

Deloitte's TMT Predictions 2024

Three themes highlight this year's report: generative AI, sustainability, and monetization.

ARTICLE • 10 MINUTE READ

It's only been a year since generative AI surged to prominence as the power of AI to create images, videos, code, and text captured both public imagination and the attention of business leaders. We expect 2024 to be a transitional year: Tech companies are upgrading their software and services with generative AI, and companies across industries are trying to benefit from it. Some of the initial generative AI exuberance has faded, and a pressing question remains: What will happen in 2024? We look at generative AI from four key angles in this report: hardware, software, services, and regulation.

It's likely that 2023 will be the hottest year in recorded history, and some climate scientists are convinced that the rate of warming is accelerating.¹ Human reliance on tech devices, services, and connections is accelerating, too. How can tech use fewer resources to produce and operate? How can tech help make other industries more efficient? One answer is improved sustainability. We look at sustainability in three ways: from an industry perspective (its value in semiconductor, telecom, and agricultural technology), a technology perspective (its ability to track sustainability goals), and a self-sufficiency perspective (its ability to deal with key material shortages through recycling e-waste).

We've grouped our 19 topics across four categories this year: generative AI; sustainability; media, entertainment, and sports; and telecom and technology.

Across all four sections, we have examples of technologies and content that have gathered user bases and audiences but may be struggling to make profits or even revenues. Streaming providers may need to shift from market share to profitability. Women's elite sports (with a struggle that may be ending) are on track to surpass a billion dollars in revenues. Software companies embedding AI face the challenge of recouping costs; podcasters seek to increase revenue per user; while device makers and service providers work to determine how to charge for satellite communication features.

It's an exciting time to be making predictions. It may be a cliché to say that things are changing faster than ever before, but the emergence of multiple trends that could help humans think better and faster, grapple with climate change, and convert billions of users to billions (or tens of billions) of dollars in revenues and profits could make 2024 TMT's most exhilarating year yet.

Here is an overview of each prediction across each category.

Generative AI

Everyone's talking about generative AI, but what will companies do with it? And how will they do it? The answers could indicate how big the generative AI market might get.

Generative AI and enterprise software: What's the revenue uplift potential?

Gen AI is coming to enterprise software, but expect competition between vendors who want to charge per user and IT departments that believe generative AI features should be free.

We predict almost all enterprise software companies will embed generative AI in at least some of their products this year. There will likely be a mix of pricing models: explicit per-user-per-month pricing, consumption-based pricing, a hybrid approach, and implicit pricing or free, at least for now. We predict that the revenue uplift for enterprise software companies will be at a US\$10 billion run rate by the end of 2024. That is lower than some estimates, but still notable for the first year of a new market.

Gen AI chip demand fans a semi tailwind...for now

The generative AI chip market is growing fast, and together with other AI-enabling chips could represent half of the value of all semiconductors sold by 2027.

Deloitte predicts that the market for specialized chips optimized for generative AI will be valued at over US\$50 billion in 2024, up from close to nothing in 2022. But there are some who fear a generative AI chip bubble, especially if enterprise use cases fail to materialize. Nonetheless, companies will likely need AI chips, including generative AI chips, as they often regard secure and reliable supply chains to be important for innovation, economic success, and national security. Generative AI chip revenue could reach US\$400 billion in 2027, and along with other AI-enabling chips, it could represent half of the value of all semiconductors sold, but more supply and new entrants make that unlikely.

Taking control: Generative AI trains on private, enterprise data

More companies, seeking to avoid the risk of models trained on public data, are expected to train generative AI on their own data to enhance productivity, optimize costs, and unlock complex insights.

While generative AI models trained on public data have grabbed the headlines, their capabilities—and challenges—are stoking demand for more customized and constrained models trained on private enterprise data. In 2024, Deloitte predicts that enterprise spending on generative AI will grow by 30%. More companies are expected to develop their own generative AI models to drive greater productivity, optimize costs, and unlock novel insights and innovations.

Walking the tightrope: As generative AI meets EU regulation, pragmatism is likely

2024 is likely to see a balance between regulatory compliance and fostering innovation in generative AI; clear regulation enables enterprises and vendors to proceed with certainty.

Well-crafted rules can help unlock the potential of any market. In the case of generative AI, the existence of clear regulatory conditions could give confidence to vendors, enterprise customers, and end-users to invest in the technology. The European Union is expected to set the stage for global regulation of generative AI in 2024, influencing its own markets and serving as a template for other regions. In 2024, two regulations are expected to help shape generative AI: the General Data Protection Regulation (GDPR)² and the European Union's upcoming AI Act.³ These regulations concern various issues, including individual consent, rectification, erasure, bias mitigation, and copyright usage.

Sustainability

As more tech devices and faster connections are produced, companies should consider reducing the energy and raw materials required. This requires innovation and commitment. And tech and connectivity can help solve huge problems, like how to feed the growing population with less water and energy.

A raw deal: Will materials shortages, supply chain challenges threaten tech's future?

In the face of rising trade issues and skyrocketing demand, e-waste recycling and digital supply networks and a holistic approach to supply chain sustainability could help in the medium term.

Deloitte predicts that multiple regions will run short of gallium and possibly germanium in 2024 and may start seeing shortages of rare earth elements (REEs) by 2025. The impending shortage across dozens of different REEs at the same time could affect various industries. Enabled in part by these tech raw materials, those industries are estimated to be worth over US\$160 billion annually in market value. If trade restrictions between China and the West escalate further, the tech and chip industry could consider bolstering supply chain resilience by sharpening their focus and increasing investments in e-waste recycling, digital supply networks, stockpiling, and sustainable semiconductor manufacturing.

Semiconductor sustainability: Chips take a smaller byte out of resources

Modern, new greenfield plants could help improve the industry scorecard, but manufacturing transformation can help both the greenfield plants and existing brownfield plants do better for energy, water, and process gas use.

Consumption of energy, water, and process gases go up and down in tandem with the rise and fall in chip manufacturing activity and revenue. Although absolute use of energy, water, and process gases are expected to grow as semi revenues grow in 2024, Deloitte predicts year-over-year declines in average water intensity and energy intensity, as well as continued shift in renewable energy mix, across all leading chipmakers. From a sustainability perspective, manufacturing transformation can help both greenfield plants and existing brownfield plants improve the chip industry's scorecard.

Dialing down the carbon: Telco sustainability surges on the back of four new trends

Telecommunications companies can reduce their carbon footprint by shutting down copper wire and 3G wireless networks, changing their field service fleets to EVs, and switching to 5G radio gear that has a "snooze button."

Telecom is not the worst carbon culprit, connecting 95% of the planet and being responsible for only 2% of emissions,⁴ but more could be done. Moving from copper wire networks to fiber optics can cut energy use by 80% or more.⁵ Older wireless networks like 3G use a lot of power while serving few customers.⁶ They're being shut down too. Electrifying field service fleets is accelerating.⁷ 5G radios can use a lot of power, but new technologies let them go to sleep when not in use, saving telcos gigawatt hours of electricity and megatons of CO2 equivalent annually.⁸

Regulations take effect: ESG reporting software sales are expected to soar in 2024

Pushed by investors, regulators, and employees, many more companies will likely systematize their ESG tracking and reporting with standardized software tools.

Sales of software solutions that help companies track and report on environmental, social, and governance (ESG) metrics will likely surpass US\$1 billion this year as EU and US reporting regulations—along with reporting requirements in Asia, Australia, and the United Kingdom—take effect and as more investments require ESG disclosures. Forecasts for ESG-reporting software estimate a compound annual growth rate of 19% to 30% over the next five years. Deloitte predicts the tipping point will be reached in 2024, with growth accelerating to over 30% and revenues climbing from under an estimated US\$800 million in 2023 to just over US\$1 billion in 2024.

On solid ground: AgTech is driving sustainable farming and is expected to harvest US\$18 billion in 2024 revenues

Agriculture technology is designed to let the producers and farmers grow more food using less pesticides, energy, water, and resources, enhancing farm yields.

Deloitte predicts the agricultural technology revenue opportunity—including Internet of Things endpoints and connectivity devices for precision crop farming, livestock management, and agricultural equipment tracking—will be US\$18 billion globally in 2024, with a 19% CAGR between 2020 and 2024. While there's a need to address the looming food crisis and agricultural production-driven climate issues, scaling food production using current methods is resource-intensive, costly, and less efficient. AgTech solutions could help by enabling farmers around the world to grow more food with less environmental impact, enhance crop yields and harvests, and manage livestock better.

Media, entertainment, and sports

People want to stay entertained, especially amid all the uncertainty in the world. But many media companies are dealing with their own uncertainties—especially how to make money as pay TV wanes, streaming gains users (but loses money), and gaming changes the entertainment landscape.

Women's elite sports: Breaking the billion-dollar barrier

Interest from fans, broadcasters, and commercial partners is driving rapid growth in the revenues of women's elite sport. How can organizations continue to accelerate growth?

Women's elite sports are expected to break through the billion-dollar revenue tape for the first time in 2024, propelled by new sponsorship deals, surging TV rights (including streaming rights), and record-breaking audiences. Revenues are expected to be three times higher than in 2021 when we last reviewed this market. Since then, momentum behind women's sports has continued to grow. The leading sports are likely to be football (soccer) and basketball, with cricket, volleyball, and rugby continuing to grow at an impressive pace. As more investment goes into women's elite sports, athletic performance and standards will likely improve, creating a virtuous circle to help supercharge revenues.

Driven to tiers: Streaming video services look to up their profitability game with viewers

In 2024, streamers are expected to charge more for premium content, fight churn with longer subscriptions, and satisfy bargain hunters with more pricing tiers.

More than a decade into the streaming video revolution, media and entertainment (M&E) companies seem to be realizing how hard it is to recoup the historic profits of the pay TV business model. In 2024, Deloitte predicts that the combined number of subscription video on demand tiers offered by the top US providers will more than double between 2022 and 2024: from an average of four options to eight. From cheap ad-supported offerings and gated content to premium tiers with instant access, streamers are expected to shift from growth at all costs to making it easier for all their subscribers to get enough value for the price. Viewers may find it harder to wade through the options, but tiering could help them get more of what they want, and less of what they don't.

Cinematic and interactive universes: Games and studios come together to bring the biggest stories to life

Shifting demographics, innovative technologies, and the desire for high-performing IP are driving the evolution of storytelling.

In 2023, some of the largest and most successful film and TV franchises began as video games, and some of the biggest video games began as film and TV shows. Now, Hollywood is looking to games for new IP that they can expand and monetize, and game companies are eyeing TV and film collaborations to help make their IP work harder and offset soaring game development costs. Whether they originate in games, TV, film, or elsewhere, some of the most compelling stories are expanding across media to engage fans, reach more audiences, and lift the broader value of their franchises.

Shuffle, subscribe, stream: Consumer audio market is expected to amass listeners in 2024, but revenues could remain modest

More consumers globally are tuning in to audio entertainment formats like podcasts, streaming audio services, audiobooks, and radio.

Deloitte predicts that more consumers worldwide will engage with audio entertainment—such as podcasts, streaming music, radio, and audiobooks—in 2024. We also predict that the global audio entertainment market will surpass US\$75 billion in revenue in 2024, up 7% across these four formats. There are opportunities for growth across audio formats, especially for podcasts, whose annual revenues per user lag the other three categories. M&E companies have opportunities in audio, including expanding global listenership, increasing ad revenues, and integrating generative AI capabilities.

Will endless low cost content do to gaming what it did to TV and film?

UGC gaming platforms that incentivize user-generated content are expected to make big payouts to creators in 2024. As this practice grows, however, it could put pressure on top-tier games and services.

Two of the most popular video games have enabled hundreds of millions of players to be 3D creators. A third heavyweight has launched new tools and incentives to stoke the fires of 3D user-generated content (3D UGC) on their own platform.⁹ Deloitte predicts that these platforms will pay out almost US\$1.5 billion to their content developers in 2024, and the number of paid independent developers on 3D UGC gaming platforms will exceed 10 million. Expanding the creator economy for games could drive greater innovation in game experiences and digital goods, stronger engagement for 3D UGC platforms, and more favorable economics for both platforms and creators. But by unleashing endless cheap 3D content, could they disrupt the entire industry?

Telecom and technology

Smartphone authentication: The killer app that can augment the smartphone's utility

Smartphones can help you securely log in, tap to pay and enter cars, buildings, and airports. As security concerns grow, the need for authentication will likely consolidate the smartphone's status as the ultimate goldilocks device: the right size, power, connectivity and trust.

In 2024, the smartphone is expected to further consolidate its status as the most successful consumer device, as it's used increasingly to prove identity, utilizing built-in fast and secure biometric authentication capabilities. Smartphones may be used trillions of times in 2024 to authenticate across an ever-widening range of actions: accessing websites, making payments online and in-store, unlocking cars, and controlling entry to physical buildings. Over the coming years, the volume of transactions authenticated by smartphone may grow to hundreds of trillions of usages per year, with smartphones usurping physical keys and passwords alike.

Signals from space: Direct-to-device satellite phone connectivity boosts coverage

Integrating satellite and terrestrial mobile networks could unlock new revenue for the satellite, semiconductor, and telecom industries.

A growing ecosystem of satellite and mobile network operators, handset manufacturers, and semiconductor companies are racing to provide global coverage for smartphones and IoT devices—without the use of specialized equipment. They are looking to connect the unconnected, improve safety and emergency responses, and expand IoT applications. Basic services for emergency communication, simple text messages, and IoT monitoring have already started. Deloitte predicts that over 200 million smartphones that can connect with satellite services will be sold in 2024. These phones are expected to contain about US\$2 billion of special chips. The question is: How will this ecosystem drive adoption?

No bump to bitrates for digital apps in the near term: Is a period of enough fixed broadband connectivity approaching?

In some parts of the world, some consumers may have all the bandwidth they need in 2024. They are still expected to want ever better internet connections, just not necessarily faster ones.

The demand for ever higher internet speeds over fixed networks has been a constant for most of the internet's history. However, this trend may be coming to an end. Deloitte predicts that, in 2024, over 90% of the most used online applications on fixed broadband networks in developed markets will have the same vendor-recommended bitrate as in 2023. We further predict that at least 80% of these applications may see no increase to advised bitrate for the period between 2023 and 2025. Over time, the bitrate for some of the most popular applications might reduce due to better compression.

Keeping it local: Cloud sovereignty a major focus of the future

More data, increased cybersecurity threats, and geopolitical tensions are expected to increase demand for cloud solutions that can operate locally. Meeting this need can protect a company's reputation, operations, and bottom line.

Government cloud (designed for stringent compliance requirements) is expected to surpass US\$41 billion, and distributed cloud (a solution for data residency) could surpass US\$7 billion. Data regulation is a complex global tapestry, and it can change frequently. What's more, localized cloud often comes with nuances, including cost, complexity, a smaller range of applications, lack of scalability, and vendor lock-ins. Companies should be agile and develop operational resilience to help protect their reputation and bottom line.

Life after debt: Venture debt funding could grow again in 2024

After a rough 2023, tech companies should expect to see a rebound in technology venture debt funding, fueling innovation.

After four years of US\$30+ billion in US venture debt activity, followed by a steep plunge in 2023 to US\$12 billion, US tech venture debt could see a modest rise to US\$14 billion to US\$16 billion, and continue growing in the future. In the short term, venture debt deals are likely to become smaller. However, we expect to see strategic venture capital deals rise in 2024, as cash-rich, mega-cap tech companies could invest or buy out smaller companies that can't raise capital or debt. These developments in the market are likely to encourage early-stage tech companies to build more stable growth over time.

By **Ariane Bucaille**
France

Kevin Westcott
United States

Gillian Crossan
United States

Endnotes

1. Oliver Milman, “*Global heating is accelerating, warns scientist who sounded climate alarm in the 80s,*” *Guardian*, November 2, 2023.
2. EUR-Lex, “*Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation),*” accessed November 16, 2023.
3. News European Parliament, “*EU AI Act: First regulation on artificial intelligence,*” June 14, 2023.
4. Iain Morris, “*The carbon-belching ICT sector must do better on the environment,*” Light Reading, September 2, 2022.
5. Grace Langham, “*Decommissioning legacy networks will be key to reducing operators’ energy usage,*” Analysys Mason, August 17, 2022.
6. ABI Research, “*Purchasing renewable energy removes CO2 emissions equal to 20 million barrels of oil a year for leading telco operators,*” April 21, 2022.
7. Yvonne Rene de Cotret, “*Electrified fleets pave the way to emissions reduction,*” Deloitte Canada, February 15, 2023.
8. Ericsson, “*Rogers tackles network energy consumption,*” accessed October 11, 2023.
9. Jay Peters, “*Epic now rewards Fortnite creators for how long people play their experiences,*” Verge, June 1, 2023.

Acknowledgements

Cover image by: **Manya Kuzemchenko**

Generative AI and enterprise software: What's the revenue uplift potential?

Gen AI is coming to enterprise software, but expect competition between vendors who want to charge per user and IT departments that believe generative AI features should be free.

ARTICLE • 7 MINUTE READ

At the start of 2023, excitement about generative AI prompted three big questions for enterprises: Will gen AI be embedded in enterprise software? How will vendors charge extra for gen AI tools inside software? And, how big an uplift in revenues will gen AI create for the enterprise software industry?

As we enter 2024, Deloitte has three projections: First, we predict almost every enterprise software company will embed gen AI in at least some of their products. Second, we predict there will be a mix of pricing models: explicit per-seat pricing (per user per month or PUPM), consumption-based pricing, a hybrid approach, implicit pricing (retaining whatever model they have today but charging more), or free, at least for now. Finally, we predict that the revenue uplift for enterprise software companies (in addition to the cloud providers of gen AI processing capacity) will be approaching a US\$10 billion run rate by the end of 2024.

