, there were dozens of different public announcements regarding enterprise and education digital reality. Deloitte Global has analyzed these public announcements and found some important trends.

HEADSETS ARE SHARED TECHNOLOGY, NOT PERSONAL

Although some enterprise technology tools such as the PC and the smartphone are personal (one per employee), other tools such as printers and LED projectors are shared among many employees: Tens or hundreds of employees use the same device as needed, and for only a brief period of time per use. The latter pattern of use is emerging as typical for XR headsets. For instance, in 2018, Walmart obtained 17,000 Oculus Go entry-level VR headsets, sending four units to all of its supercenters and two units to smaller locations. Using these 17,000 headsets, the company was able to train over a million employees on more than 45 different modules about new technology, customer service/empathy, and compliance. On average, almost 60 employees used each headset.

NOT EVERY EMPLOYEE MAY NEED A HEADSET

It may be obvious that some types of workers, such as office workers, have little need for XR headsets. But even in jobs where digital reality headsets might be useful, not every employee may require one. One of the more common uses for XR headsets in the workplace is to onboard new hires only, rather than using them to support existing employees.

NOT EVERY LOCATION NEEDS HEADSETS

The Canadian province of Saskatchewan has 1.2 million residents, 40% of whom live in remote areas. It has deployed two pairs of Google AR headsets in each of 11 communities so that onsite medical practitioners can consult in real time with experts in urban medical centers. The tool has proved “invaluable for wound treatment,” according to Ivar Mendez, unified head of the Department of Surgery at the University of Saskatchewan. But as useful as digital reality headsets can be in remote areas, they are not needed for such applications in urban areas, where a majority of the world’s population lives.

HEADSETS MAY BE ONLY A SMALL PART OF A PROJECT’S TOTAL COST

The costliest publicly announced XR project to date is the 2018 US$480 million deal between the United States Army and Microsoft using the MR HoloLens. This agreement was not for off-the-shelf headsets, but for customized devices with thermal sensing and night vision used not just for training but on the battlefield as well. Although there have been discussions of follow-on orders for 40,219 headsets costing over US$2 billion over several years, the initial deal covered only 2,500 headsets over two years. Even if these headsets cost 10 times more than off-the-shelf HoloLens units, or about US$30,000 each, the total hardware value of 2,500 headsets would be a mere US$75 million. In other words, the software, services, and development portions of the overall MR solution likely represented more than five-sixths of the total contract value, and the headsets themselves only about 16%.

DIGITAL REALITY HEADSETS ARE STILL IN THEIR EARLY DAYS, AND ARE MOSTLY USED IN PILOTS

More than half of the public announcements concerning XR headsets include the words “pilot,” “trial,” or “test.” The training, enterprise, and education market for headsets is still relatively nascent, and it is, therefore, not surprising that individual companies and schools have only bought tens or hundreds of units. That said, as the Walmart and US Army stories illustrate, follow-on purchases can be on the order of tens of thousands of units.
Putting all of that together, the headset markets that are moving fastest right now are in immersive training, especially where real-world training would be dangerous, difficult, or expensive; for frontline health care workers; for use in retail (consumer-facing, but still an enterprise use case); and for building digital reality strategies across the domains of hardware, software, and services.

What about education?

As mentioned earlier, the market for educational XR is poised to be among the fastest-growing XR segments over the next few years. Admittedly, this growth is off a very small base. Educational uses of digital reality have been embryonic between 2015 and 2020; according to one report, the global education XR market was only US$0.68 billion in 2019, and the headset component of that (as distinct from software, content, and services) is likely less than US$100 million. Our research on headset announcements and partners uncovered no large education-specific pilots. Some small pilots do exist, but these use only a few headsets. For example, Brock University in Canada was using VR in classrooms prior to the pandemic, but it only has six headsets.

That said, as schools and colleges have been shut down due to the pandemic, XR and XR headsets are proving a valuable tool. As one example, a Canadian postsecondary institution is using VR for welding and automotive painting vocational programs, and has found the technology so useful and safe that it “will likely continue to use it when COVID-19 restrictions ease.” And one university in Kentucky has been teaching classes during the pandemic with HTC Vive VR headsets (although it has only 18 of them). Other schools in California, Michigan, and Mexico also use a variety of digital reality headsets for MBAs and other programs.

THE BOTTOM LINE

As enterprise and education XR headset sales grow, it is worth keeping in mind that these sales are likely to continue to represent a minority of total digital reality project spend compared to software, development, content, and services. Over time, however, the hardware component will likely grow as a percentage of project value, as many of the other costs tend to be upfront while headsets become more material as pilots turn into full deployments.