That's lower than the US\$14 trillion (not a typo) that one fund manager has predicted for gen AI software by 2030,¹ but still significant, even though it's a fraction of the US\$1.6 trillion in global enterprise IT spending projected for 2024,² and also smaller than the expected hardware uplift for chips and servers that perform gen AI of more than US\$50 billion in 2024.³ Given the excitement about gen AI enterprise software tools in 2023, how did we arrive at this number, and why isn't it higher?

The market potential for 2025 and beyond looks robust, and the revenue uplift for enterprise software companies will likely be tens of billions of dollars. But 2024 is effectively a transition year. The various kinds of enterprise software tools that are expected to include gen AI are not launching until late 2023 or early 2024. Some companies have cautioned analysts that adoption and revenues in the second half of the year will be much stronger than in the first half.⁴ More than 70% of companies are experimenting with gen AI, but less than 20% are willing to spend more on it.⁵ Putting that all together for 2024, we predict a revenue uplift approaching a \$10 billion run rate by the end of 2024.

We predict that for most enterprises, the "gateway to gen AI" will likely be through the three categories outlined below wherein gen AI features are embedded in existing software, often going unnoticed by users.

Broad enterprise productivity software suites: There were predicted to be 1.14 billion knowledge workers worldwide in 2023,⁶ which suggests a total addressable market of almost US\$400 billion annually, assuming every knowledge worker needs at least one gen AI-powered set of enterprise software tools and that they pay US\$30 PUPM.

Enterprise software tools: There are various software tools such as database and analysis solutions, enterprise resource planning (ERP) solutions, customer relationship management (CRM) solutions, creative and document management solutions, and more. As of the time of writing, Deloitte analysis of publicly available announcements suggests that 100% of the 50 largest enterprise software companies are planning to offer a version of their software that has gen AI features, with some using PUPM pricing models, usage pricing, or offering for free, at least for now.⁷

Engineering, design, and software development tools: There are a number of new tools in which gen AI is not merely an enhancement but is, in fact, core to the new product. Multiple chip design companies are offering gen AI-enhanced versions that design chips (one of our 2023 TMT Predictions),⁸ perform functional verification, and test silicon.⁹ In the computer-aided design space, many players offer gen AI features.¹⁰ And some software development tools charge around US\$10 PUPM.¹¹

Gen AI is expensive to operate but its value to customers may not yet be clear

At one level, it makes sense that software companies want to charge for gen AI. In addition to growing revenues, offering gen AI is expensive. They are often spending billions of dollars either buying the chips that are needed for gen AI training or needing to buy instances from the cloud or chip companies. As an example, looking at some of the largest cloud players, they are expected to spend from 3%–13% of their 2023 capex on AI.¹² Further, the operating costs are non-trivial, with estimates that each gen AI query costs from US\$0.01 to US\$0.36 per query. As an example, one service that costs US\$10 PUPM is rumored to be losing US\$20 monthly, with some users costing the provider more than \$80.¹³ Both chip pricing and operating costs are expected to come down sharply over time but likely not until the current gen AI chip shortage eases, which Deloitte predicts will not occur sooner than the second half of 2024 (see [gen AI chips prediction](#).)

Meanwhile, at least some buyers of enterprise software are pushing back. To quote a June 2023 US survey of buyers, although the long-term potential of gen AI features was very high, some respondents said that AI features were “table stakes” (that is, every vendor needed to offer it) but “good luck trying to get me to pay for it.”¹⁴ Other customers may not yet see the value of gen AI in their workflows: They may not want to pay to use it but might take free trials that later shift to paid services.

The bottom line

At a high level, companies likely won't pay for gen AI-enhanced tools unless they generate a positive ROI. A research study from the fall of 2023 suggests that ROI might be very strong, with knowledge workers using gen AI (direct access, not through embedded AI inside enterprise software) to do more, faster, and at a higher quality than those not using the tools.¹⁵ If improvements of those magnitudes are seen in the real world and across multiple industries, it seems likely that the uplift in revenue over the longer term could be much larger than the early innings of 2024 might suggest.

However, if ROI gains are lower (or take time to be demonstrated) vendors are likely to see slow adoption or buyer pushback on pricing. One potential alternative between high per-seat-per-month pricing and free “table stakes” gen AI could be a hybrid model: a relatively low per-seat month price (<US\$10) but combined with a consumption charge (aka usage-based pricing), which would allow vendors to recapture some of their per-request operating costs.¹⁶ Effectively, the more you use it, the more you pay.

Regulation and concerns around privacy, IP ownership, accuracy/confabulation, and more could be roadblocks. Any one of these could be enough to slow or even halt the adoption of gen AI-enhanced enterprise software solutions. Some proposed EU rules are so restrictive that many current-generation AI software tools might not be allowed in that market. (See Paul Lee’s [2024 Prediction on regulation](#).) One potential solution for some of those barriers could be for companies to build their own models, then train and run them on gen AI cloud services. This is likely to be a multibillion-dollar market for processing, software, and services over time. (See Chris Arkenberg’s [2024 Prediction on Private LLMs](#).)

Currently, leading gen AI accelerator chips are in shortage and on allocation (see Duncan’s [2024 Prediction on gen AI chips](#)), and this may be complicating the ability of companies to meet demand for gen AI-inside features. They need thousands or even tens of thousands of these US\$40,000 chips to meet anticipated demand, and some companies cannot get enough (or any) gen AI via the cloud.¹⁷ Capacity is expected to ramp into H1 2024, but in the first part of that year, gen AI-inside software revenues could be capacity-constrained.¹⁸ That chip scarcity may allow enterprise software companies to charge higher prices for gen AI features, as users of the software likely cannot build their own gen AI solutions due to chip scarcity.

But it may then become likely that gen AI accelerator chip prices will fall in the next 18–24 months. That could happen abruptly as supply comes online and new entrants emerge. Multiple players are expected to announce new data center and edge processing gen AI chips,¹⁹ and although it is unclear what market share these chips could garner, alternative chips are likely to reduce the current hardware shortages and high prices.

Some larger companies that have the capacity to build their own solutions either on top of hardware they buy, or through cloud gen AI capacity, may be deferring building/buying gen AI capacity until the prices come down. They are seemingly happy to be fast followers rather than paying up to be at the leading edge. Relatively few companies are expected to buy the hardware—many will get gen AI from the bigger cloud providers. Further, companies that are planning to use the hybrid model may see customer pushback on usage-based pricing if costs fall sharply, and may need to reduce that pricing.

A gen AI revenue uplift of up to US\$10 billion is a significant positive for vendors. But it should be placed in context: Global spending on cloud services is large and still growing, but that growth will likely slow. Public cloud was a US\$546 billion industry in 2022, up 22% from the year before,²⁰ but growth was down to 16% in Q2 of 2023,²¹ and although major cloud players are launching gen AI as a service, it’s unclear from publicly reported numbers just how much money they are making from those services in 2023 or will in 2024. Will gen AI revenues for cloud companies be big enough and fast enough to re-accelerate overall growth to the 20% annual range?

Another question is whether companies will pay for more than one gen AI-inside enterprise software per employee. Most knowledge workers use multiple software tools, and at US\$10–30 per month, cumulative “gen AI stacking” spending could be more than US\$100 monthly. Could we see a new class of software that sits on top of everything else and does gen AI on everything ... a universal co-pilot?

Ultimately, generative AI features have to show efficacy for a number of tasks very quickly and are contributing to enormous amounts of spending and strategic planning from large companies. Providers are confronting their own costs while figuring out pricing for customers, both for pure-play gen AI processing via the cloud and gen AI-enabled service features. End users for their part may take some time to determine how these capabilities translate into direct value for their businesses, but likely will soon realize gains and develop a better sense of its value and how much they’re willing to pay for it. From what we’ve seen so far, this could raise prices across the board.

By **Duncan Stewart**
Canada

Bariş Sarer
United States

Gillian Crossan
United States

Jeff Loucks
United States

Endnotes

1. Trevor Jennewine, “*Cathie Wood says artificial intelligence (AI) software may be a \$14 trillion market: 2 superb growth stocks to buy now and hold through the boom*”, *The Motley Fool*, September 10, 2023.
2. Gartner, “*Gartner says more than half of enterprise IT spending in key market segments will shift to the cloud by 2025*”, press release, February 9, 2022.
3. Duncan Stewart, Christie Simons, Brandon Kulik, Gillian Crossan, *Gen AI chip demand fans a semi tailwind ... for now*, Deloitte Insights, November 2023.
4. Deloitte analysis of quarterly earnings releases from public enterprise software companies and analyst reports in September and October 2023.
5. Carl Franzen, “*More than 70% of companies are experimenting with generative AI, but few are willing to commit more spending*”, VentureBeat, July 25, 2023.
6. Gartner, “*Gartner says worldwide social software and collaboration revenue to nearly double by 2023*”, press release, September 24, 2019.
7. Deloitte analysis of company announcements from June 2023 to October 2023.
8. Jeff Loucks, Duncan Stewart, Christie Simons, and Brandon Kulik, *AI in chip design: Semiconductor companies are using AI to design better chips faster, cheaper, and more efficiently*, Deloitte Insights, November 30, 2022.
9. Anton Shilov, “*Synopsys intros AI-powered EDA suite to accelerate chip design and cut costs*”, AnandTech, March 30, 2023.
10. Kevin, “*The role of artificial intelligence (AI) in the CAD industry*”, Scan2CAD blog, March 22, 2023.
11. Loraine Lawson, “*GitLab all in on AI: CEO predicts increased demand for coders*”, The NewStack, June 9, 2023.
12. Counterpoint, “*AI drives cloud player capex amid cautious overall spend*”, press release, July 27, 2023.
13. Tom Dotan and Deepa Seetharaman, “*Big Tech Struggles to Turn AI Hype Into Profits*”, *Wall Street Journal*, October 9, 2023.
14. Karl Keirsted, et. al., “Ears to the Ground – Unvarnished Feedback on GenAI Adoption and Trends from Large Enterprises through AI Start-Ups”, UBS Global Research and Evidence Lab, June 7, 2023.
15. Fabrizio Dell'Acqua, Edward McFowland, Ethan R. Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayner, François Candelon, and Karim R. Lakhani, “*Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality*,” SSRN, September 18, 2023.

16. Puneet Gupta, “*How any SaaS company can monetize generative AI*”, Tech Crunch, August 21, 2023.
17. Erin Griffith, “*The desperate hunt for the A.I. boom’s most indispensable prize*”, *New York Times*, August 16, 2023.
18. Dashveenjit Kaur, “*The genAI explosion is driving the chip industry up*”, Techwire Asia, September 11, 2023.
19. Kyle Wiggers and Devin Coldewey, “*This week in AI: The generative AI boom drives demand for custom chips*”, TechCrunch, September 11, 2023.
20. Leigh McGowran, “*Public cloud services revenue surged past \$500bn last year*”, Silicon Republic, July 7, 2023.
21. Canalys, “*Global cloud services market growth slows to 16% in Q2 2023*”, press release, August 10, 2023.

Acknowledgements

The authors would like to thank **Rohan Gupta, Chris Arkenberg, David Jarvis, Ankit Dhameja, and Karthik Ramachandran.**

Cover image by: **Manya Kuzemchenko**

Gen AI chip demand fans a semi tailwind ... for now

The generative AI chip market is growing fast, and together with other AI-enabling chips could represent half of the value of all semiconductors sold by 2027.

ARTICLE • 8 MINUTE READ

Deloitte predicts that the market for specialized chips optimized for generative AI will be over US\$50 billion in 2024. From close to zero in 2022, that sum is expected to make up two thirds of all AI chip sales in the year. Deloitte predicts that total AI chip sales in 2024 will be 11% of the predicted global chip market of US\$576 billion.¹ Recent AI chip market forecasts in 2027 range from an aggressive US\$400 billion to a more conservative US\$110 billion. For several reasons, the more conservative estimates may be more realistic.²

On the other hand, there are those who fear a generative AI chip bubble: Sales will be massive in 2023 and 2024, but actual enterprise generative AI use cases could fail to materialize, and in 2025, AI chip demand could collapse, similar to what we've seen in crypto mining chips in 2018 and 2021.³

Even at the lower end of the nonbubble range, AI chips would be a very large part of the semiconductor market, and a needed tailwind given anticipated sluggish demand from traditional stalwarts such as smartphones⁴ and PCs⁵ and more mature data center chips.

Understanding AI chip demand

At a high level, generative AI is similar to many other kinds of AI used in recent years: It's a form of machine learning, specifically deep learning combined with neural networks. But there are important differences, and the major chip companies as well as others have built or are building chips that are specially optimized for generative AI, as older AI chips are too slow or too inefficient to do it well and are lacking the right kind of design and memory.⁶

In spring 2023, the more advanced special chips were selling for about US\$40,000 each.⁷ There was strong demand for a million or more chips. Coveted chips were in severe manufacturing shortage and allocation (mainly due to a lack of advanced packaging capacity), and were a chokepoint for the rollout of generative AI offerings from thousands of companies.⁸ Many of the companies that make these chips can't make them fast enough, and that imbalance is expected to continue well into 2024.⁹ Demand is high, supply is constrained, and prices are high.

Perhaps even more important are the geopolitical implications: The special generative AI chips require a host of advanced technologies from all over the world, although they are fabricated mostly in Asia now and are likely to be highly concentrated there in the future. These chips are increasingly subject to trade restrictions from the US, Europe, and their Asian allies for China and Russia.¹⁰ Although China can develop its own generative AI data sets and software, it may be more difficult for the country to buy or make the most advanced chips needed for cutting-edge AI processing in the next five years. It's also unclear to what extent China may have been able to advance its chip production in spite of restrictions. For example, in September of 2023, a Chinese chip manufacturer produced a chip (for smartphones) that was made on 7nm process node. The chip was both smaller and two to three generations behind the processes used by leading generative AI chips, but it was closer than most Western analysts believed possible, given Western sanctions.¹¹

Getting technical

At the heart of the current state of the art generative AI hardware is a rack-scale board made up of different kinds of chips and interconnections.

Those boards are a mix of central processing units combined with a specialized (very large and very advanced process node) graphics processing unit (GPUs) that is in a special kind of packaging with a special kind of high speed memory.¹² In one example, the GPU is a piece of silicon called a die of over 800 mm² (which is very large), consisting of 80 billion transistors, packaged with a very large, very fast, high bandwidth memory (HBM3) in what is called 2.5D advanced packaging.¹³ This can be done either at the end of the foundry process or at the start of the back-end assembly and test process by an outsourced assembly and test player.¹⁴

Inside the data centers where most of these generative AI accelerators will be located, it's often necessary to move large chunks of data over short distances as rapidly as possible, using special networking chips.¹⁵ These communications chips are not used solely for generative AI applications, but generative AI is one of the biggest drivers for their use at present, and likely worth single digit billions of dollars in 2024.¹⁶

Finally, generative AI chips use a lot of power, approximately 10 kilowatts per board, and multiple units would produce far more heat than air cooling could cope with, so the market for liquid cooling is likely to be US\$2 to 3 billion in spending in 2024, growing at about 25% annually.¹⁷ Those high power draws also could require new high voltage power supplies.¹⁸ Using higher voltages could offer significant efficiency gains and is likely a sub billion dollar annual market across a number of smaller players.¹⁹

The bottom line

Deloitte is relatively confident in our prediction for the 2024 generative AI-driven market opportunity of about US\$50 billion, but what happens after the current high demand and high prices are met by higher supply and new entrants is unclear.

The 2027 numbers mentioned earlier (up to US\$400 billion) are potentially important to the global semiconductor industry, and from reputable sources, but there are several reasons why they might be too optimistic.

First, the summer 2023 market for generative AI GPUs is marked by having essentially a single designer, who in turn relies on a single supplier that is capacity constrained.²⁰ Meanwhile, buyers are trying to secure as many chips as they can to build processing capacity for anticipated consumer and enterprise use of generative AI training and inference.²¹ As a result, pricing could be roughly as high as it might ever be. As that supplier builds more capacity, or as new competitors enter the market, prices are more likely to decline than to stay where they are, thus impacting revenues for 2025 and beyond.

Second, when chip customers are on allocation and are unlikely to receive complete orders, they often over order. Knowing that orders are being cut back by 75%, for instance, though they'd need 25,000 chips, some might ask for 100,000 chips, rather than the 25,000 which is their "true demand." Once supply and AI chip demand are more in balance, buyers may get more chips than they need and then pull back just as new capacity is coming online. This is part of the semi "bullwhip effect" and a contributor to the extreme cyclicality the chip industry has historically seen.²²

Third, currently all training and almost all generative AI inferences are done using the same data center generative AI chips. But it's likely that over time a significant part of the generative AI inference will be done on edge processors.²³ These could be smaller GPUs or CPUs or new application-specific integrated circuits and could come from existing generative AI chip companies, or new entrants, including both traditional edge processing chip companies, but also companies not traditionally known for designing chips.²⁴ Doing more at the edge could expand the market, or cause prices of data center generative AI chips to fall.

Finally, as mentioned earlier, there are those who worry about a generative AI chip bubble, with a robust 2023 and 2024 and a weaker 2025: This view is not consensus, but worth being mindful of, given the possibility of boom and bust.

While it's hard to say with certainty, it may be probable that a combination of a higher and more diversified supply, lower than predicted AI chip demand, moving inference to edge processors, and lower prices could make the 2027 AI chip market closer to the lower end of the potential US\$110-400 billion range—still more than double 2024 levels.

Regardless of whether the market is US\$100 billion or US\$400 billion, companies are likely to need AI chips, especially generative AI chips, and to regard secure supply and reliable supply chains as critical for innovation, economic success, and national security.

And herein lies a challenge for the United States and Europe. Although multiple chipmakers are building advanced node plants capable of making cutting edge CPUs and GPUs for AI and generative AI,²⁵ there is not enough existing packaging capacity in Europe or the United States from either front end or back end companies.²⁶ Equally, there are no significant existing or planned HBM or HBM3e plants in either the United States or Europe.²⁷ Although the generative AI dies could be made domestically, they would likely need to be sent to Asia (Southeast Asia, South Korea, or Taiwan) for both the HBM3 memory *and* the advanced packaging portions of the process.

Both the European Chips Act and the US CHIPS and Science Act have money set aside for advanced packaging and advanced memory investments,²⁸ but it is unclear if it will be enough for each region to become self-sufficient in packaging generative AI chips.

The final implication of the growth in generative AI chips is China. At present, the United States, Netherlands, and Japan all have export controls in effect that prevent China from purchasing advanced node chips of all types, including generative AI chips, as well as the know-how to do so.²⁹ Amid concerns that future export controls may target less advanced chips,³⁰ leading Chinese internet companies ordered US\$5 billion of generative AI chips in August 2023, ahead of further potential US restrictions.³¹

If generative AI will be as important to innovation, economic growth, and national security in 2027 as it appears to be today, and if China is restricted from purchasing advanced AI chips and the tools needed to build their own advanced chips, it could have further impact on global economics, including the potential for raw material export restrictions of needed elements ([see raw materials prediction](#)) and other negative effects that could impact global growth.

By **Duncan Stewart**
Canada

Brandon Kulik
United States

Christie Simons
United States

Gillian Crossan
United States

Endnotes

1. World Semiconductor Trade Statistics, “[WSTS Semiconductor Market Forecast Spring 2023](#),” May 2023.
2. Analysis based on data sourced from multiple publicly available sources: Martin Baccardax, “[Nvidia jumps higher as Mizuho analysts see \\$300 billion AI chip potential](#),” *The Street*, July 24, 2023; Patrick Seitz, “[Intel on track with data center chip lineup, touts play in artificial intelligence](#),” *Investor’s Business Daily*, March 30, 2023; World Semiconductor Trade Statistics, “[WSTS Semiconductor Market Forecast Spring 2023](#);” Deborah Yao, “[Analysts’ take: Nvidia widens its total addressable market](#),” *AI Business*, June 1, 2023.
3. *The Economist*, “[Crypto-miners are probably to blame for the graphics-chip shortage](#),” June 19, 2021.
4. International Data Corporation (IDC), “[Weak consumer demand continues to delay a recovery for the smartphone market, according to IDC](#),” press release, May 31, 2023.
5. IDC, “[PC and Tablet market face further decline before a rebound in 2024, according to IDC](#),” press release, June 13, 2023.
6. Samuel K. Moore, “[Nvidia’s Next GPU Shows That Transformers Are Transforming AI](#),” *IEE Spectrum*, April 8, 2023.
7. Kif Leswing, “[Nvidia’s top A.I. chips are selling for more than \\$40,000 on eBay](#),” *CNBC*, April 14, 2023.
8. GPU Utils, “[Nvidia H100 GPUs: supply and demand](#),” July 2023 (updated August 2023), accessed September 15, 2023.
9. Lucas Mearian, “[Chip industry strains to meet AI-fueled demands — will smaller LLMs help?](#),” *Computerworld*, September 28, 2023.
10. Rita Liao, “[China’s AI firms might further lose chip access in new US ban](#),” *TechCrunch*, June 28, 2023.
11. Jeff Pau, “[SMIC bypasses US curbs to make 7nm chips](#),” *Asia Times*, September 5, 2023.
12. Dylan Patel, Myron Xie, Gerald Wong, and George Cozma, “[AI Capacity Constraints - CoWoS and HBM Supply Chain](#),” *Semi Analysis*, July 5, 2023.
13. Ibid.
14. Semiconductor Engineering, “[Advanced Packaging](#),” accessed November 14, 2023.
15. Brian T. Horowitz, “[AI Workloads Spur Competition in Networking Chips](#),” *Network Computing*, July 13, 2023.
16. Deloitte analysis of AI networking chip market.

17. Deane Dray, Jonathan Atkin, et al., *RBC Imagine: Datacenter Liquid Cooling Market Overview*, June 21, 2023.
18. Steve Taranovich, “*Data centers feel the power density pinch*,” *Electronic Design*, August 6, 2021.
19. Dylan Patel, Myron Xie, Gerald Wong, George Cozma, “*Energizing AI: Power delivery competition heats up Vicor, MPS, Delta, ADI, Renesas*,” *Semi Analysis*, August 2, 2023.
20. Dylan Patel, Myron Xie, Gerald Wong, and George Cozma, “*AI Capacity Constraints - CoWoS and HBM Supply Chain*.”
21. Ibid.
22. Chris Richard, Dan Hamling, Duncan Stewart, and Karthik Ramachandran, *Five fixes for the semiconductor chip shortage*, Deloitte Insights, December 6, 2021.
23. Lucas Mearian, “*Chip industry strains to meet AI-fueled demands — will smaller LLMs help?*”
24. Ibid.
25. Michelle Adams, “*Where Are All the New Semiconductor Fabs in North America & Europe?*” *Z2Data*, September 12, 2023.
26. Duncan Stewart, Karthik Ramachandran and Brandon Kulik, *Chipping in to boost production: US and Europe move toward greater self-sufficiency and resilient supply chains*, Deloitte Insights, April 24, 2023.
27. Anton Shilov, “*Memory makers on track to double HBM output in 2023*,” *AnandTech*, August 9, 2023.
28. Sheryl Miles, “*CHIPS Act implementation requires strong focus on 'Advanced Packaging'*,” *Electronic Specifier*, October 11, 2022.
29. Anirban Ghoshal, “*US wins support from Japan and Netherlands to clip China’s chip industry*,” *COMPUTERWORLD*, January 30, 2023.
30. Andrew Ross Sorkin, Ravi Mattu, Bernhard Warner, Sarah Kessler, Michael J. de la Merced, and Lauren Hersch, “*The A.I. chips war could heat up this summer*,” *The New York Times*, June 28, 2023.
31. Kanjyik Ghosh and Stephen Nellis, “*China’s internet giants order \$5 bln of Nvidia chips to power AI ambitions -FT*,” *Reuters*, August 10, 2023.

Acknowledgements

The authors would like to thank **Uwe Lambrette, Jan Nicholas, Dan Hamling, Karthik Ramachandran, Chris Arkenberg, Bariş Sarer, and Rohan Gupta.**

Cover image by: **Manya Kuzemchenko**

Taking control: Generative AI trains on private, enterprise data

More companies, seeking to avoid the risk of models trained on public data, are expected to train generative AI on their own data to enhance productivity, optimize costs, and unlock complex insights.

ARTICLE • 10 MINUTE READ

In 2023, generative AI came out of the shadows. Grabbing headlines and driving an explosion of startups, it likely reshaped the strategic roadmaps of some of the world's largest companies. For the first time, AI systems became conversational, creative, and even seemingly emotional, able to render remarkable imagery and return deep and comprehensive (if not entirely accurate) answers to complex queries. In a matter of months, the capabilities of large language models (LLMs) and visual diffusion models provoked international debate about their potential impact on global economics and geopolitics.¹

Although this initial wave of generative AI has been primarily consumer-facing and trained on public data, a deeper groundswell is building beneath private models that include more proprietary and domain-specific data. Companies that have been accumulating data for years now have an opportunity to unlock more of its value with generative AI. Doing so effectively could help solve for some of the current challenges of public models but will likely require thoughtful investments and decision-making.

Deloitte predicts that, in 2024, enterprise spending on generative AI will grow by 30%, from an estimated US\$16 billion in 2023.² While enthusiasm has been high, enterprises have mostly been cautiously experimenting, trying to figure out the specific value of generative AI for their businesses and the costs of deploying, scaling, and operating them effectively.³

Still, the market is expanding, and more enterprises are allocating budgets to generative AI. In 2024, much of their generative AI spending is expected to be paid to leading cloud service providers for training models and providing computation for user queries, as well as to data scientists that will help bridge company data to foundational models. However, 2024 could also see growth in more on-premise graphic processing unit (GPU) data centers as larger businesses—and government entities—seek to bring more generative AI capabilities in-house and under their control, mirroring the prior lifecycle for digital transformation to cloud, then hybrid, and into data centers. The main limitations to growth will likely be access to talent—and for some, to GPUs⁴—but companies may also wrestle with unclear use cases and issues relating to data quality.

Pros and cons of public models

The year ahead will likely see somewhat less exuberance for generative AI by a more reasoned assessment of its capabilities and costs. Users and use cases are expected to help clarify where its strengths lie and where it may be unfit or simply untrustworthy. Challenges faced by early public models, like factual errors, “hallucinations” (where the model fabricates something that may sound right⁵), copyright, and fair use, are being confronted by those providers while further incentivizing more private models.⁶

Because generative models have required such massive volumes of training data, the first wave of public models were mainly trained on the largest data set available: the public internet.⁷ Thus, these models have also absorbed the many biases, contradictions, inaccuracies, and uncertainties of the internet itself. In some ways, this has enabled them to converse on a remarkable array of topics and to exhibit surprisingly creative, poetic, and even seemingly emotional behaviors. But it has also required work to normalize results, avoid toxic output, and reinforce more accurate and preferred responses.

When pressed for facts, models trained on public data, such as social network posts, may fabricate them.⁸ And they can do so with authority, causing many users to believe their assertions without properly fact-checking the results. Popular LLMs were not designed to be factually accurate, but rather statistically accurate. They are very good at guessing what a typical human expects to come next in a sentence. This capability, combined with a model’s “temperature”—the amount of randomness allowed in a model’s response⁹—can introduce hallucinations, leading to a headline, for example, of a lawyer presenting “case law” from a generated legal brief that was made up by the model.¹⁰ However, this capability also fuels its creativity—for example, using visual diffusion models to generate novel character design for video games.¹¹

Publicly trained models have also run afoul of laws regarding copyright and fair use, with lawsuits mounting from those who see their own works reflected in generative outputs.¹² This has been especially problematic for diffusion models that generate images based on public training sets that include copywritten works.¹³ In response, some providers are enabling websites to cloak their content from being scraped as training data, potentially adding to the challenges of public models seeking training sets.¹⁴ And although copyright laws can vary by market, some can make AI-derived works indefensible, either for being overly derivative of prior art or for not being human enough to merit copyright.¹⁵ However, artists and copyright holders may be challenged to prove derivation from training sets that include billions of diverse inputs.¹⁶ Additionally, companies may be concerned about losing control of their own data if they add it to public models. Data leakage can happen when data used in training sets becomes visible to users—accidentally or by adversarial prompt engineering.¹⁷ For all these reasons, many businesses have been hesitant to adopt publicly trained generative AI.¹⁸

Leading providers of generative AI are also reckoning with these challenges and feeling pressure to evolve their business models.¹⁹ They face lawsuits and regulations for all the above reasons, while spending capital to train and tune models that are supporting millions of daily user prompts.²⁰ With enormous costs of computing for training models and inference at scale, hyperscale data centers may be capable providers and also be able to bear the brunt of costs and responsibility.

From consumer-facing to private domains

Because the fundamental capabilities of generative AI are compelling and relying on public solutions can introduce unwanted risks, more companies are looking to deploy their own models trained on their own private data.²¹ Doing so can avoid copyright and usage issues while enabling companies to develop bespoke solutions designed to produce desired behaviors and trustworthy results.

For many media and entertainment companies, generative AI has already disrupted content creation, enabling anyone to generate text, audio, and images. However, the most common tools that have enabled this disruption were trained on the public web, provoking lawsuits by authors and artists that believe their own works were included without consent or remuneration.²² To avoid such usage issues, both Adobe Systems²³ and Getty Images²⁴ have launched generative AI solutions that are trained on their own licensed visual content—the photographs and digital images they have amassed over their years of operation. When these tools generate new images, the results are explicitly within the licensing and reuse agreements of their content libraries. This can help them avoid copyright challenges while extending pathways for creators to license and monetize their own work in private training sets.

Still, companies will be required to abide by the leading practices and regulations governing the kinds of data being used, such as personally identifiable data or medical information. Those companies merging private and public data may similarly be challenged to integrate them effectively, while adhering to data privacy and copyright laws. Nevertheless, these are conversational learning systems, perhaps in their early days, that are showing potential to find and amplify value in data.

If data is “the new oil,” as many have said, LLMs and diffusion models may offer a higher-performance engine to help make it work. Many companies have accumulated large amounts of data that generative AI can help them operationalize. Generative AI can offer them a better lens on their data, combining a conversational and visual interface with the ability to reckon with vast troves of data far beyond human reasoning. Gazing across 2024, more companies may see the influence of generative AI not only on their operations and product lines, but also within their C-suite and boardrooms.

The bottom line

More companies are looking to use generative AI to help drive productivity and optimize costs. Using generative AI capabilities, they may also be able to unlock more value in their data by surfacing complex insights, ferreting out errors and fraud, reducing risks in decision-making, seeking optimizations, predicting opportunities, and even amplifying creative innovation. Some are already developing domain-specific solutions that may likely show results in the coming year.²⁵ Indeed, more companies are beginning to unlock the competitive advantages of generative AI, so there may be risks in waiting. But there are many considerations for the costs of development and operations, where to deploy different parts of the value chain, and how to set guardrails and ensure accurate and trustworthy results.

Training with private data can avoid some of the pitfalls but may still require efforts to make generative AI outputs trustworthy and accurate. Constraining training sets to be more domain-specific can narrow their range of responses. Reinforcement learning²⁶ and human feedback²⁷ are already helping steer models toward preferred behaviors, but companies that know their data best should be the ones leading the development of reward models and policy optimization.²⁸ Such efforts can help tackle hallucinations and bias, although they might have their own limitations.²⁹ By optimizing for specific results, the novelty and creativity of a model can degrade over time.³⁰ Done well, feedback can enable greater domain expertise and superhuman reasoning capabilities within those domains.³¹

Companies planning to develop their own models should consider the costs of doing so. Models may be relatively easy to develop, especially with new, open-source models entering the market. Depending on the use case, a given company should try to understand how large of a model may be necessary, how much data they might need to train it effectively, and how much computation may be necessary to get it up and running. Companies may have diverse data sets of differing qualities that should be conditioned and brought together into a database.³² Data will then need to be organized. As they are familiar with their own data, companies may be the best equipped to guide the accurate labeling of training sets.

Generative AI models can have billions of parameters that may require training on very large data sets. This can necessitate a lot of computation.³³ Companies may need to work with hyperscale cloud providers—and plan to pay for the cycles—or purchase their own hardware, which can be expensive to buy and operate.³⁴ Training may be the most expensive part, but trained models should then respond to queries. If query workloads are large, inference costs can go up as well. This means that companies should carefully consider the costs of talent, computation, and time to develop, deploy, and operate a model, compared to the anticipated path to return on investment. Having a clear set of goals in mind, and a roadmap of implementations to get there, can keep projects on track while surfacing gains or losses early.

Using computation and expertise can also prompt considerations for deployment and collaborators. It may make sense to work with existing cloud providers. As companies scale, or if they have proprietary or sensitive data, they may choose a hybrid or on-premise data center approach. If they do, they should be thoughtful about redundancy and security, as they would with any other critical services—and maybe more so; that is because a compromised system could divulge deep intelligence about the company’s data, or an adversarial attack could cause a trusted AI to deliver manipulated behaviors to stakeholders.

In most cases, an ecosystem approach could benefit by distributing investment, expertise, and risk. However, each company should consider the best approach to the outcomes they’re trying to achieve. There are different pathways, but the “right” path should reflect a company’s unique needs for cost, performance, security, data types, and strategic objectives. Generative AI is a fast-moving and highly funded field that is just beginning to reveal its use cases, opportunities, and implications.

AI in the C-suite and the boardroom

Looking to the future, what might it mean if companies have their own intelligent learning systems? What does an AI-native organization look like? How much is it business-aligned versus human-aligned? What are the implications of having a conversational LLM that can see things in your data—or patterns of your competitors—that you can’t? Companies may soon have multiple agents that have moved into numerous workflows not just for operations but also planning and decision making.

As these systems establish value and trust, they could move further up the hierarchy of decision-making, potentially becoming a conversational voice in the C-suite or the boardroom.³⁵ Such possibilities were often considered science fiction, but in 2024, they’ll seem much closer and should be anticipated.

Ultimately, business leaders will be tasked with experimentation and careful planning to determine what generative AI can do for their bottom line. Will the capabilities of generative AI enable truly differentiated financial performance and competitive advantage? And if so, for how long might those competitive advantages last? Will they become the new table stakes for business performance? Stepping back, what signals might emerge to show whether generative AI is incremental or revolutionary?

By **Chris Arkenberg**
United States

Bariş Sarer
United States

Gillian Crossan
United States

Rohan Gupta
United States

Endnotes

1. David Solomon and Eric Schmidt, "[The future of generative AI](#)," Goldman Sachs, September 13, 2023.
2. Michael Shirer, "[IDC forecasts spending on GenAI solutions will reach \\$143 billion in 2027 with a five-year compound annual growth rate of 73.3%](#)," IDC, October 16, 2023.
3. Katyanna Quach, "[Despite the hype, generative AI is not a significant chunk of enterprise cloud spend](#)," Register, September 12, 2023.
4. Lucas Mearian, "[Chip industry strains to meet AI-fueled demands — will smaller LLMs help?](#)," Computerworld, September 28, 2023
5. Janakiram MSV, "[How to reduce the hallucinations from large language models](#)," New Stack, June 9, 2023.
6. Tiana Garbett and James G. Gatto, "[Generative AI and copyright – some recent denials and unanswered questions](#)," *National Law Review* 13, no. 319 (2023).
7. Sharon Goldman, "[Generative AI's secret sauce – data scraping – comes under attack](#)," VentureBeat, July 6, 2023.
8. Sascha Heyer, "[Generative AI – understand and mitigate hallucinations in LLMs](#)," Medium, June 13, 2023.
9. Sascha Heyer, "[Generative AI – mastering the language model parameters for better output](#)," Medium, June 12, 2023.
10. Benjamin Weiser and Nate Shweber, "[The ChatGPT lawyer explains himself](#)," *New York Times*, June 8, 2023.
11. Shannon Liao, "[A.I. may help design your favorite video game character](#)," *New York Times*, May 22, 2023.
12. "[From ChatGPT to Getty v. Stability AI: A running list of key AI-lawsuits](#)," Fashion Law, October 19, 2023.
13. James Vincent, "[Getty Images sues AI art generator Stable Diffusion in the US for copyright infringement](#)," Verge, February 6, 2023.
14. Danielle Romain, "[An update on web publisher controls](#)," Keyword, Google, September 28, 2023.
15. Christopher Hutton, "[Generative AI set for era-defining clash with copyright law](#)," Washington Examiner, April 20, 2023.
16. Blake Brittain, "[US judge finds flaws in artist's lawsuit against AI companies](#)," Reuters, June 19, 2023.