What could boost enterprise headset sales even more would be if they follow the same trajectory as several other workplace devices. In the 1980s, some businesses had a single portable computer or radiotelephone/cellular phone for communal use. Over time, these devices became seen as so useful that every employee had to have their own, and price points dropped so significantly that doing so became affordable. A similar shift in both perceived utility and price for digital reality headsets could increase the market by 10 times or more.

It is unclear how XR headsets will fare postpandemic. It is possible that some of the enterprise and education use of digital reality headsets will be a blip: The headsets will be used during lockdown periods, and then discarded when things return to normal. However, for most organizations, it seems more likely that the COVID-19 period will be a crucible in which XR headsets prove their usefulness, spurring continuing growth. After all, if it’s too dangerous, too difficult, or too expensive to train in the real world, why wouldn’t you train in a virtual one?

Price will be one important factor spurring growth. It is anticipated that both existing manufacturers and new entrants will introduce high-quality digital reality headsets under the magic US$1,000 price point. Other reasons to anticipate increase in the use of digital reality include:
• Hard numbers on some enterprise VR programs so far show that they improve productivity by an average of 32%.²⁷

• Academic research suggests that AR is better than video in workplace settings. Employees prefer it, their problem-solving improves, they make fewer errors, and they perceive it as more efficient than a standard video call.²⁸

• AR training yields a 75% learning retention rate, higher than almost any other form of training. (Lectures and reading have only a 5%–10% retention rate.)²⁹ Another study showed that AR results in greater knowledge transfer and more than doubles learning outcomes.³⁰

• VR appeals to a variety of learning styles,³¹ and is especially useful for training that requires repetition and retention.³²

• For learning in dangerous environments (such as for firefighters), VR is safer and less risky for employees and students.³³

With VR, people do not need to travel to access training devices, and they do not need to bring heavy equipment to a special training location.³⁴

**If it’s too dangerous, too difficult, or too expensive to train in the real world, why wouldn’t you train in a virtual one?**

Companies and educational institutions looking to deploy XR can consider several best practices:³⁵

• **Make it impactful.** XR is about business outcomes and ROI, not about building shiny objects. Digital reality should solve problems in ways that were not otherwise possible.

• **Make it engaging.** Technology for technology’s sake isn’t helpful to anyone. It has to fit into the way humans work in order to achieve better outcomes.

• **Make it flexible and scalable.** A well-designed solution is built to evolve with new developments in the technology.

• **Make it work with change management.** XR is going to have far-reaching effects on workplaces and schools in the years to come, which will require new ways of working and thinking.

• **Make it easy.** Digital reality headsets should be more plug-and-play than they are now, especially for consumers, but also for enterprise and education.

• **Make it physically attractive.** Many early XR headsets were too large, uncomfortable, or ungainly. Physical appeal matters even more for the consumer market, but it applies to both enterprise and education markets too.

Not all work and learning are suitable for XR—for example, to receive fine-grained tactile feedback when practicing surgery or delicate mechanical operations. Yet the future may see advances in haptics as the technology further develops.³⁶ And as it does, we can expect digital reality to become more widespread among businesses and educational institutions alike, transforming the way we get the job—or the learning—done.


4. IDC, “Worldwide spending on augmented and virtual reality forecast to deliver strong growth through 2024.”

5. Ibid.

6. Ibid.


15. MacRumors, “Apple glasses,” October 22, 2020. *TMT Predictions 2021* is an independent publication and has not been authorized, sponsored, or otherwise approved by Apple Inc.


28. XMReality, “Field study shows that remote guidance using augmented reality (AR) is more effective than a standard video call,” press release, October 30, 2018.


32. DeMarinis et al., Real learning in a virtual world.


OF ALL THE activities that COVID-19 brought online, video doctor’s visits may be the one that caused the most personal trepidation. After all, how can a doctor take your blood pressure, examine your throat, or evaluate a skin tumor over Zoom or Skype? But as it turns out, many consumers (and doctors) have been quick to change their minds about video visits’ efficacy and appeal, and they are now prepared to do it that way for the long term.

We predict that the percentage of total visits to doctors that are done virtually via video will rise to 5% globally in 2021, up from an estimated 1% in 2019.1 While 5% may not sound like much, consider that 8.5 billion doctor’s visits, worth a total of approximately US$500 billion, took place in the Organisation for Economic Co-operation and Development (OECD) 36 countries in 2019 alone. Five percent of that would translate into more than 400 million video visits and about US$25 billion in value, depending on how much doctors are paid (either directly by a patient, by insurance, or by national health insurance) for video visits compared to in-person ones.