17. Jaydeep Borkar, "[What can we learn from data leakage and unlearning for law?](#)," Cornell University, July 19, 2023.
18. Carl Franzen, "[More than 70% of companies are experimenting with generative AI, but few are willing to commit more spending](#)," VentureBeat, July 25, 2023.
19. Medium, "[Why Gen AI adoption among businesses will look radically different in 2024](#)," September 13, 2023.
20. Will Oremus, "[AI chatbots lose money every time you use them. That is a problem.](#)," *Washington Post*, June 5, 2023.
21. *Economist*, "[AI is setting off a great scramble for data](#)," August 13, 2023.
22. Christopher J. Valente, Michael J. Stortz, Amy Wong, Peter E. Soskin, and Michael W. Meredith, "[Recent trends in generative artificial intelligence litigation in the United States](#)," K&L Gates, September 5, 2023.
23. Ashley Still, "[Reimagining our video and audio tools with Adobe Firefly](#)," Adobe Blog, April 17, 2023.
24. Getty Images Newsroom, "[Getty Images launches commercially safe generative AI offering](#)," September 25, 2023.
25. Jamiel Sheikh, "[Bloomberg uses its vast data to create new finance AI](#)," Forbes, April 5, 2023.
26. Cameron Hashemi-Pour, "[What is reinforcement learning?](#)," TechTarget, August 2023.
27. Jan Leike, Miljan Martic, and Shane Legg, "[Learning through human feedback](#)," Google DeepMind, June 12, 2017.
28. Dimitriy Konyrev, "[Reinforcement learning with human feedback \(RLHF\) for LLMs](#)," SuperAnnotate, April 27, 2023.
29. Ben Dickson, "[The challenges of reinforcement learning from human feedback \(RLHF\)](#)," TechTalks, September 4, 2023.
30. Jithin James, "[The impact of temperature in LLMs: Balancing determinism and creativity](#)," Medium, July 12, 2023.
31. Leike, Martic, and Legg, "[Learning through human feedback.](#)"
32. Tom Davenport and Maryam Alavi, "[How to train generative AI using your company's data](#)," *Harvard Business Review*, July 6, 2023.
33. Sid Sheth, "[Generative AI drives an explosion in compute: The looming need for sustainable AI](#)," SiliconAngle, February 5, 2023.

34. Guido Appenzeller, Matt Bornstein, and Martin Cassado, “*Navigating the high cost of AI compute*,” Andreessen Horowitz, April 27, 2023.

35. Stanley McChrystal, “*AI has entered the situation room*,” Foreign Policy, June 19, 2023.

Acknowledgements

Cover image by: **Manya Kuzemchenko**

Walking the tightrope: As generative AI meets EU regulation, pragmatism is likely

2024 is likely to see a balance between regulatory compliance and fostering innovation in generative AI; clear regulation enables enterprises and vendors to proceed with certainty.

ARTICLE • 11 MINUTE READ

The importance of well-crafted rules should not be overstated when it comes to unlocking the potential of any market. In the case of generative AI, the absence of clear regulatory conditions may cause vendors, enterprise customers, and end users to hesitate. However, the European Union (EU) is expected to set the stage for global regulation of generative AI in 2024, not only influencing its own markets but also serving as a template for other regions.

In 2024 two EU regulations are expected to help shape the growth of the generative AI market in the region and further afield. These are the General Data Protection Regulation (GDPR),¹ which has been applicable since 2018, and the upcoming EU AI Act (AIA), expected to be agreed to in early 2024. As generative AI opens up debates on how to manage issues of individual consent, rectification, erasure, bias mitigation, and copyright usage, the industry's trajectory could be shaped by how organizations and regulators view, enforce, and manage areas of contention.

Despite potential challenges, collaboration in the form of open and transparent conversations between industry and regulators is likely to result in a pragmatic approach that balances regulatory compliance with fostering innovation in generative AI. This would continue the pattern of discussions held in 2023, which saw interventions by regulators in the European Union and other markets. Vendors have adjusted their approach to meet with regulators' requests; regulators have enabled innovation.² By addressing the concerns raised by EU regulations in 2024, while promoting the benefits of core technologies, the generative AI market is expected to continue to evolve productively.

Existing and drafted EU regulations are likely to influence generative AI globally

This prediction is focused on EU regulations on generative AI as it is likely to be among the first set of agreed-to regulations that will have a global impact.³ In recent years, there has been a clear "Brussels effect,"⁴ with EU regulations having global ramifications, and we expect a similar impact from EU regulations that would cover generative AI.⁵ The extraterritorial impact is likely to have varied impacts:

1. **It is directly applicable to vendors operating from any market that are selling into, or targeting users in, EU countries.** Organizations that are noncompliant are likely to be subject to material fines.
2. **Other markets may use EU regulation as a template.** EU regulation has, for example, influenced India's Digital Personal Data Protection Act, 2023⁶ and equivalent regulations in Brazil and California. The AIA may have informed a US Senate bipartisan blueprint for AI that includes elements such as the licensing of high-risk applications like face recognition and public disclosure of training data used in foundation models.⁷
3. **Multinationals, as well as those complying within the European Union, may apply their AI governance (including that specific to generative AI) globally** for a more standardized approach aligned to leading practices, based on compliance with EU regulations.

The majority of EU regulation pertaining to generative AI should become relatively clear by the first quarter of 2024.

In 2024, the direction of European regulation on generative AI is likely to become far clearer. At this time, the industry should have sight of the agreed text of the AIA, which complements the GDPR.⁸ All companies looking to offer or deploy generative AI solutions should monitor developments in the AIA while also maintaining compliance with the GDPR.

The process for final agreement is in three stages; at the time of writing, two stages had been finalized, with the third and final phase pending the outcome of the “dialogue” between the EU Council, Parliament, and Council.

- The Council of the European Union finalized its position in December 2022,⁹ at which point generative AI had only just reached mainstream awareness following the launch of ChatGPT.
- The EU Parliament finalized its position in June 2023, and this included specific regulation for generative AI.¹⁰ This prediction's references to the AIA reflect mostly the status as of this point in time.
- The final version of the AIA, expected in early 2024, may include variations to the Parliamentary position. There will then be a further two years before the AIA is applicable.

There are specific terms that are applicable to the European Union's regulation of generative AI, and it is important to define these. The critical components and types of players that the European Union has defined for the purposes of regulating generative AI within the AIA are:

- **Foundation models (FMs):** These are AI models that are trained on data at scale such as OpenAI's GPT models or Google's PaLM 2.¹¹ These models can be applied to a wide range of tasks, and as such, differ from narrow AI models. Per the European Union's definition, a foundation model could be used in a GPAI or other more specific AI models.¹²
- **General purpose AI (GPAI):** The EU regulation defines this as an AI system¹³ that is designed to perform “generally applicable functions” and does not have an “intended purpose.”¹⁴ It can be used in a “plurality of contexts” and in a “plurality of other AI systems.” Core capabilities of a GPAI would include recognition (e.g., of images or speech), generation (currently most commonly of text or images), pattern detection, and translation.
- **Generative AI:** This is defined as an AI system created specifically to generate outputs in a range of formats. The best-known generative AI applications include ChatGPT, Snap AI, Google Bard, and Microsoft's M365 Copilot.

There are two types of entities that will be in scope:

- **Providers:** A person, public authority, agency, or other body that develops or commissions an AI system with a view to making this publicly available, for a charge or for free.
- **Deployers:** A person, public authority, agency, or any other body using an AI system under its authority. In some contexts, a deployer may be considered a provider. This would be the case if the deployer uses an AI system for a high-risk application.

This prediction will firstly focus on the GDPR, whose obligations are known, and then on the AIA, whose shape is forming but not yet finalized.

Generative AI and the European Union's GDPR

Generative AI is expected to need to comply with the GDPR on processing of personal data. The GDPR, which came into effect in May 2018,¹⁵ defines the rights of “data subjects”—that is, individuals whose personal data being processed could be used to identify those people.

A fundamental tenet of EU regulation is that individuals' personal data use is grounded on applicable legal grounds, with lawfulness of processing to be maintained for each processing activity.¹⁶

This requirement may seem to clash with the core approach of generative AI, which is based on foundation models. Each model is trained on massive quantities of raw data—the more the better. A large proportion of this data—the exact share varying by model—may require consent per some interpretations of EU law. The largest foundation models may have been trained on petabytes (thousands of gigabytes, or GB) of data.¹⁷ Earlier ones, including GPT-3, were trained on 570 GB of data.¹⁸ Generative AI applications in any medium—text, image, code, or other—create content using the knowledge within each foundation model.

Given the vast number of people whose data may have been used, obtaining individual consent, where required, becomes a complex exercise. Furthermore, as each foundation model supports an effectively infinite number and range of applications, requesting permissions for each additional purpose is even more unrealistic.

However, obtaining individual consent might not be mandatory. “Legitimate interest” may prove to be a sufficient “lawful basis” that permits training of the foundation models that drive generative AI.¹⁹ A legitimate interest exists when there is a compelling reason for processing and that processing of the data is the only approach to achieve the desired outcome.²⁰ Regulators are likely wanting to see that organizations have conducted the appropriate evaluations to ensure that claimed legitimate interests and individuals' rights and freedoms are balanced.

Furthermore, obtaining individual permissions may well be considered a “disproportionate effort.” An acceptable middle way may be mass-market communication. This was one of the steps requested in April 2023 by the Italian regulator, the Garante, of OpenAI to permit it to reinstate service.²¹ It placed an obligation on the data controller (the nominated person responsible for the foundation model) to launch an awareness campaign in broadcast and online media. This was meant to inform users that personal data may have been used and explain how such data could be deleted via an online tool.

Regulators might view positively that the intent of training is specifically to create better inferential capability that can then be deployed in generative applications (such as OpenAI's ChatGPT, Stability.ai's DreamStudio, or Adobe's Firefly).

The European Data Protection Board may provide more clarity on the issue of consent, among other contentious areas, in 2024.²²

The GDPR tenets of rectification, erasure, and the right to be forgotten are applicable to the foundation models that underpin generative AI

GDPR includes a suite of rights with regards to personal data. If data is incorrect, an individual can ask for it to be corrected. If the data subject no longer wants their personal data to be associated with or processed by that organization, they can ask for it to be deleted. These rights have been well known since GDPR came into force. Addressing such requests may cost organizations thousands of dollars.

The foundation models that underpin generative AI are trained on myriad websites that may contain errors. The training process is a single event during which errors can be absorbed into the model. Updating the model to reflect rectifications or other changes could be done most accurately by retraining the model, but this implies substantial costs and time.²³

The approach that's likely to be used to satisfy this requirement will be to use negative feedback loops to fine-tune the model.²⁴ If an original data point is determined to be wrong, weighting applied to the erroneous data point can be changed to minimize the likelihood of that data point reappearing. Feedback loops are imperfect but may be considered appropriate. That said, it is not certain how this approach might work in the case of class action challenges, which may require large swathes of data to be deleted.

Data minimization and statistical accuracy

The idea of data minimization is that collection of personal information should be limited to what is strictly relevant and necessary to achieve a specific task, and as soon as this is complete, the data should be deleted.²⁵ This approach may seem to be at odds with foundation models, with their efficacy related to how much data they can query, with more being better.

However, the principle of data minimization may still be compatible with generative AI if data is de-personalized, for example, by using approaches such as pseudonymization (swapping personal identifiers with placeholder data, which reduces, but does not eliminate, data protection risks) and anonymization (deleting identifiers, which means data is no longer “personal”).²⁶ Using these approaches, the volume of training data can be maintained, but full anonymization may be challenging. Organizations should have an appropriate framework in place to assess and explain to and assure the regulator how they determine what is necessary.

The size of foundation models is linked to statistical accuracy, which is an element of proposed EU regulation included in the AIA.²⁷ In an AI context, accuracy refers to the quality of outputs generated. With a foundation model, the greater the volume of good training data, the more accurate the results should be.²⁸

The next part of the prediction considers the possible impact of the AIA on generative AI.

AIA obligations on foundation models, per the EU Parliament's agreement

As mentioned earlier, the EU Parliament finalized its position in June 2023, and this included specific regulation for generative AI. The final version of the AIA, expected in early 2024, may include variations to the Parliamentary position.

The Parliamentary agreement included the following elements:

- Foundation models should be registered in an EU database.
- Models should be tested extensively to have appropriate levels of predictability, interpretability, corrigibility, safety, and cybersecurity for the entirety of the model's expected life cycle.
- Design, testing, and analysis should identify and reduce risks during the model's development.
- Datasets used in training models should have sufficient data governance standards. Data sources should be assessed for data quality and bias.
- Energy usage should be minimized and monitorable across the model's life cycle.
- Extensive, accessible technical documentation should be available to downstream providers, to enable their compliance. This documentation should be available for a decade from commercial launch.
- A quality management system should ensure and document compliance.

Additionally, providers of FMs used in generative AI systems and providers who specialize an FM into a generative AI system should:

- Comply with additional transparency obligations, including the specific labeling of outputs as AI-generated.
- Ensure safeguards against the generation of outputs that breach EU law.
- Document and publish summaries of training data that is protected by copyright.

Bias may be mitigable

The AIA aims to minimize bias within AI systems. This includes the suppression of human bias. Foundation models may have been trained on biased content such as biased text related to gender, race, or sexual preference for example.

Training data is also likely to include language biases, with most content written in English, with additional biases resulting from the preponderance of content ingested from writers of a specific gender, ethnicity, social class, degree of education, and income group.²⁹ Historical biases used to train foundation models could, therefore, generate content that repeats or even accentuates those biases.

Regulators are likely to require that biases be mitigated via any of a variety of techniques, including weighting or the inclusion of synthetic data that can balance out bias.³⁰ Data controllers—which could be both the AI developer and the AI deployer—are likely to be asked to document “traceability,” which explains steps taken.³¹

Copyright: More clarification on permitted behaviors required

In 2024, further clarity is likely to be needed regarding the use of copyright content.³²

Existing EU law may permit usage of copyright data for training, specifically “instances of text and data mining that do not involve acts of reproduction or where the reproductions made fall under the mandatory exception for temporary acts of reproduction.”³³ The AIA draft requires that copyright works used for training be listed.

The EU recently, via the Digital Single Market Directive,³⁴ introduced permissions for use of text and data mining for scientific research and for commercial lawful use; although for commercial use there is a right to “opt out” of that permission. Content owners, including several media companies, have exercised that right to opt their data out of AI training.³⁵ As of April 2023, more than a billion items had been removed from a training set for the Stable Diffusion v3 model.³⁶

The largest foundation models may be categorized as systemic

The AIA focuses on risk assessment of each application. This runs counter to the general-purpose nature of foundation models.

However, there may be a distinction between systemic foundation models (SFM)—those whose impact represents a system risk—and others, following an approach used in the EU Digital Services Act when categorizing types of online platforms and search engines.³⁷ Designation as an SFM is likely to be made according to the quantity of computing resources required to train the model, the type and cost of training inputs used, and its likely market impact. SFMs are likely to have a greater degree of due diligence obligations.³⁸

Another possible outcome is that the AIA may establish some baseline requirements applicable to all foundation models (e.g., around transparency and technical documentation), with additional requirements if foundation models are used for high-risk use cases.

The bottom line

European regulation matters. It is likely to have extraterritorial and regional impacts. At first glance, several existing principles of EU regulations that apply to digital services may have seemed to present major obstacles to the growth of the generative AI market. Indeed, some commentators had likely expected generative AI to be incompatible with EU guidelines.

How generative AI will shape up in the years ahead and what impact it could have is still unknown. It may be several years before the scale and nature of its impact is certain. In 2024, and beyond, vendors and regulators are likely to aspire to collaborate to attain an outcome that works for consumers, enterprises, vendors, and society in general. Governments are acutely aware of the importance of enabling innovation in generative AI—for example, via regulatory sandboxes.³⁹

In 2024, as generative AI applications evolve and the resulting legal challenges become clearer, the direction of the regulatory response may become more evident. Generative AI is likely to remain an emerging sector this year, which can make it hard for regulation to be explicit at this stage. There will likely still be core questions to address, such as the responsibilities for providers of generative AI versus deployers, when each is a separate entity.

By

Paul Lee
United Kingdom

Lucia Lucchini
United Kingdom

Michelle Seng Ah Lee
United Kingdom

Endnotes

1. European Union (EU), *Directive 95/46/EC (General Data Protection Regulation)*, April 27, 2016.
2. European Commission (EC), “*First regulatory sandbox on Artificial Intelligence presented*,” last updated January 30, 2023; Spanish Ministry of Finance, “*Approved statute of the Agencia Española de Supervisión de la Inteligencia Artificial (AESIA)*,” August 22, 2023.
3. Anna Gamvros, Edward Yau, and Steven Chong, “*China finalises its Generative AI Regulation*,” Norton Rose Fulbright, July 25, 2023.
4. Charlotte Siegmann and Markus Anderljung, *The Brussels Effect and artificial intelligence: How EU regulation will impact the global AI market*, Centre for the Governance of AI, 2021.
5. Tatjana Evas, *European framework on ethical aspects of artificial intelligence, robotics and related technologies*, European Parliamentary Research Service (EPRS), 2020.
6. India Ministry of Law and Justice, *The Digital Personal Data Protection Act, 2023, Gazette of India*, August 11, 2023; Raktima Roy and Gabriela Zanfir-Fortuna, “*The Digital Personal Data Protection Act of India, explained*,” Future of Privacy Forum, August 15, 2023.
7. Khari Johnson, “*Senators want ChatGPT-level AI to require a government license*,” *Wired*, September 9, 2023.
8. *European Parliament News*, “*EU AI Act: First regulation on artificial intelligence*,” last updated June 14, 2023.
9. Spanish Presidency of the Council of the European Union, “*The EU pioneers regulation of artificial intelligence*,” October 22, 2023.
10. *European Parliament News*, “*EU AI Act: First regulation on artificial intelligence.*”
11. Rick Merritt, “*What are foundation models?*,” *Nvidia Blog*, March 13, 2023; Jacob Devlin et al., “*BERT: Pre-training of deep bidirectional transformers for language understanding*,” Cornell University Arxiv, last updated May 24, 2019; Google AI, “*AI across Google: PaLM 2*,” accessed November 20, 2023.
12. Elliot Jones, “*Explainer: What is a foundation model?*,” Ada Lovelace Institute, July 17, 2023.
13. Council of the European Union, “*Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*,” November 29, 2021; Carlos Ignacio Gutierrez, Anthony Aguirre, and Risto Uuk, “*The European Union could rethink its definition of General Purpose AI Systems (GPAIS)*,” OECD.AI, November 7, 2022.

14. Council of the European Union, “*Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts*,” May 13, 2022.
15. Council of the European Union, “*The General Data Protection Regulation*,” last updated September 1, 2022.
16. Norwegian Consumer Council, *Ghost in the machine: Addressing the consumer harms of generative AI*, June 2023.
17. Kim Martineau, “*IBM and NASA team up to spur new discoveries about our planet*,” *IBM Blog*, February 1, 2023.
18. Merritt, “*What are foundation models?*”
19. Pablo Rodrigo Trigo Kramcsak, “*Can legitimate interest be an appropriate lawful basis for processing Artificial Intelligence training datasets?*,” *Computer Law & Security Review* 48 (April 2023).
20. Information Commissioner’s Office (ICO), “*Legitimate interests*,” accessed November 20, 2023.
21. Italian Data Protection Authority, “*The Guarantor for the Protection of Personal Data*,” March 30, 2023.
22. European Data Protection Board (EDPB), “*EDPB resolves dispute on transfers by Meta and creates task force on ChatGPT*,” press release, April 13, 2023.
23. Will Knight, “*OpenAI’s CEO says the age of giant AI models is already over*,” *Wired*, April 17, 2023.
24. Haziqa Sajid, “*The AI feedback loop: Maintaining model production quality in the age of AI-generated content*,” *Unite.ai*, July 25, 2023.
25. European Data Protection Supervisor, “*Data minimization*,” accessed November 20, 2023.
26. Ireland Data Protection Commission, “*Apply anonymity and pseudonymity*,” accessed November 20, 2023; ICO, “*Chapter 3: Pseudonymisation*,” *Draft Anonymisation, Pseudonymisation and Privacy Enhancing Technologies Guidance*, February 2022.
27. European Parliament, *Artificial Intelligence Act*, accessed November 20, 2023.
28. Rishi Bommasani et al., *On the opportunities and risks of foundation models*, Center for Research on Foundation Models (CRFM), Stanford Institute for Human-Centered Artificial Intelligence (HAI), Stanford University, July 12, 2022.
29. Aldo Lamberti, “*Tackling bias in large ML models: The role of synthetic data*,” *Syntheticus*, July 31, 2023.

30. Syntheticus, *Synthetic data 101: What is it, how it works, and what it's used for*, accessed November 20, 2023.
31. Suhas Maddali, "*How to address data bias in machine learning*," *Towards Data Science*, July 27, 2022.
32. Atsuki Mizuguchi, "*Legal issues in generative AI under Japanese law*," *Nishimura & Asahi*, July 11 2023.
33. European Parliament, "*Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC*," *Official Journal of the European Union*, May 17, 2019.
34. Ibid.
35. João Pedro Quintais, "*Generative AI, copyright and the AI Act*," *Kluwer Copyright Blog*, May 9, 2023.
36. Kyle Wiggers, "*Spawning lays out plans for letting creators opt out of generative AI training*," *TechCrunch+*, May 3, 2023.
37. European Union, *The Digital Services Act (DSA)*, accessed November 20, 2023.
38. J. Scott Marcus, "*Adapting the European Union AI Act to deal with generative artificial intelligence*," *Bruegel*, July 19, 2023.
39. EC, "*First regulatory sandbox on Artificial Intelligence presented*," June 27, 2022; Spanish Ministry of Finance, "*Approved statute of the Agencia Española de Supervisión de la Inteligencia Artificial (AESIA)*."

Acknowledgements

The authors would like to thank **Nick Seeber, Lukas Kruger, Suchitra Nair, Ben Stanton, Robert MacDougall, Joanne Conway**, and **Isabel Parker** for their contributions to this article.

Cover image by: **Manya Kuzemchenko**

A raw deal: Will materials shortages, supply chain challenges threaten tech's future?

In the face of rising trade issues and skyrocketing demand, e-waste recycling and digital supply networks and a holistic approach to supply chain sustainability could help in the medium term.

ARTICLE • 8 MINUTE READ

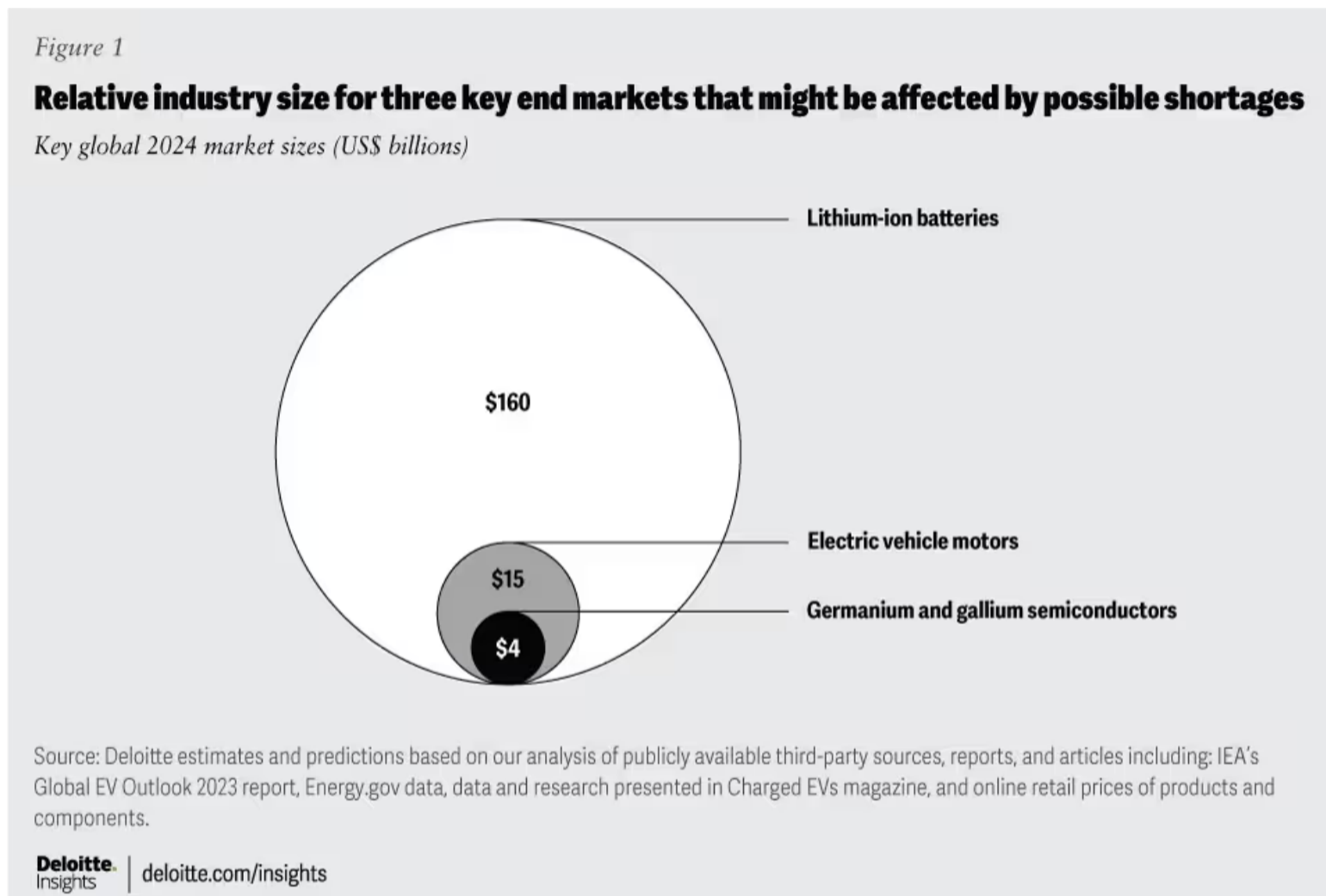
You can't make certain chips, motors, batteries, or a host of other key technologies without certain materials. Deloitte predicts that multiple regions will run short of gallium and possibly germanium as soon as 2024, impacting chipmakers. By 2025, we may start seeing shortages of rare earth elements (REEs) for magnets in electric car motors and more, as well as lithium and cobalt for batteries.¹ Actions can be taken, however, that can help avoid shortages in the near, medium, and long term.

How we got here

Some of these raw material shortages were more predictable: Anticipated growth in battery electric vehicles (BEVs) is forecast to drive demand for lithium and cobalt above the predicted production ramp for years, which could affect BEVs, but also consumer and enterprise devices such as laptops, tablets, and smartphones.² However, geopolitical tensions between China and western economies have recently impacted supply chains, with the potential for future impacts, as well.

First, China is subject to export controls that limit its ability to purchase critical chips, as well as the technologies and software needed to make the chips domestically³ (see [gen AI chips prediction](#)). In July, China announced that it would begin to impose its own controls on the export of germanium and gallium.⁴ Second, some analysts have expressed concerns that China may also restrict the export of the 17 elements that make up REEs (used for electronics, clean energy, aerospace, automotive, and defense) in 2024, leading to possible shortages for western manufacturers that year or the year after.⁵ Further, some of the alternate sources of tech-important raw materials are in geographies in which political, regulatory, or social factors may decrease their reliability in the long term.⁶

There have been raw materials shortages before that affected the tech industry. For example, the tantalum shortage of 2000 impacted capacitor supply.⁷ But what could be unprecedented in 2024 and 2025 is that these shortages could possibly be across dozens of different raw materials at the same time.⁸ Also unprecedented, the industries enabled by these materials play a much bigger role in the economy than they used to. Such industries are estimated to be worth over US\$160 billion annually, or (in the case of semiconductors or electric vehicle (EV) motors) are worth smaller dollar amounts but are still critical for innovation, growth, and national security (figure 1).⁹



When speaking of these shortages, time frames matter

In the very short term (late 2023 and early 2024) many companies that need germanium and gallium have inventories that are expected to last them into the first half of 2024. After that, gallium may be the bigger challenge to acquire, as there is a large and reliable source of germanium in British Columbia.¹⁰

In the longer term, mines and smelters can be built. Gallium is mainly produced as a byproduct of aluminum production from bauxite,¹¹ and there are bauxite deposits in dozens of countries and on every continent.¹² Equally, rare earth deposits are often not that rare, and mines are planned for Australia, Angola, Afghanistan, Canada, and the US over the next few years.¹³

Between now and then there are three big things that could be done to help mitigate supply chain vulnerabilities to existing or probable material shortages.

- **E-waste recycling:** Globally, the electronics industry throws away a huge volume of valuable elements, collectively worth an estimated US\$50+ billion annually.¹⁴ Many of those elements are often vulnerable to supply chain disruptions. A rapid increase in recycling could help boost supply, and a push to innovate e-waste recycling technologies could help to accelerate this even more.
- **Digital supply networks (DSNs):** These allow industries to do more with less: DSNs can anticipate and plan for raw material shortages and make them less severe by getting exactly the right minerals to the right place at the right time with less waste. They are not expected to stop shortages from happening, but they can help make them less severe or protracted. Building DSNs is often synergistic with other sustainable supply chain efforts including, for example, collaboration with and incentivization of supplier networks to reduce other forms of waste, including GHG emissions (see [semi sustainability prediction](#)).
- **Stockpiling:** As an example, the US currently has strategic reserves of petroleum, grain, bullion, helium, and raw materials critical for military defense.¹⁵ Would it make sense to create analogous reserves of critical raw materials for EVs and semiconductors?

Not just elementary, my dear Watson

Thus far, we have entirely discussed elements. But tech and semi supply chains are not just about raw materials. They are also about:

1. Refined, processed, and purified materials. Low purity neon or silicon are plentiful. But the gas necessary for lasers in semi manufacturing and the silicon for ingots need to be ultrapure, and sources of the ultrapure materials are highly concentrated: Ukraine is the source of about 50% of semiconductor grade neon,¹⁶ while China is the source of 80% of global polysilicon (used for both semi manufacturing and solar power) supply.¹⁷
2. Manufactured specialty compounds. It might be an epoxy or resin, a special cleaning fluid or gas, or type of plastic.¹⁸ Any geopolitical uncertainty, or an earthquake, typhoon, hurricane, fire, flood, drought, or pandemic could cause significant and lasting supply chain interruption.

Strategies such as recycling, DSNs, and in general more sustainable semi manufacturing (see [semi sustainability prediction](#)), plus stockpiling could help. While building a new mine takes five to 10 years, building a new factory, for example, takes approximately two to three years.¹⁹

The bottom line

The world produced 57.4 million metric tons of e-waste in 2022,²⁰ of which only 17% was formally collected and recycled,²¹ although there is significant regional variation, with the United States recycling about 15% in 2019 and the EU about 42.5% as of 2022.²² On current trends, the e-waste number is predicted to rise to 75 million tons by 2030.²³ There are a variety of e-waste sources: Large appliances alone make up more than half of e-waste in the European Union, but IT, telecom, and consumer electronics make up almost 30% as of 2020.²⁴ The International Telecommunication Union set a goal for e-waste recycling to reach 30% globally by 2023, but that goal will likely not be achieved.²⁵

Why isn't more e-waste being recycled? At a high level, it's often because it's unprofitable. Recycling doesn't make money; it costs money. That's mainly because each device contains a mix of recyclable elements, often chemically bound to other elements or compounds, and physically bound to other elements or compounds, such as solders, plastics, and ceramics.²⁶ Extracting metals requires work, energy, chemical, and physical processes, and often results in dangerous or harmful waste.²⁷ Much of the current recycling is done outside where the toxic e-waste is produced. For example, e-waste from the developed world is often shipped to and processed in the developing world, especially impacting women and children's health: between 2.9 and 12.9 million women may be at risk from exposure to toxic e-waste through their work in the informal waste sector, and around 18 million children are employed in waste processing industries.²⁸

Although there's an estimated US\$15 billion in valuable metals in e-waste annually (mainly from printed circuit boards),²⁹ the costs are sometimes higher, resulting in the current, relatively low percentage of e-waste being recycled. However, if there was a significant push for supply chain resilience reasons (rather than purely economic ones) for REEs, lithium, cobalt, and semiconductors, we could see two benefits: The percentage of e-waste being recycled would likely climb, and more of the recycling may need to occur on shore in the European Union and North America, helping reduce environmental harm to other countries.

A number of niche recycling and circular logistics, process, and business model providers have emerged, and represent a foundation upon which such scaling of recycling can build. The US government has recently announced various incentives for extracting lithium and REEs from e-waste.³⁰

Of the elemental challenges, REEs are one of the biggest, as the magnets enabled by REEs are found in EVs, wind turbines, defense systems, and much more.³¹ An interesting longer-term solution may be found in Vietnam. According to the United States Geological Survey (USGS), Vietnam has the second largest rare earth deposits at 22 million tons, after only China.³² And it has already started to ramp up production, with output of raw REEs in 2022 of 4,300 tons, over ten times higher than the year before, and with a 2030 goal of 2 million tons by 2030.³³

How much would a strategic US or EU gallium stockpile cost? Low purity gallium costs about US\$280 per kilogram, while 99.99999% pure gallium costs about US\$450 per kilogram, and gold is roughly US\$66,000 per kg.³⁴ Large users of gallium for electronics purposes use “dozens of tons” per year,³⁵ so a three-year reserve would be approximately \$20 million (unpurified) to \$30 million (purified.)

Not only are REEs and other important metals, such as gallium and germanium, mined in China, but they are also usually smelted there, too. If trade restrictions between China and the West escalate further, and as the technology industry increasingly uses various hard-to-get elements, building supply chain resilience likely requires reducing source material concentration and building more mines and smelters on shore or near shore in the long term. In the near term, resilience can also be boosted via investments in e-waste recycling, digital supply networks, and stockpiling.

These initiatives will take billions of dollars of investment, but the industries that rely on these materials are worth a hundred times more in annual revenues.