The relationship of this growth to COVID-19 is clear. In April 2020, 43.5% of all US Medicare primary care visits were via telehealth; prepandemic, this figure stood at just 0.1%.4 While telehealth also includes phone calls, emails, and nonvideo software solutions, the number of visits by video rose greatly. The number of people using the Department of Veteran’s Affairs Video Connect system rose to 120,000 per week, compared to 10,000 per week in the same period in 2019.3 Meanwhile, in the spring of 2020, video consultation services in France rose by 40%–100%.4 And in a May 2020 survey, 14% of Canadians said they would choose a video doctor’s visit where possible going forward.5

More video and other types of virtual visits mean more business for the companies providing the technologies to support them. We predict that the market for pure-play telehealth virtual visit solutions will reach US$8 billion in 2021.6 Partially driven by the growth in virtual visits, we also expect that more than US$33 billion of medical-grade home health care technology (mainly therapeutic and monitoring solutions, which can include medically approved consumer products such as smart watches) will be sold in 2021, up almost 20% over 2019.7

What happened?

The technology for video visits has been around for years. However, several factors—COVID-19 foremost among them—are converging today to drive higher usage.

COVID-19 LEFT PEOPLE NO CHOICE, AND REGULATORY BARRIERS WERE LOWERED

In March 2020 alone, authorities in the United Kingdom, United States, and Germany loosened regulatory barriers, modified rules around privacy, and endorsed telemedicine, including video visits.8 To quote one UK doctor: “We’re basically witnessing 10 years of change in one week.”9
COVID-19 also forced people to learn how to use the software. During the pandemic, literally hundreds of millions of people who had never used video calling software and hardware before used it for the first time as they worked from home during lockdowns. Even before the pandemic, video calling was relatively easy to use ... but many people hadn’t. Postpandemic, almost everyone is now a veteran at setting it up, getting lighting levels right, muting and unmuting themselves, and so on.

Importantly, the newly skilled at video calling includes tens of millions of people over age 65, who visit doctors more frequently than younger individuals. Although only 17% of the population, the 65-plus demographic accounts for more than 30% of all doctor’s visits in the United States; in 2016, people age 65 and up made 80% more office visits than the average number of visits among the general US population. Historically, research has shown that elderly users, even when they have the right tools and connections, are still less likely to use digital applications, in part due to unfamiliarity. COVID-19 provided the impetus for them to change that. To stay in touch with children, grandchildren, and friends, the 65+ population underwent a rapid and forced training on video hardware and software. For them to be able to use it for virtual video visits with physicians is a fringe benefit that can drive substantial growth in the video visit market.

DEVICES AND CONNECTIONS ARE REACHING CRITICAL MASS, ESPECIALLY AMONG THE ELDERLY

The generational divide in digital device ownership has been rapidly narrowing in the last five years. As more older people become equipped to come online, their ability to engage in virtual visits will increase—an important consideration for those with limited mobility or other constraints that make it difficult to go to the doctor in person.

**FIGURE 1**

The ownership generation gap is narrowing in the United Kingdom for devices other than laptops

UK device ownership by age

<table>
<thead>
<tr>
<th></th>
<th>2015</th>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age 18–75</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smartphone</td>
<td>51%</td>
<td>82%</td>
</tr>
<tr>
<td>Tablet</td>
<td>60%</td>
<td>66%</td>
</tr>
<tr>
<td>Laptop</td>
<td>66%</td>
<td>66%</td>
</tr>
<tr>
<td>Any of smartphone, laptop, or tablet</td>
<td>79%</td>
<td>79%</td>
</tr>
</tbody>
</table>

Historically, although the devices needed for video calls (laptops, smartphones, and tablets) have been broadly ubiquitous, their ownership has been much less widespread among those over 65. However, as figure 1 illustrates, this has begun to change. In 2015, one in three Britons age 65 to 75 did not own a device capable of supporting a medical video call. By 2020 that figure had fallen to 1 in 25 as more older adults bought tablets and especially smartphones.

Of course, devices that are not connected are not useful, but here again the picture for those over 65 has markedly improved. In the five years between 2014 and 2019, overall internet usage in the United States rose six percentage points, but internet usage for people over 65 rose 16 percentage points. Although not all senior Americans were connected in 2019, based on the growth rate in penetration and the pandemic likely over three out of four were by mid-2020, which is likely a critical mass for enabling widespread medical video visits among the 65-plus population.12

**CONNECTIVITY’S REACH AND SPEED ARE IMPROVING**

Virtual visits’ growth will depend partly on the extent to which more of the world’s population becomes connected. As of 2017, about 12% of rural residents in the United States lacked access to fixed internet service of up to 10 megabits per second (Mbps) download and 1 Mbps upload.13 Furthermore, these speeds are maximum speeds: During periods of heavy use, when they are shared by multiple users in a home, speeds can be much slower. Most virtual video visit applications require at least 0.5 Mbps upload speeds, meaning that at times when connectivity is slower than the maximum for these users, their connections would not be able to support a video doctor’s visit.

It is also worth noting that access to connectivity is lower among certain populations besides those in rural areas. “Digital exclusion” is higher for those living in social housing, from lower-income groups, with visual and other disabilities, who are homeless or unemployed, who have lower levels of education, and whose first language is not the country’s native language(s).14 Because of this, these populations will likely be slower to adopt telemedicine in general and video visits specifically.

The connectivity situation, however, is slowly improving. Governments around the world, working with network operators, are trying to get more citizens connected to the internet, and at higher speeds, especially in rural areas. As telemedicine becomes increasingly important in delivering health care (once again, especially in rural areas), we can expect those initiatives to accelerate.15 It is also possible that low Earth orbit satellite constellations may be able to provide high enough speeds for ubiquitous global coverage.
though these networks are not yet in full service, and many questions about them remain, especially around affordability.16

5G will further accelerate telemedicine, as it allows for faster transmission of large image and high-quality video files, better augmented reality/virtual reality and spatial computing, and more reliable connections with guaranteed quality of service. In some cases, 5G could even allow telemedicine to move beyond diagnosis and monitoring, enabling doctors to perform actual procedures and surgeries using ultra-low latency (under 10 millisecond) virtual technology.17

COVID-19 MAY MAKE VIDEO CALLING THE DEFAULT
Historically, most electronic communications for consumers and businesses have been email/messaging or voice-only calls, with video only used when absolutely required. Although we don’t have hard data for how that may be changing now, we see signs that, as one commentator has observed, “Video calling is the new normal, and it’s all because of coronavirus.”18 Some of video calling’s appeal may be due to the medium’s novelty, but if a preference for video over nonvideo methods is indeed rising, it has two critical implications for medical video visits. First, if we see a permanent shift to video over email and voice calls, then our prediction of 5% of all visits being video will likely be much too low. Second, video may make virtual visits much more medically effective. Obviously, a video call helps a health professional assess a laceration or rash, but it also has the benefit of showing the patient’s and health professional’s faces. A seminal 1979 study, long before video calling was widespread, concluded that “Effective nonverbal communication—facial expression, voice tone, etc.—is essential for successful patient-practitioner interaction.”19

Sizing the global doctor visit market
Let’s return to our prediction for the revenue generated by video doctor’s visits. Why do we believe that video visits will be worth about US$25 billion in 2021?

The market for physical visits with doctors is very large, but although there are some sources that give the number of annual visits for a given country, no single source tallying global visits exists. Using two different approaches to size the market, we believe that patient visits to doctors (both physical and virtual) are likely to generate more than US$700 billion globally in 2021. OECD member states will account for more than half a trillion dollars of this figure in 2021; if 5% of these are virtual, that works out to about US$25 billion for video visits.

METHODOLOGY 1: ADDING UP PER-COUNTRY COST PER VISIT
The OECD publishes data on annual per capita doctor consultations by country.20 We multiplied the latest available data by each country’s estimated 2020 population to determine the number of annual doctor visits per country (figure 3). This analysis suggests that across the 36 OECD countries, with a total population of 1.31 billion people, more than 8.7 billion visits take place every year.
FIGURE 3

More than 8.7 billion physical and virtual doctor's visits globally take place each year