By **Duncan Stewart**
Canada

Gillian Crossan
United States

Ariane Bucaille
France

Christie Simons
United States

Endnotes

1. Joe McDonald, “*Threatened by shortages, electric car makers race for supplies of lithium for batteries,*” *AP News*, June 27, 2023.
2. Jessica Shankleman, Tom Biesheuvel, Joe Ryan, and Dave Merrill, “*We’re Going to Need More Lithium,*” *Bloomberg*, September 17, 2017.
3. Bureau of Industry and Security (U.S. Department of Commerce), “*Commerce strengthens restrictions on advanced computing semiconductors, semiconductor manufacturing equipment, and supercomputing items to countries of concern,*” press release, October 17, 2023; European Parliament, “*EU AI Act: first regulation on artificial intelligence,*” news release, June 14, 2023.
4. Hannah Ziady and Xiaofei Xu, “*China hits back in the chip war, imposing export curbs on crucial raw materials,*” *CNN Business*, July 3, 2023.
5. Mai Nguyen and Eric Onstad, “*China’s rare earths dominance in focus after it limits germanium and gallium exports,*” *Reuters*, October 20, 2023.
6. U.S. Department of Energy, “*Critical materials assessment,*” July 2023.
7. Gina Roos, “*Tantalum capacitor suppliers still wary about ability to meet demand,*” *EETimes*, June 2, 2001.
8. U.S. Department of Energy, *op. cit.*
9. Deloitte estimates and predictions based on our analysis of publicly available third party sources, reports, and articles including: IEA, *Global EV Outlook 2023*, 2023; Energy.gov data; data and research presented in Charged EVs magazine; and online retail prices of products and components.
10. Reuters, “*What are Gallium and Germanium and which countries are producers?*” *Reuters*, July 7, 2023.
11. Nora Foley and Brian Jaskula, “*Gallium – A smart metal,*” US Geological Survey, March 2013.
12. Ruth F. Schulte and Nora K. Foley, “*Compilation of Gallium resource data for Bauxite deposits*” US Geological Survey, 2014.
13. Jason Mitchell, “*China’s stranglehold of the rare earths supply chain will last another decade,*” *Investment Monitor*, April 26, 2022.
14. Vanessa Forti, Cornelis Peter Baldé, Ruediger Kuehr, Garam Bel, “*The Global E-waste Monitor 2020 – Quantities, flows, and the circular economy potential,*” United Nations Institute for Training and Research, 2020.
15. Wikipedia, “*Strategic reserves of the United States,*” accessed October 26, 2023.

16. Alexandra Alper, "[Russia's attack on Ukraine halts half of world's neon output for chips](#)", Reuters, March 11, 2022.
17. Kelly Pickerel, "[China's share of world's polysilicon production grows from 30% to 80% in just one decade](#)," Solar Power World, April 27, 2022.
18. Deloitte employee (semiconductor manufacturing expert), interview, July 2023.
19. Deloitte employee (manufacturing expert), interview, July 2023.
20. WEEE Forum, "[International E-Waste Day: 57.4M Tonnes Expected in 2021](#)," accessed October 3, 2023.
21. UN environment programme, "[How disposable tech is feeding an e-waste crisis](#)," November 21, 2022.
22. Alex Barshai, "[Precious metals recovery from e-waste](#)," emew clean technologies, blog post, December 28, 2022.
23. Carolyn Gramling, "[Earth's annual e-waste could grow to 75 million metric tons by 2030](#)," *ScienceNews*, July 2, 2020.
24. European Parliament News, "[E-waste in the EU: facts and figures \(infographic\)](#)," April 12, 2023.
25. ITU, "[Global E-waste Monitor 2020](#)," accessed October 3, 2023.
26. Barshai, *op. cit.*
27. Ibid.
28. World Health Organization, "[Children and digital dumpsites: E-Waste exposure and child health](#)," 2021.
29. Barshai, *op. cit.*
30. The White House, "[Fact sheet: Securing a Made in America supply chain for critical minerals](#)," press release, February 22, 2022.
31. Office of Fossil Energy and Carbon Management, "[Rare Earth Elements](#)," accessed October 26, 2023.
32. United States Geological Service, "[2022 Mineral commodity summaries – Rare Earths](#)," accessed October 26, 2023.
33. Reuters, "[Vietnam to up annual raw rare earths output to 2m tonnes by 2030](#)," Nikkei Asia, July 25, 2023.
34. Spot prices as of October 26, 2023. See: Shanghai Metals Market, "[Latest update in the SMM Indium/Germanium/Gallium Market](#)," accessed October 26, 2023.

35. Ephrem Joseph, “*Global semiconductor industry feels the heat as China plans gallium export controls,*” Proactive, July 20, 2023.

Acknowledgements

The authors would like to thank **Karthik Ramachandran, Dan Hamling, Jan Nicholas, Bobby Mitra, Mark LaViolette, Iain Nicklin,** and **Steve Watkins.**

Cover image by: **Manya Kuzemchenko**

Semiconductor sustainability: Chips take a smaller byte out of resources

Modern, new greenfield plants could help improve the industry scorecard, but manufacturing transformation can help both the greenfield plants and existing brownfield plants do better for energy, water, and process gas use.

ARTICLE • 11 MINUTE READ

The semiconductor manufacturing industry is notoriously volatile: As of fall 2023, it is in the middle of the seventh downturn since 1990.¹ Measured in US dollars, chip industry revenue is expected to decline 10% in 2023 and is expected to rise 12% in 2024.² Maybe unsurprisingly, when the chip industry shrinks, its use of energy, water, and process gases with high global warming potential (GWP), all go down. And when the industry grows, GWP goes up. Absolute measures of sustainability are often unhelpful, especially in an industry that—despite its volatility—is expected to grow, and is predicted to surpass US\$1 trillion in 2030,³ almost double the anticipated industry revenues of US\$515 billion in 2023.

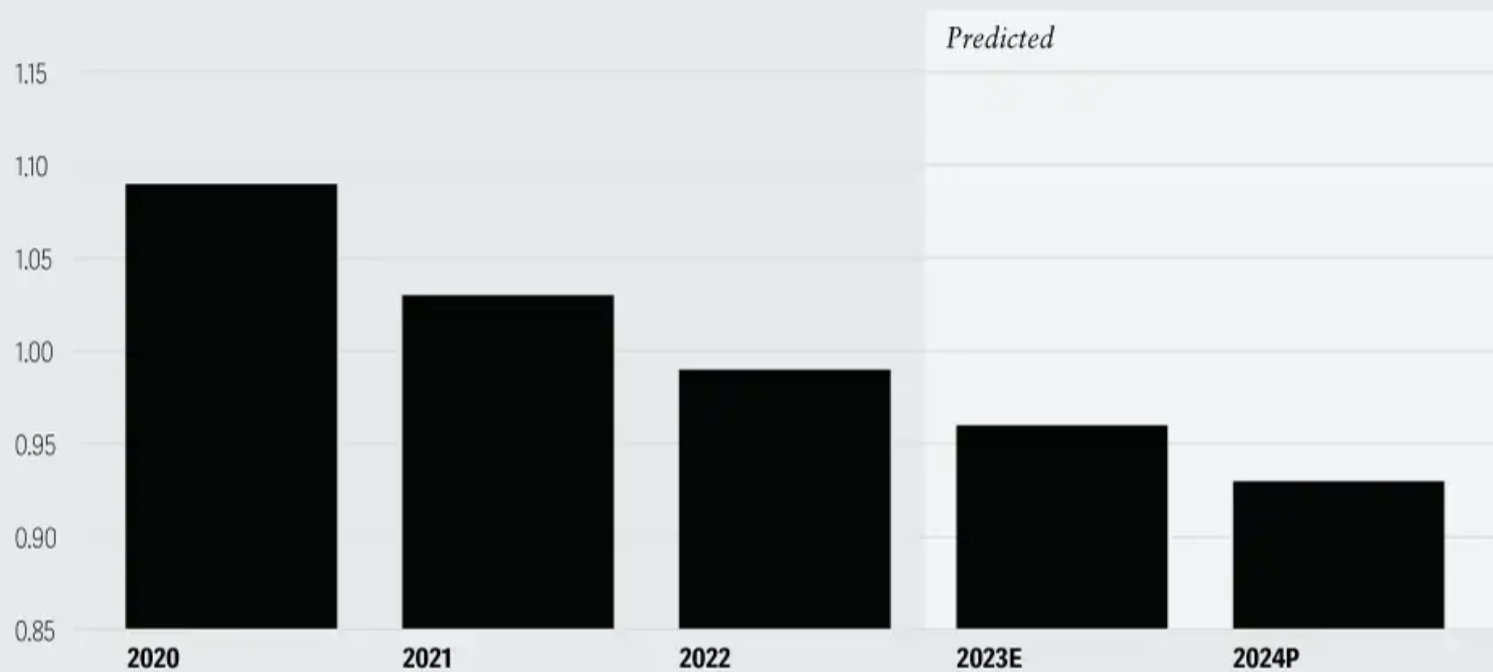
Instead, a better yardstick may be resource intensity: For every dollar of revenue, how much energy, how much water, and how much high GWP process gases are going to be used next year, compared with this year? Deloitte predicts that there will be a year-over-year decline in average water intensity (figure 1) and energy intensity in 2024, as well as declines in energy intensity and growth in the percentage of energy used by leading chipmakers that is renewable.

Figure 1

Resource intensity is a better way to measure industry progress on its climate goals

Water intensity, two-year moving average, 2020-2024

● Water intensity (kilo of water consumed/US\$ dollar of revenue)



Notes: E indicates estimated value; P denotes predicted value. Water intensity (two-year moving average) is computed as a ratio of aggregate water use (in millions of tons) to total semiconductor industry revenue. Actual water intensity for 2021 and 2022 dropped 6% and 4% YoY respectively, and we estimate a (relatively conservative) 3% YoY decline in 2023 and predict another 3% decline in water intensity for 2024, based on ongoing efforts by semi companies to improve water use and reduce intensity.

Sources: Data on water use gathered from publicly available corporate social responsibility (CSR) reports of 11 listed semiconductor companies (from across North America, Asia, and Europe; includes IDMs and foundries, and represents memory, logic, and analog) that have significant manufacturing footprint, globally. Revenue data based on WSTS annual reported numbers (for 2020-2022) and estimated/projected revenue (for 2023-2024) for the overall semiconductor industry.

Deloitte Insights | deloitte.com/insights

Some of the improvement in resource use intensity is likely part of an ongoing, decade-long trend as the industry has recently been trying to get better at semiconductor sustainability. Some of it may come from a significant growth in brand new chip plants (greenfields). As might be expected, the various equipment, tools, and processes in a new plant (all other things being equal) are often more sustainable than equivalent technologies that might be five, 10, or even 20 years old. That said, newer plants using advanced node technologies pose a sustainability challenge for the industry: Moving from a mature technology such as 28 nm manufacturing to advanced node manufacturing at 2 nm needs 3.5 times as much energy, 2.3 times as much water, and emits 2.5 times as much greenhouse gases, and this trend is expected to continue as processes become ever more advanced.⁴ Interestingly, even more of industry progress towards greater sustainability will come from implementing manufacturing transformation on those older plants (brownfields): Deloitte predicts that a full manufacturing transformation project can significantly reduce the intensity of energy, water, and process gas use over a multiyear period.

But there's still more to do

Looking at the lifetime energy and resource use of chips, manufacturing is only a part of the challenge. And the energy used by chips after they have been manufactured (for example, in power hungry data centers doing generative AI) can be a material factor. Equally, resource extraction, test and packaging, distribution, life cycle, and end of life are all important parts of the semiconductor sustainability equation.

While some chipmakers have set aggressive 2030 carbon zero and other sustainability targets, there is wide variation globally. In general, European Union-headquartered companies tend to have the most aggressive 2030 targets, and while some US-headquartered organizations have similarly ambitious targets, others are aiming for 2040 or later.⁵ Outside of Singapore, most Asian-headquartered chipmakers are setting targets for 2050 and beyond, or not at all.⁶ That said, in September of 2023, a leading Asian chipmaker moved up its commitment to use 100% renewable energy by 10 years, to 2040.⁷

The chip industry was responsible for approximately 0.2% of global carbon dioxide equivalent emissions in 2021.⁸ If it doesn't want to double to 0.4% by 2030 as the industry doubles in size, it should improve both greenfield plants and transform brownfield plants.

Energy consumption

Making chips is energy intensive. Melting silicon, lithographing with high-powered lasers, creating and maintaining vacuums, and endless cleaning takes a lot of electricity. Semi fabs can use up to 100 megawatt-hours per hour,⁹ the equivalent of over 80,000 typical North American homes. With that said, there are only about 500 open fabs at present.¹⁰ Although there's new plant construction between now and 2025, it's only 41 fabs globally.¹¹ Moreover, semiconductor companies have applied novel chip design techniques and advanced process technologies, for instance, using low-leakage transistors and low-power systems, and altering system power modes (such as off, idle, drowsy when select modules or IPs are not active). These have helped reduce energy needs across end-use devices and systems across industries, but with the expanding manufacturing footprint, the industry should look at other ways to optimize the use of resources and lower emissions.

The industry has two levers to pull in terms of reducing their energy use and associated carbon footprint. First, they're trying to become more energy efficient, but progress in that is often slow, especially as chip companies resort to increasingly advanced manufacturing technologies to push the limits of semiconductor manufacturing.¹²

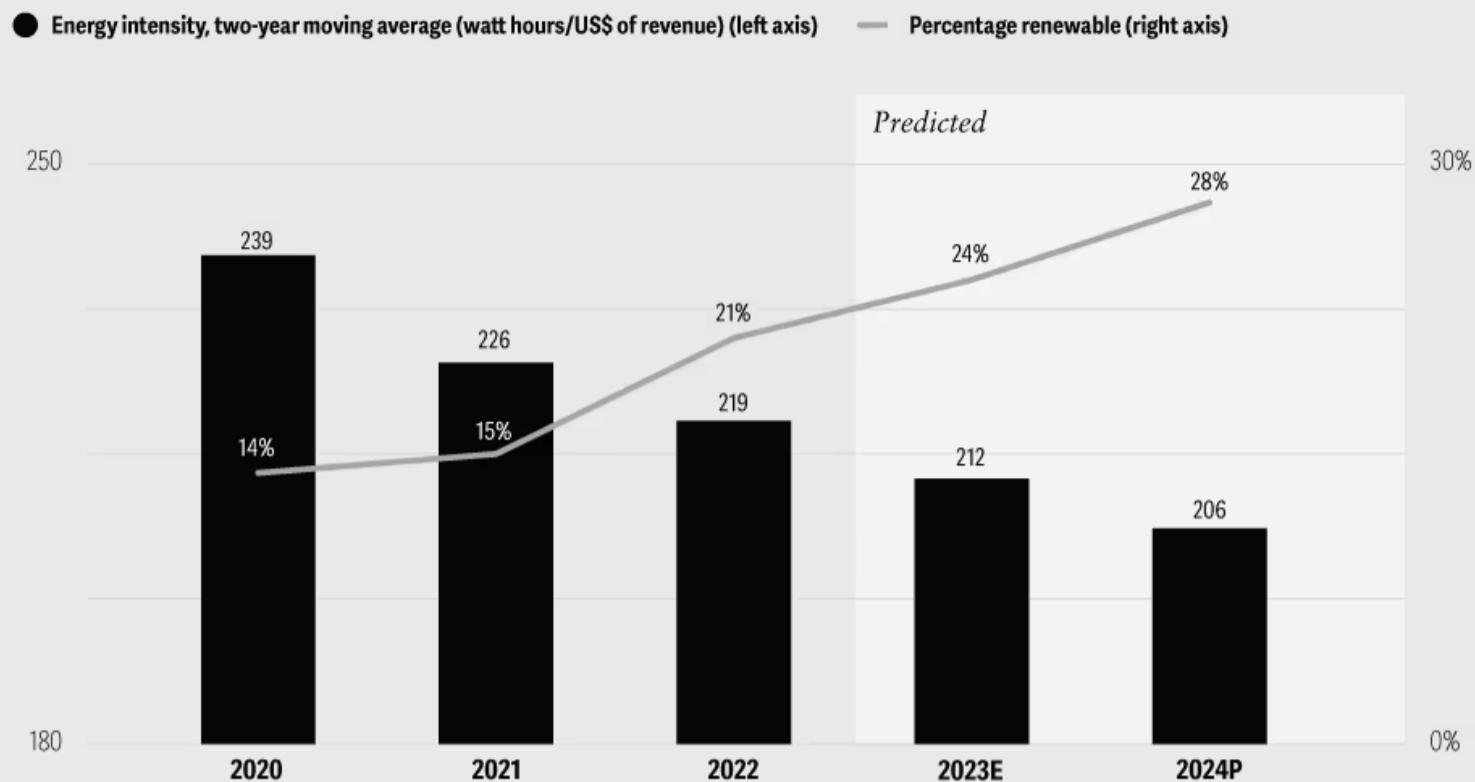
One thing they can consider is to make things (in addition to the actual chip manufacturing process) more energy efficient, faster: Leadership in Energy and Environmental Design buildings are mature technologies that the industry has relied on for a decade to improve its sustainability.¹³ They can also look to increase their use of renewable energy. As an example, 93% of a major America-headquartered chipmaker's energy needs were met by renewables as of its 2022–23 fiscal year.¹⁴ However, across the three largest players, renewables were only 28% of the energy mix in 2022, up five percentage points over 2021.¹⁵

Combining both levers has led to the fab portion of the industry reducing its energy intensity (watt-hours/dollar) from almost 240 in 2020 to 219 in 2022 and a predicted 206 in 2024 (figure 2).¹⁶ Further, the percentage of energy used that was renewable is predicted to rise even faster than energy intensity falls: Renewable energy is predicted to be 28% of the mix by 2024, double the percentage from 2020.

Figure 2

Renewables are key to improving energy intensity

Energy intensity and renewable energy use, 2020-2024



Notes: E indicates estimated value; P denotes predicted value. Energy intensity (two-year moving average) is computed as a ratio of total energy consumption (in gigawatt hours) to total semiconductor industry revenue. Renewable energy (RE) % use is computed as total RE consumption (in gigawatt hours) as a percentage of overall energy consumption across all sources. Actual energy intensity for 2021 and 2022 dropped 5% and 3% YoY, respectively, and we estimate a further 3% drop in 2023 and predict another 3% decline in energy intensity for 2024. This article highlights our research-based assumptions and rationale to back the prediction. Similarly, RE percentage use has increased by an average three percentage points (pps) every year between 2020 and 2022, and given the ongoing efforts of semi companies to grow their RE use, we predict RE percentage of total energy consumption will continue to go up further by 3 to 4 pps, annually, in 2023 and 2024.