Total number of annual doctor’s visits by country, 2019 or latest available data

<table>
<thead>
<tr>
<th>Country</th>
<th>Annual visits (millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan</td>
<td>1594</td>
</tr>
<tr>
<td>United States</td>
<td>930</td>
</tr>
<tr>
<td>South Korea</td>
<td>866</td>
</tr>
<tr>
<td>Germany</td>
<td>829</td>
</tr>
<tr>
<td>Turkey</td>
<td>801</td>
</tr>
<tr>
<td>Italy</td>
<td>411</td>
</tr>
<tr>
<td>France</td>
<td>385</td>
</tr>
<tr>
<td>Mexico</td>
<td>361</td>
</tr>
<tr>
<td>Spain</td>
<td>341</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>340</td>
</tr>
<tr>
<td>Poland</td>
<td>288</td>
</tr>
<tr>
<td>Canada</td>
<td>253</td>
</tr>
<tr>
<td>Australia</td>
<td>199</td>
</tr>
<tr>
<td>Netherlands</td>
<td>154</td>
</tr>
<tr>
<td>Hungary</td>
<td>103</td>
</tr>
<tr>
<td>Colombia</td>
<td>94</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>88</td>
</tr>
<tr>
<td>Belgium</td>
<td>83</td>
</tr>
<tr>
<td>Chile</td>
<td>73</td>
</tr>
<tr>
<td>Portugal</td>
<td>71</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>60</td>
</tr>
<tr>
<td>Austria</td>
<td>59</td>
</tr>
<tr>
<td>Israel</td>
<td>54</td>
</tr>
<tr>
<td>Switzerland</td>
<td>37</td>
</tr>
<tr>
<td>Greece</td>
<td>34</td>
</tr>
<tr>
<td>Sweden</td>
<td>27</td>
</tr>
<tr>
<td>Lithuania</td>
<td>27</td>
</tr>
<tr>
<td>Ireland</td>
<td>25</td>
</tr>
<tr>
<td>Finland</td>
<td>24</td>
</tr>
<tr>
<td>Norway</td>
<td>24</td>
</tr>
<tr>
<td>Denmark</td>
<td>22</td>
</tr>
<tr>
<td>New Zealand</td>
<td>18</td>
</tr>
<tr>
<td>Slovenia</td>
<td>14</td>
</tr>
<tr>
<td>Latvia</td>
<td>11</td>
</tr>
<tr>
<td>Estonia</td>
<td>7</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>4</td>
</tr>
</tbody>
</table>

Source: OECD health care data, with Deloitte research data for missing countries.
In some countries, patients, insurers, or both pay specific prices for doctor’s visits. In other countries with state medical care, where the average patient pays nothing or very little, we assumed an average visit cost of what would be paid by a noncovered visitor for a 15-minute consultation with a general practitioner (GP). This cost number is almost certainly a minimum: Specialist visits would cost much more.

Based on a variety of sources for the 36 OECD countries, we believe the weighted average per-visit cost is about US$61, although it varies widely between countries. Multiplying each country’s number of visits by the cost per visit for that country yields a total of about $520 billion.

**METHODOLOGY 2: CALCULATING VISIT REVENUE AS A PERCENTAGE OF GDP**

Our second methodology for estimating the market size for doctor’s visits used the top-down approach of calculating doctor’s visit revenue as a percentage of GDP.

The OECD 36 members’ nominal GDP in 2018 was cumulatively US$53 trillion. In that same year, OECD members spent an average of 8.8% of their GDP on health care. However, expenditure varies by country. Applying each country’s health care expenditure percentage to its GDP yields a total health care spend of US$6.6 trillion across all 36 OECD countries.

OECD does have data for spending on general primary care (which excludes all hospital care as well as other primary care services such as dental, preventive, and home-based curative care), but only for 22 of the 36 countries, and only for 2016. For those countries and that year, general primary care (aka doctor’s visits) represented an average of 6.8% of total health care spending.

Assuming that the rate is approximately at that level across all 36 countries, general primary care spending totaled approximately US$450 billion for OECD members in 2016. If we include spending on specialist visits, it seems likely that total OECD spending on all doctor’s visits was more than US$500 billion, or more or less in line with the number produced using methodology 1.

**Applying each country’s health care expenditure percentage to its GDP yields a total health care spend of US$6.6 trillion across all 36 OECD countries.**

Annual patient visit data also is available for some non-OECD countries. For example, in Brazil, the average person went to the doctor 2.8 times per year in 2017; with a population of over 212 million, that means that Brazilian doctors had nearly 600 million visits. Even assuming a cost of only US$25 per visit, that would add another US$15 billion to the total. And although we lack data for very large markets such as Russia, China, and India, at even one or two annual visits per person in these countries, the market would be billions of visits larger. In sum, the global market for doctor’s visits could approach more than 12 billion visits per year, to the tune of US$700 billion.
THE BOTTOM LINE

Although we do not expect video visits to stay at pandemic levels, they will almost certainly not return to the prepandemic rate of about 1%. Evidence suggests that many caregivers agree: A summer 2020 webinar survey of US health care professionals found that only 5% anticipated virtual visits to return to prepandemic levels.26

One big reason that virtual visits are likely to persist is that patients like them. In a recent survey, nearly half of Americans said they prefer health care professionals who offer phone- or web conference-based consultations.27 Virtual visits tend to be more efficient, reducing visit time by 20%.28 They reduce the wait time for seeing a specialist (pre–COVID-19 studies show that median wait times for specialist consultations were down 50% in New York City and 75% in San Francisco following virtual consultations).29 They eliminate the need to travel to and from a doctor’s office. They are also seen as safer.30 Further, having video visits offer value beyond just convenience will likely enhance adoption over time.

That said, there are still some patients who do not see video visits as equal to physical ones. In an April 2020 survey, 66% of respondents believe that a doctor or nurse needs to physically examine them to understand their health needs, and 56% don’t think they get the same quality of care/value from a virtual visit as from an in-person visit.31

More broadly, although many patients, health professionals, insurers, and regulators already like video visits, driving adoption higher than pandemic levels will require buy-in from not just some, but many or all of these stakeholders. Health care providers and the health care ecosystem have considered video visits and other kinds of virtual health as a substitutive channel for in-person care delivery. “Next reality” transformation requires capabilities to position video visits as an integral channel for care management and as a way to drive reduction in the total cost of care.

Doctors and medical professionals are still learning how to optimize video technology and their own behavior for new models of care. As just one example, health care professionals should adapt their learning and training to go from a bedside manner to a “webside” manner. They also should look strongly at more proactive care with wearable and “nearables” (smart objects: everyday items with small, wireless computing devices attached to them) and more ubiquitous team-based solutions that also support caregivers.

Whether insurers and governments continue to reimburse for virtual video visits will matter a great deal in markets where insurers play a key role. Historically, many insurers have not paid for virtual visits at the same rate (or at all) as in-person visits. Once again, however, COVID-19 has prompted the situation to change. Many US insurers and the US Center for Medicare and Medicaid Services (CMS) relaxed rules around virtual visit reimbursement due to the pandemic. According to one US analysis, only 0.2% of the medical claims filed in March 2019 were telehealth-related; in March 2020, that number had risen to 7.5%.32 Two-thirds of health professionals surveyed in a pre–COVID-19 2020 survey said that “top accelerators [for virtual health adoption] included overcoming regulatory barriers such as licensing restrictions and restrictions on allowing for site-neutral payments, along with implementing payment methods that reward better health outcomes.”33
Investors and companies should expect higher levels of investment in areas related to virtual health, as well as merger and acquisition (M&A) activity. In the first two quarters of 2020, health innovation funding globally was up by 19% compared to the same period in 2019, reaching a new record of US$9.1 billion. And in August of 2020, two telemedicine industry leaders became one, as Teladoc spent US$18.5 billion to purchase Livongo. Although they won’t all be megadeals of that magnitude, more telemedicine M&A deals are probable.

The growth of video visits will likely have implications for other industries as well. The telecom industry, for instance, will have a large role to play in making virtual health care as widely available as possible. Although 90% of adult Americans and 73% of Americans over the age of 65 are connected to the internet, both numbers should be higher in order to make universal access to video visits possible. The numbers are roughly similar in other developed countries, but they are lower in developing countries and in rural areas globally.

In addition, sectors that produce technology that can be used for medical monitoring will likely benefit from video visits’ growth. For instance, although smart watch sales declined to US$25 billion in 2020, we anticipate that they will reach US$64 billion by 2024, in part due to their use in medical applications. This in turn will drive change in the health care industry: Wearables such as smart watches have the potential to reduce hospital costs by 16% over the next five years. But wearables also need to be used appropriately. As an example, the US Food and Drug Administration–approved Apple Watch wrist wearable device is a very useful tool for monitoring those who have been diagnosed with atrial fibrillation, but are much less useful at screening populations of presumed healthy people for the condition due to their high false positive rate.

No one today expects a doctor, black bag in hand, to make house calls. But thanks to video visits, it’s now possible for patients to receive medical care at home once again. While video visits may never completely replace in-person consultations, we expect that over time, for those visits where they are appropriate, they will become as ordinary and acceptable an option as going to a doctor’s office is today.
Endnotes

6. Daily Chronicle, “Astonishing growth in telehealth virtual visits market which is growing exponentially with the major key players in the market within the forecast period 2020-2027,” August 28, 2020.
9. Ibid.
15. Alex Schulte, Melissa Majerol, and Jessica Nadler, Narrowing the rural-urban health divide: Bringing virtual health to rural communities, Deloitte Insights, November 27, 2019.
20. “This indicator presents data on the number of consultations patients have with doctors in a given year. Consultations with doctors can take place in doctors’ offices or clinics, in hospital outpatient departments, or, in some cases, in patients’ own homes. Consultations with doctors refer to the number of contacts with physicians, both generalists and specialists. There are variations across countries in the coverage of different
types of consultations, notably in outpatient departments of hospitals. The data come from administrative sources or surveys, depending on the country. This indicator is measured per capita.” See: OECD, “Data: Doctors’ consultations,” accessed September 28, 2020.