Sources: For energy intensity, data on energy consumption was gathered from publicly available CSR reports of 11 listed semiconductor companies that have a significant manufacturing footprint, globally (from across North America, Asia, and Europe; including IDMs and foundries, and represents memory, logic, and analog products). Revenue data based on WSTS annual reported numbers (for 2020-2022) and estimated/projected revenue (for 2023-2024) for the overall semiconductor industry. RE % trend line is based on the aggregate data of 8 listed semiconductor companies (from across North America, Asia, and Europe, including IDMs and foundries; and represents memory, logic, and analog products) that reported their annual RE consumption and have a significant manufacturing footprint, globally.

Deloitte Insights | deloitte.com/insights

Water “usage”

The global chip industry used 264 billion gallons (about 1 trillion liters) of water in 2019.¹⁷ But although some water is lost to evaporation or other causes, depending on geography and chipmaker, all of that water is not “used”: A major America-headquartered chipmaker used 16 billion gallons of water in 2021 but returned 13 billion gallons of it (over 80%) and doubled their water savings from two years earlier.¹⁸ Chipmakers across Taiwan averaged 85% water recycling between 2016 and 2020.¹⁹

Most water use in the semiconductor industry is for the manufacturing process (76%), but a lot of water is also used for cooling towers (9%) and scrubbers (11%).²⁰ The biggest part of those scrubbers is in process gas abatement (see next section), which could present significant opportunities for water reduction: Switching the abatement systems to idle mode, when not actively processing, reduces water usage by 98%.²¹ Improvements could also be made in reducing both process water and water for cooling.

Process gases

The chip industry uses several gases, some of which have very high GWP. These are primarily fluorinated gases, namely perfluorocarbons (PFCs), hydrofluorocarbons, nitrogen trifluoride (NF₃), and sulfur hexafluoride (SF₆), used both in etching and cleaning,²² and nitrogen oxide gas that’s used in deposition and purging processes.²³ As an example, SF₆ has a GWP 23, 500 times higher than CO₂.²⁴ According to the US Environmental Protection Agency, historically, “anywhere between 10 and 80 percent of the fluorinated greenhouse gases pass through the manufacturing tool chambers unreacted and are released into the air.”²⁵

There are three main avenues for reducing the impact of these gases: process improvements/source reduction, alternative chemicals, and destruction technologies (often referred to as ‘abatement’).²⁶ In general, much of the low hanging fruit opportunities for improvement in the first category have likely been made, but progress continues to be made at the margins, and this is where manufacturing transformation can potentially have an impact. There have been some successes in finding alternatives: A number of PFCs were replaced by NF₃, which, while still problematic, was an improvement.²⁷ However, finding and qualifying for the manufacturing processes alternative gases can be a slow process, and only a few breakthroughs appear imminent, such as G1.²⁸ Abatement remains the workhorse here. Capturing and destroying (usually by burning or conversion) as much of the high GWP process gases is the key. As an example, abating 99% of NF₃ is likely both achievable and better than 95%.²⁹ By and large, process gases are not often reused or recycled, due to issues around purity, cost, and ability to be integrated into the sub-fab physical footprint.³⁰

PFAS

Although not the focus of this prediction, the chip industry also uses or produces many perfluoroalkyl and polyfluoroalkyl substances (PFAS). In 2023, a restriction proposal on their use was made by the chemical authorities in five EU countries, calling for their eventual elimination. This is expected to come into force in 2025–2026.³¹ The semiconductor industry association SEMI, in response to both the EU proposal and various existing or proposed US regulations, is looking at reducing the use of PFAS and finding alternatives.³²

Manufacturing transformation

Today’s chip fabs are like a forest with trunks, leaf canopies, and vast roots. In addition to all the machines and clean rooms at surface level, there’s a canopy of pipes and ducts overhead, and an even more complex set of pumps, abatement systems, scrubbers, and transformers underneath the floor (the sub fab). This ecosystem has many parts that can be hard to access or monitor in real time, and by modeling, adding connected sensors everywhere, and continuously monitoring the use of energy, the use of water and process gases can be made more efficient. Enabling technologies include digital twins, generative AI, and private 5G networks. Leaks can be detected, and systems can be idled or powered down when not in use. It can cost hundreds of millions of dollars to take a decade-old chip plant and transform it, but the gains in sustainability, as well as lower costs and higher efficiency are likely worth it both for the bottom line and the planet.

In general, this idea is integral to the 6D building information modeling (BIM) concept, which encompasses the inclusion of a sixth dimension, specifically the performance and sustainability of the building. BIM is the process of creating and managing information on a construction process, and by incorporating this additional dimension, the virtual model is designed to depict not just the physical, temporal, and cost-related aspects of the building, but also its environmental and social implications. Consequently, this could become a valuable resource for evaluating how the building affects its surroundings and community, and for recognizing opportunities for enhancement.³³

Gross versus net: Chips are greener than you think

Making a trillion dollars' worth of chips in 2030 is going to have an environmental impact in terms of energy, water, and greenhouse gases. The industry is working to make that impact as low as possible, but it may be worth noting that chips could help enable sustainability gains. Road and air travel can be avoided by using video conference technology enabled by chips, computers are the standard tool for discovering and designing drugs,³⁴ hyperscalers are investing in more renewable energy sources to power chips in data centers,³⁵ and food can be grown more sustainably (see our [AgTech prediction](#)). Some might argue that the environmental benefits of chips are greater than the environmental costs of making semiconductors. On the other hand, others believe that: 1) This might be true of chipmaking itself, but factoring in resource extraction, test and packaging, distribution, life cycle and end-of-life considerations paints a less beneficial picture;³⁶ and 2) The Jevons paradox (aka rebound effect) could lead to worse sustainability outcomes. If chips get made more efficiently, we will just make and use more of them, potentially leading to a worse overall outcome.³⁷

The bottom line

For semiconductor companies, environmental awareness is its own reward, being more sustainable is good, and is increasingly being required by what is sometimes called the 5Cs framework: capital (investors,) compliance (regulators,) constituents (such as employees), communities, and creativity (innovation). But being more sustainable is often also better for reducing costs, can help in the competition for semiconductor talent, and can reduce semi supply chain vulnerabilities.

Dedicated environmental, social, and governance (ESG) funds are US\$8 trillion today and are predicted to be as much as US\$30 trillion by 2030.³⁸ Even outside of dedicated funds, asset managers are increasingly using ESG screening tools when building their portfolios, and this includes chipmakers. One additional challenge is from regulators. Currently, most public companies report scope 1 and scope 2 emissions (direct and indirect energy use) but not scope 3 (the supply chain, both upstream and downstream). It's possible that regulators, both in Europe and the United States, may require scope 3 disclosures. As is well known after the pandemic chip shortage, almost every industry has chips in their supply chain, and chipmakers would likely be asked by the customers to have the best sustainability profile possible.

Energy, gases, and (usually to a lesser extent) water are expensive, and becoming more so. Reducing these input costs can have a positive effect on the bottom line.

Equally, semiconductor companies are attempting to build greenfield plants both in traditional strongholds such as Asia, but increasingly in the United States and Europe.³⁹ And there is a global competition for talent: The semi industry is competing for scarce technically skilled talent with multiple other industries and should have a positive environmental track record.⁴⁰ Workers, especially younger workers, prefer working for companies with the best sustainability track records. As a 2023 Deloitte survey of Gen Z and Millennials found, "one in six have already changed jobs or industries due to climate concerns, with another quarter planning to do so in the future."⁴¹

Finally, reducing reliance on energy and water significantly expands the areas where chip plants can be located. Droughts have recently affected chipmaking in multiple regions, and as one headline put it, "No water, no microchips."⁴² Equally, chipmakers in Asia and the United States were subject to power outages due to climate change, and chip plants are exceptionally reliant on uninterrupted power.⁴³ Raw material dependencies (see [prediction on raw material and supply chains](#)) and transport disruptions and investing in semiconductor sustainability can offer material supply chain resilience benefits.

By **Duncan Stewart**
Canada

Dr. Bobby Mitra
United States

Karthik Ramachandran
India

Christie Simons
United States

Endnotes

1. Chris Richard, Dan Hamling, Duncan Stewart, and Karthik Ramachadran, *Five fixes for the semiconductor chip shortage*, Deloitte Insights, December 6, 2021.
2. World Semiconductor Trade Statistics (WSTS), “*WSTS semiconductor market forecast spring 2023*,” news release, accessed November 2, 2023.
3. Vyra Wu, “*Global semiconductor market to exceed US\$1 trillion in 2030, at CAGR of 7%, says DIGITIMES Research*,” *DIGITIMES Asia*, January 10, 2023.
4. Marie Garcia Bardon and Bertrand Parvais, “*The environmental footprint of logic CMOS technologies*,” *EE Times*, December 14, 2020.
5. Sarah Barry James, Stefan Modrich, and Sydney Price, “*Path to net-zero: US chipmakers balance growth vs. going green*,” S&P Global Market Intelligence, June 13, 2022.
6. Ibid.
7. Cheng Ting-Fang and Katherine Creel, “*Taiwan Semiconductor Manufacturing Company moves up 100% green energy goal by 10 years*,” *Nikkei Asia*, September 15, 2023.
8. In 2021, global scope 1 and scope 2 emissions for the semiconductor industry were estimated at 76.5 megatons carbon dioxide equivalent. Global emissions for 2021 were 37.9 gigatons carbon dioxide equivalent, so semiconductor emissions were 0.2%. See: Maxime Pelcat, *Green house gas emissions of semiconductor manufacturing in 2021*, University of Rennes, June 1, 2023.
9. Christel Galbrun-Noel, “*How to improve power reliability for semiconductor fabs*,” Schneider Electric blog, November 15, 2021.
10. 492 plants based on *Wikipedia’s consolidated list of semiconductor fabrication plants*, accessed September 14, 2023.
11. SemiMedia, “*41 new fabs to be added globally from 2022 to 2025*,” November 4, 2022.
12. Bardon and Parvais, “*The environmental footprint*.”
13. Deloitte analysis of sustainability reports from multiple semi companies.
14. Intel, *2022–23 corporate responsibility report*, accessed September 14, 2023.
15. Analysis based on data reported in publicly available corporate sustainability reports of select semiconductor companies.
16. See source and methodology notes for figure 2.

17. Shannon Davis, “[Water supply challenges for the semiconductor industry](#),” *Semiconductor Digest*, October 24, 2022.
18. Editorial, “[Intel achieves net positive water in three countries](#),” Intel, July 13, 2022.
19. Davis, “[Water supply challenges](#).”
20. Ibid.
21. Ibid.
22. US Environmental Protection Agency (EPA), “[Semiconductor industry](#),” accessed September 14, 2023.
23. Generon, “[Using nitrogen gas in the semiconductor manufacturing process](#),” accessed September 14, 2023.
24. Mike Czerniak, “[The time is now: Sustainable semiconductor manufacturing](#),” *Semiconductor Digest*, November 2021, pp: 16–19.
25. US EPA, “[Semiconductor industry](#).”
26. Ibid.
27. Mike Czerniak, “[The time is now: Sustainable semiconductor manufacturing](#),” *Semiconductor Digest*, November 2021.
28. Samsung is developing G₁ (or G1) as one of the alternative gases with low global warming potential and replacing perfluorocarbon gases in some products. See: Samsung, [A journey towards a sustainable future: Samsung Electronics sustainability report 2023](#), accessed September 14, 2023.
29. Marie Garcia Bardon and Bertrand Parvais, “[The environmental footprint of logic CMOS technologies](#),” *EE Times*, November, 2020.
30. Chris Bailey, “[Recovery and recycling of process gases: What are the options?](#),” *Semiconductor Digest*, accessed November 2, 2023.
31. Eurofins Scientific, “[Perfluoroalkyl and polyfluoroalkyl substances \(PFAS\) restriction proposal: The largest substances ban project ever in Europe](#),” accessed September 16, 2023.
32. SEMI, “[PFAS Explainer: The semiconductor industry responds](#),” accessed September 16, 2023.
33. The BIM Engineers, “[From 3D BIM to 7D BIM](#),” June 8, 2023.
34. Wikipedia, “[Drug design](#),” accessed October 25, 2023.
35. Rick Johnston, “[How data centers can use renewable energy to increase sustainability and reduce costs](#),” Device 42, Inc., April 5, 2023.

36. Interview with Deloitte semiconductor sustainability practitioners, July and August, 2023.
37. Jaume Freire González, “*The Jevons paradox and rebound effect: Are we implementing the right energy and climate change policies?*,” The OECD Forum Network, September 22, 2022.
38. Jag Alexeyev, *ESG and sustainable investment outlook: US\$30 trillion by 2030 on the way to net-zero*, Broadridge Financial Solutions, Inc., 2021.
39. Deloitte, “*2023 semiconductor industry outlook*,” accessed November 2, 2023.
40. Karen Weisz, Christie Simons, Brandon Kulik, Duncan Stewart, and Teresa Lewis, “*The global semiconductor talent shortage*,” Deloitte, accessed November 2, 2023, p. 7.
41. Deloitte, “*2023 Gen Z and Millennial survey*,” accessed November 2, 2023.
42. Emanuela Barbiroglio, “*No water no microchips: What is happening in Taiwan?*,” *Forbes*, May 31, 2021.
43. Analysis based on publicly available information sourced from *EDN* (2001), *Silicon Expert* (2021), and *CNBC* (2022) showed how power outages and disruptions affected fab operations and chip production at different points in time in the United States and Asia.

Acknowledgements

The authors would like to thank **Jan Nicholas, Dan Hamling, Steve Watkins, Iain Nicklin, Nicholas Wyver, Negin Rood,** and **Sathiya S.**

Cover image by: **Manya Kuzemchenko**

Dialing down the carbon: Telco sustainability surges on the back of four new trends

Telecommunications companies can reduce their carbon footprint by shutting down copper wire and 3G wireless networks, changing their field service fleets to EVs, and switching to 5G radio gear that has a “snooze button.”

ARTICLE • 8 MINUTE READ

Deloitte predicts that telcos worldwide will be able to reduce their carbon footprint by 2%, or 12 million tons of carbon dioxide equivalent (CO₂e) in 2024 and do the same again in 2025.¹ The telco industry is not a particularly egregious emitter: It connects over 95% of the planet’s people over wired and wireless networks and yet is directly responsible for only about 1.6% of 2022’s global emissions of 37 billion tons CO₂e, or about 600 million tons.² That said, telcos are publicly listed, consumer-facing, for-profit companies, and improving sustainability can help comply with regulations, attracts subscribers, and helps the bottom line.

Some of that reduction is expected to come from strategies that telcos have been using for years, such as buying more renewable energy (some telcos are already at 100% renewable energy, but many more expect to get 50% of their energy from renewables by 2025),³ running data centers more efficiently (natural cooling can reduce energy use by 30%),⁴ getting consumers to keep phones longer,⁵ and encouraging the growth in the secondhand phone market—a US\$50 billion run rate industry in 2023.⁶ None of this is new.

What is new—or at least becoming more common in 2024—are four other solutions.

Copper wireline switch off

Switching copper off can reduce carbon footprint while potentially saving billions of dollars annually for carriers. Readers who are old enough will recall that landline telephones worked even when the power went off. That’s because the twisted pair of copper wires that ran from the phone company central office into homes not only delivered the audio signal for voice communication, but also a 48-volt direct current to power the phone and the ringer. As more homes have moved to fiber optic cable or cut the landline via fixed wireless access, the percentage of consumers who need to connect over copper, either for voice or data connection over DSL, has been dropping for decades. Not only is fiber easier to maintain and requires fewer field service calls than copper, but it also uses up to 80% less energy, both on the twisted pairs to homes and central office exchanges.⁷

Singapore has already switched off its copper network, but we expect copper retirement to accelerate in 2024, with networks in Asia/Pacific (Japan, Australia, New Zealand) and 10 markets in Europe (Norway, Sweden, Spain, Portugal, Estonia, France, Italy, the United Kingdom, and Switzerland) moving forward on shutting down copper.⁸ By 2025, at least some countries will see close to 100% switch off, and many more will be at 25% to 50%.⁹

3G wireless switch off

Speaking of networks that now have few users and use a disproportionate amount of power ... Remember 3G wireless networks? In addition to being able to use that 3G spectrum for 4G or 5G (aka refarming), there are significant efficiency and carbon gains to shutting 3G networks down: One report estimates that shutting down 3G reduces energy costs by 15%.¹⁰

Introduced in 2001, 3G peaked in many markets around 2012, and has gradually been replaced by 4G and now 5G networks offering higher speeds and vastly superior efficiency/bits transmitted. 3G networks are still heavily used in parts of the developing world: Smartphone users in Angola and Ethiopia spent 40% of their time on 3G in 2023.¹¹ But in other markets, 3G was shut down years ago, including all of Taiwan from 2017 to 2018. In the last six years, 3G has been switched off by 58 operators across 22 countries (averaging about 10 operators and four countries annually), and we expect to see shutdowns accelerate to 15 operators across six countries in 2024 and a further 26 operators across 10 countries in 2025.¹²

Making field service fleets greener

Telcos have big fleets of commercial vehicles and making those more sustainable can have a big effect. A commercial vehicle emits up to 21x as much carbon dioxide as a light-duty passenger vehicle, and although commercial vehicles represent only 20% of all vehicles, they make up 60% of road transport emissions.¹³

For a single operator, maintaining tens of thousands of cellphone towers and connecting millions of households and businesses requires a massive field service fleet of cars, trucks, and vans. As an example, one UK operator has a fleet of 33,000 vehicles, which is the second largest commercial fleet in the United Kingdom and is responsible for over two-thirds of its direct (scope 1) emissions.¹⁴ As charging networks get larger, and as more electric vehicles (EVs) such as vans and trucks are being manufactured, the pace of electrification is accelerating. A French operator had 3,000 EVs in its fleet by the end of 2022 with plans to double that by 2025, and also took delivery of the first hydrogen-powered van in 2023.¹⁵

Next-generation 5G gear gets greener

5G radios can use a lot of power, but new technologies let them go to sleep when not in use, saving telcos gigawatt-hours of electricity and megatons of CO₂e annually.¹⁶

There are a host of technologies that allow 5G to do more than 4G, one of which is an active antenna unit (AAU). Instead of a passive antenna, 5G AAUs are capable, using a technology called massive MIMO, of beamforming, which allows for much higher speeds, capacities, spectral efficiency, and lower cost per megabit, but at the cost of increased power consumption.¹⁷ Most of the power draw from the various baseband units (BBUs) and remote radio units (RRUs) is the same for 4G and 5G. The big changes in capability and power consumption come from AAU drawing 4.2 kilowatts, which is over 60% of the power consumption of a 4G base station.¹⁸ With additional frequencies supported, power consumption can climb even further, up to 20 kilowatts. This not only requires improved power supplies—costing billions of dollars in upgrading or retrofitting for a large network¹⁹—but also has a negative effect on a telco's carbon footprint.