21. The weighted average cost is weighted by the number of annual visits per country times the price per visit paid in that country.


32. Worthington, “The future of telehealth and telemedicine.”


36. GlobalData, “Wearable tech market set to grow 137% by 2024 but smartwatches to see a 10% decline in revenue this year due to shipment delays and tighter consumer wallets, says GlobalData,” August 13, 2020.


38. Mark Terry, “Apple Watch atrial fibrillation study has high rate of false positives,” Bio Space, March 18, 2019. Apple Watch is a trademark of Apple Inc., registered in the United States and other countries. *TMT Predictions 2021* is an independent publication and has not been authorized, sponsored, or otherwise approved by Apple Inc.
About the authors

Kenneth Abrams  |  kabrams@deloitte.com
Kenneth Abrams, MD, is Deloitte's chief medical officer and a managing director in Deloitte's Strategy practice. An anesthesiologist with more than 30 years of experience as a practicing physician and physician executive, Abrams is known as a thought leader in clinical strategy, operations and performance improvement, virtual health, and clinical integration.

Chris Arkenberg  |  carkenberg@deloitte.com
Chris Arkenberg is a research manager with Deloitte's Center for Technology, Media & Telecommunications. He has dedicated his career to exploring how people and organizations interact with transformational technologies.

Ariane Bucaille  |  abucaillle@deloitte.fr
Ariane Bucaille is Deloitte's global Technology, Media, & Telecommunications industry (TMT) leader and also leads the TMT practice and the TMT Audit practice in France. She has more than 20 years of experience and is a chartered and certified public accountant.

Cornelia Calugar-Pop  |  ccalugarpop@deloitte.co.uk
Cornelia Calugar-Pop is the lead Technology, Media & Telecommunications (TMT) researcher at Deloitte UK, where she conducts research and analyzes trends in the TMT space. Her responsibilities include production of regular analyses on topical issues, publication of thought leadership reports, research program management, and internal research support.

Allan Cook  |  allcook@deloitte.com
Allan Cook is the Digital Reality business leader for Deloitte Consulting LLP and the global Technology, Media & Telecommunications sector leader for operations transformation. He focuses on advising organizations on strategies and implementation plans around virtual reality, augmented reality, mixed reality, and 360-degree immersive experience.

Naima Hoque Essing  |  nhoqueessing@deloitte.com
Naima Hoque Essing is a research manager in the Deloitte Center for Technology, Media & Telecommunications, Deloitte Services LP. Her research focuses on the impact of emerging technology, business, and regulatory trends on industries and enterprises.
Bill Fera | bfera@deloitte.com
Bill Fera, MD, specializes in technology-enabled transformation to support the advancement of population health strategies. As a practicing physician, health system executive, and consultant, Fera has worked across health plans and health systems to drive toward a value-based, patient-centered model of care.

David Jarvis | davjarvis@deloitte.com
David Jarvis is a senior research manager with Deloitte's Center for Technology, Media & Telecommunications. He has more than 12 years of experience in the technology industry focusing on emerging business and technology issues and the potential impacts of longer-term change.

Patrick Jehu | pjehu@deloitte.com
Patrick Jehu is the cloud leader for Deloitte Consulting LLP's Technology, Media & Telecommunications industry practice as well as the Technology sector. He has steered large-scale digital and enterprise transformations, helping to enable enhanced market competitiveness for some of the world's most influential brands.

Dan Jones | danjones@deloitte.co.uk
Dan Jones leads Deloitte's work in sports around the world, advising high-profile clients on strategic, commercial, financial, regulatory, organizational, and structural matters. He is a regular speaker and expert commentator on sports business and edits Deloitte's Annual Review of Football Finance and the Football Money League.

Paul Lee | paullee@deloitte.co.uk
Paul Lee is a UK partner and the global head of research for the technology, media, and telecommunications (TMT) industry at Deloitte. In addition to running the TMT research team globally, Lee manages the industry research team for Deloitte UK.

Dan Littmann | dlittmann@deloitte.com
Dan Littmann specializes in helping clients within the telecommunications industry define growth strategies and accelerate the launch of new products and services. He has presented and led primary research efforts on the evolving digital ecosystem and its impact on wireless and wireline network providers.

Michael Liu | jliu@deloitte.com.cn
Michael Liu is a partner in Deloitte China, where he leads the Cloud and Technology Strategy/Transformation practice and serves as the North China Regional Consulting leader. He has more than 20 years of experience providing technology strategy, cloud, and digital transformation services.
Jeff Loucks   |   jloucks@deloitte.com
Jeff Loucks is the founder and executive director of Deloitte's Center for Technology, Media, and Telecommunications (the TMT Center). Using original research and deep industry expertise, Loucks and the TMT Center develop insights that reveal emerging trends so executives can anticipate change rather than reacting to it.

Sanket Nesargi   |   snesargi@deloitte.com
Sanket Nesargi is a managing director in Deloitte’s Technology, Media & Telecommunications (TMT) practice. He specializes in 5G and edge networks and their applicability to enable use cases across multiple industries.

Nobuo Okubo   |   nookubo@tohmatsu.co.jp
Nobuo Okubo is Deloitte’s global leader for Technology and a partner in Deloitte Japan. He has over 25 years of experience in technology and management consulting, focusing on the electronics industry.

Suhas Raviprakash   |   sraviprakash@deloitte.com
Suhas Raviprakash is a senior analyst at Deloitte Support Services India Pvt. Ltd. and part of the UK TMT Insight team. His research focuses on consumers’ digital behavior and the impact of COVID-19 on various verticals such as media production, 5G, e-bikes, gaming, and misinformation.

Duncan Stewart   |   dunstewart@deloitte.ca
Duncan Stewart is the director of research for the Technology, Media & Telecommunications (TMT) industry for Deloitte Canada. He presents regularly at conferences and to companies on marketing, technology, consumer trends, and the longer-term TMT outlook.

Anil Kumar Tarigoppula   |   antarigoppula@deloitte.com
Anil Kumar Tarigoppula is an assistant manager with Deloitte Support Services India Pvt. Ltd. and is part of the UK TMT Insight team. With more than 10 years of industry experience, his areas of expertise are survey data analysis, forecasting, and telecom infrastructure.

Kevin Westcott   |   kewestcott@deloitte.com
Kevin Westcott, vice chairman, is the leader of Deloitte’s US Technology, Media & Telecommunications (TMT) practice and the global Telecommunications, Media, and Entertainment (TME) practice. His industry experience spans film, television, home entertainment, broadcasting, over-the-top, publishing, licensing, and games.

Izzy Wray   |   iwray@deloitte.co.uk
Izzy Wray is a manager in Deloitte's Sports Business Group and leads its work in women's sports, with a particular focus on its commercial potential. She works on a broad range of strategic, commercial, financial, and governance projects for a variety of clients.
Contact us

Our insights can help you take advantage of change. If you’re looking for fresh ideas to address your challenges, we should talk.

Industry leadership

Ariane Bucaille
Global TMT industry leader | Partner | Deloitte France
abucaille@deloitte.fr

Ariane Bucaille is Deloitte’s global TMT leader and also serves as TMT leader for Deloitte France.

Kevin Westcott
Global TM&E sector leader | Partner | Deloitte LP
kewestcott@deloitte.com

Kevin Westcott serves as Deloitte’s global telecommunications, media and entertainment (TM&E) leader. In the United States, he is also a vice chairman and leads the TMT practice.

Nobuo Okubo
Global technology sector leader | Partner | Deloitte Japan
nookubo@tohmatsu.co.jp

Nobuo Okubo is Deloitte’s global technology sector leader and a partner with Deloitte Japan.
About the Deloitte Center for Technology, Media & Telecommunications

Deloitte's Center for Technology, Media & Telecommunications (TMT) conducts research and develops insights to help business leaders see their options more clearly. Beneath the surface of new technologies and trends, the center’s research will help executives simplify complex business issues and frame smart questions that can help companies compete—and win—both today and in the near future. The center can serve as a trusted adviser to help executives better discern risk and reward, capture opportunities, and solve tough challenges amid the rapidly evolving TMT landscape.

Connect
To learn more about the Center for Technology, Media & Telecommunications and to stay up to date on our latest research and insights, please visit www.deloitte.com/us/tmtcenter.

Subscribe
To receive TMT email communications, please subscribe at https://my.deloitte.com/subscriptions.html and select your areas of interest.

Engage
Follow us on Twitter at: @DeloitteTMT.