If fully utilized, that 5G power consumption could deliver many millions of bits of data to thousands of users, and 5G networks are 90% more efficient than 4G when measured in bits per kilowatt.²⁰

But when not fully utilized, 5G is using a lot of power and may not be serving enough customers to justify it. And first-generation 5G equipment doesn't do well at going into sleep mode (also known as low power mode or deep idle mode) and adapting its power consumption to the usage on a flexible basis. There are nonflexible solutions: Some telcos have resorted to literally turning off their 5G networks at night when utilization is low and forcing customers to rely on the legacy 4G networks only.²¹

That is about to change in 2024, with all of Ericsson, Nokia, Samsung, and Huawei, (which combined make up around 87% of the global market for 5G radio access networks)²² introducing 5G RRUs, BBUs, and AAUs that use much less energy, often powered by AI algorithms and new semiconductors.²³ Across the board, they claim that power consumption for second generation 5G gear is 20% to 50% lower than the previous generation,²⁴ and up to 94% lower during periods of low traffic.²⁵ There are also versions that are aimed at the 5G Internet of Things (IoT) market, that have a reduced capability and use much less power than full 5G, and even less power than equivalent 4G solutions.²⁶

At a high level, 5G networks are often well deployed in most major markets, and most operators are not likely to rip out older 5G radios and antennas for newer sustainable ones. For regions that are only just beginning to deploy 5G, the new equipment should be used, and as older gear fails over time, it can be replaced with more efficient equipment.

The biggest contributor to improving the environmental footprint of 5G radios in 2024 and 2025 will likely be in 5G deployments that have been slow up until now. As telecom infrastructure companies intensify their 5G coverage with more towers and antennas, they can likely buy more energy-efficient gear.

The bottom line

It isn't just telcos and their twisted copper pairs that can be switched off and replaced with more efficient fiber optics. Cable companies use coaxial cable to deliver their signals for TV and internet. It's unpowered, so shutting it off saves less energy. However, according to one study, moving to passive optical fiber to the home network uses about 9% less electricity than a coaxial cable solution (for comparison, the fiber solution uses 37% less electricity.)²⁷

Also, although 5G networks have the most power-hungry antenna technology, they are currently often underutilized and architecturally are the easiest to power down when underutilized due to their advanced core (the 5G New Radio standard).²⁸ That isn't to say that sleep mode technology can't be used for 4G, 3G, or even 2G networks. The energy savings are not quite as large, and as 2G and 3G networks are shut off, most of the focus is expected to be on 4G.

Equally, although the focus may be on the 3G shutdown, there are also still 2G networks running, primarily for machine to machine or IoT communications, such as smart meters.²⁹ They don't use as much power as 3G, but they are also expected to be shut down over the next decade. Some countries are moving more quickly on 2G shutdowns (France by 2025) while others are slower (the United Kingdom by 2033).³⁰

Telcos can do a lot to reduce their scope 1 direct emissions (electrifying their fleets), or scope 2 purchased power emissions (5G sleep mode, copper and 3G shutdowns), but scope 3 indirect value chain emissions are likely a more significant hurdle. Handsets are a big part of telcos' purchased goods and services. In the United States, carriers are responsible for 75% of all smartphones purchased each year (it is much lower in Europe), and each new smartphone has a footprint of 85 kilos of CO₂e.³¹

As we said in our *2022 TMT Predictions* on making smartphones more sustainable, "Because manufacturing accounts for almost all of a smartphone's carbon footprint, the single biggest factor that could reduce a smartphone's carbon footprint is to extend its expected lifetime."³²

Carriers can help in that, but it's tricky, as they should work with their customers to achieve these goals and can only do so much on their own. They may want to encourage customers to keep phones longer: New European Union regulations support that, requiring components and software support to last 5 years.³³ Carriers can not only offer more attractive SIM-only plans, but also lower new phone subsidies; offer refurbished phones, incentives for repair services, and rewards for trade-ins; and dispose of phones more sustainably—preferably recycling phones rather than contributing to the growth in e-waste.

By **Duncan Stewart**
Canada

Cindy Varga
Canada

Ariane Bucaille
France

Kevin Westcott
United States

Endnotes

1. Deloitte analysis of the combined effect of copper switch off, 3G switch off, fleet electrification, and reduced power 5G radios.
2. Iain Morris, “*The carbon-belching ICT sector must do better on the environment*,” Light Reading, September 2, 2022.
3. Deloitte analysis of telcos in North America, Europe, and Asia.
4. Orange, “*Two new Orange data centers supporting growth of usages and controlling energy impact*,” June 1, 2022.
5. Paul Lee, Cornelia Calugar-Pop, Ariane Bucaille, and Suhas Raviprakash, *Making smartphones sustainable: Live long and greener*, Deloitte Insights, December 1, 2021.
6. Hana Anandira, “*Secondhand smartphone sales top US\$13B*,” Mobile World Live, July 13, 2023.
7. Grace Langham, “*Decommissioning legacy networks will be key to reducing operators’ energy usage*,” Analysys Mason, August 17, 2022.
8. Deloitte analysis of multiple public sources.
9. Deloitte analysis of multiple public sources.
10. ABI Research, “*Purchasing renewable energy removes CO2 emissions equal to 20 million barrels of oil a year for leading telco operators*,” April 21, 2022.
11. Robert Wyrzykowski, “*African smartphone users see a diverse mobile experience across the continent*,” Opensignal, July 27, 2023.
12. Deloitte analysis of “*3G phaseout*” section from Wikipedia, accessed October 11, 2023.
13. Anders McKenzie, Joe Solly, and Samira Dadgar, “*Electrified fleets pave the way to emissions reduction*,” Deloitte Canada, accessed November 16, 2023.
14. Andy Wales, “*Three ways to get the UK’s electric vehicle revolution on the road*,” BT Group plc, October 19, 2021.
15. Orange, “*Environmental commitment: Net-zero carbon by 2040*,” accessed October 11, 2023.
16. Ericsson, “*Rogers tackles network energy consumption*,” accessed October 11, 2023.
17. TELCOMA Global, “*Massive MIMO—benefits and challenges*,” accessed November 7, 2023.
18. Chen Dongxu, “*5G power: Creating a green grid that slashes costs, emissions, and energy use*”, Huawei, July 2020.

19. Ibid.
20. Viavi Solutions, “[What is 5G energy consumption?](#),” accessed October 11, 2023.
21. Masha Borak, “[5G towers are consuming a lot of energy, so China Unicom is putting some of them to sleep overnight](#),” *South China Morning Post*, August 27, 2020.
22. David Manners, “[Huawei, Nokia, Ericsson to take 74.5% of 5G market this year](#),” *Electronics Weekly*, August 2, 2022.
23. Deloitte analysis of 2023 announcements from all four companies, with products being available in the second half of 2023 or early 2024.
24. Ibid.
25. Ericsson, “[Strengthened networks portfolio boosts Ericsson’s drive towards net-zero emissions](#),” press release, February 9, 2023.
26. [Ericsson](#), “[Ericsson rides the next wave of 5G with new reduced capacity solution](#),” June 19, 2023.
27. Europacable, [Fibre: The most energy-efficient solution to Europe’s bandwidth needs](#), July 2022.
28. Tim Hatt and Emanuel Kolta, [5G energy efficiencies: Green is the new black](#), GSM Association Intelligence, November 2020.
29. Tereza Krásová, “[UK telcos prepare to turn off 3G to boost energy efficiency](#),” *Light Reading*, July 27, 2023.
30. Ibid.
31. Lee, Calugar-Pop, Bucaille, and Raviprakash, [Making smartphone sustainable](#).
32. Ibid.
33. European Commission, “[Designing mobile phones and tablets to be sustainable—ecodesign](#),” accessed October 11, 2023.

Acknowledgements

The authors would like to thank **Karthik Ramachandran**, **Ankit Dhameja**, and **Paul Lee**.

Cover image by: **Manya Kuzemchenko**

Regulations take effect: ESG reporting software sales are expected to soar in 2024

Pushed by investors, regulators, and employees, many more companies will likely systematize their ESG tracking and reporting with standardized software tools.

ARTICLE • 5 MINUTE READ

Sales of software solutions that help companies track and report on environmental, social, and governance (ESG) metrics will likely surpass US\$1 billion this year, as EU and US reporting regulations—along with reporting requirements in Asia, Australia, and the UK—take effect, and as more investments require ESG disclosures.¹ Forecasts for ESG reporting software estimate a CAGR from 19% to 30% over the next five years,² and Deloitte predicts the tipping point will be reached in 2024, with growth accelerating to more than 30% and revenue climbing from under an estimated US\$800 million in 2023 to just more than US\$1 billion in 2024 (figure 1).

Emerging challenges and regulations will likely drive greater ESG reporting adoption

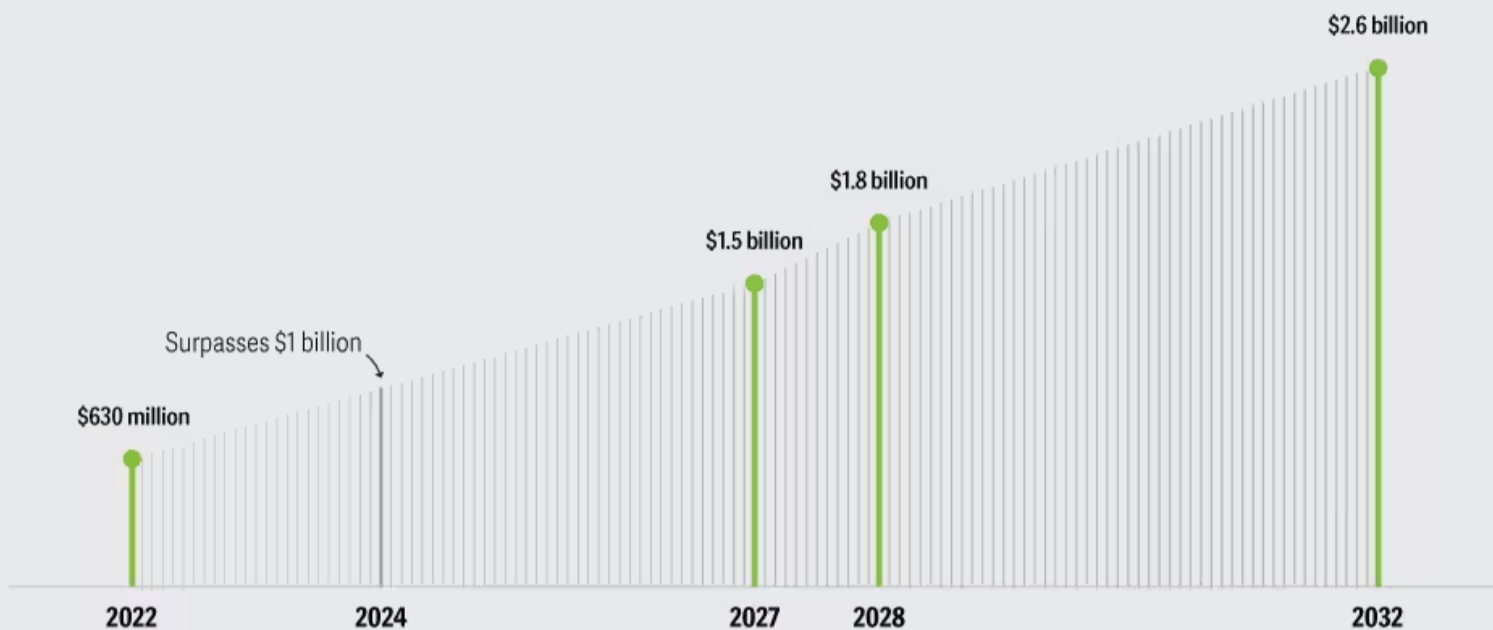
ESG reporting isn't new, per se. Many large, global companies release ESG or corporate sustainability reports voluntarily each year, highlighting their commitment to reducing carbon emissions and achieving sustainability goals.³ These reports are seen as important for investors and customers, plus potential employees who want to align with “virtuous” companies and support sustainable growth.⁴

In Deloitte's 2023 Gen Z and Millennial survey, for instance, 50% of Gen Z respondents said they are pushing their employers to drive change on environmental issues, and 42% said they'd switch jobs due to climate concerns.⁵

Figure 1

ESG tracking software sales are expected to see a huge jump, with a CAGR ranging from 13% to 30% over the next five years

Estimated ESG reporting software growth based on market size (US\$)



Source: Pitchbook data, 2023.

Deloitte Insights | deloitte.com/insights

A challenge with current ESG tracking and reporting approaches, however, is that they may not necessarily be consistent or accurate.⁶ Different ways of calculating emissions and accounting for carbon use may yield vastly different results, especially as companies are expected to move beyond “Scope 1 and Scope 2” (the emissions they generate and the emissions of the utilities they consume) and consider Scope 3 (the emissions generated by their upstream supply chains and downstream value chains) in evaluating their impact.

In more recent years, many companies have also been scrutinized for their adherence to social values such as diversity, equity, and inclusion (DEI), preservation of biodiversity, and ethical practices. Again, with dozens of voluntary frameworks to choose from, companies employ a wide range of inputs to arrive at statistics that often paint them in a positive light.⁷

In fact, the top-cited barrier to ESG adoption, according to a global survey of business leaders, is a lack of consistent and standardized data.⁸ Deloitte expects that new regulatory requirements, along with outputs designed to those specifications, will establish de facto standards and drive adoption. These regulations—primarily in the EU and United States, and in the United Kingdom, Hong Kong, New Zealand, and other countries—take effect in the 2024–2025 timeframe.⁹

The EU’s Corporate Sustainability Reporting Directive (CSRD) is an update to the 2014 Non-Financial Reporting Directive (NFRD), and it expands the number of companies required to provide sustainability disclosures from around 12,000 to more than 50,000.¹⁰ It also imposes requirements around “double materiality,” which means companies must report the impacts that ESG efforts have on their own businesses and the impacts they’ll have on the environment, human rights, social standards, and sustainability-related risk.¹¹ CSRD applies to multinational companies whose EU activities generated more than 150 million euros in annual turnover over the last two years. These European branches may have to provide consolidated reporting on their parent company’s activities to the EU, as well.

In the United States, proposed rules from the Federal Acquisition Regulatory Council require certain federal contractors to disclose their greenhouse gas (GHG) emissions and climate-related financial risk, and to set science-based targets to reduce their emissions.¹² California’s recently enacted Climate Accountability Package imposes Scope 3 reporting requirements on any company with revenues above US\$1 billion that does business in California.¹³

The SEC is working on ESG reporting requirements for many registered funds and investment advisers. Its proposed rules, “Enhanced Disclosures by Certain Investment Advisers and Investment Companies About Environmental, Social, and Governance Investment Practices,” are designed to promote consistent, comparable, and reliable information for investors who want to review a fund’s or adviser’s incorporation of ESG factors.¹⁴

As enacted, companies that are subject to the EU CSRD rules must file reports in 2025, reflecting data for FY24. The rules expand reporting requirements to mid-size and smaller companies by 2026, as well, as increasing the total addressable market and revenue opportunities.¹⁵

Finally, CSRD and proposed SEC compliance will also require third-party assurance of ESG reports. Auditors will likely have a larger role in guiding companies around ESG frameworks, standards, disclosures, and other opportunities. Taken together, these regulatory activities suggest that the time to implement a robust and comprehensive ESG tracking and reporting software solution is now.

Setting targets beyond Scope 3

In a 2021 Open Compliance and Ethics Group survey, only 9% of companies said they use ESG tracking software, which further underscores the potential for rapid uptake.¹⁶ These software solutions take a variety of approaches to calculating Scope 3 and other social impact metrics. How much carbon footprint each of a company’s suppliers and distributors adds, for instance, depends on where these entities are located, the time of year, and the efficiency of their own energy solutions.¹⁷

ESG tracking software takes company data as an input, along with vast data libraries, indexes, estimate tables, and sometimes AI to calculate carbon usage, ethics and corruption levels, and other social and environmental impacts. The size and accuracy of these libraries are one area where solutions attempt to differentiate from one another. Some incorporate HR data, as well, to report progress on DEI and economic parity goals.

The solution-provider market is crowded, comprising pure-play ESG analytics players, ERP companies that have acquired and bolted-on ESG functionality, professional services companies, and tech giants. Pricing varies widely with different structures based on the types of emissions reporting required. With close to 50,000 companies now under CSRD compliance mandates, the potential for US\$100 million+ in 2024 sales seems likely.¹⁸ A flurry of M&A activity in the space also indicates that the market is anticipating a wave of adoption in the near term.¹⁹

The bottom line

ESG tracking and reporting doesn’t have to be about ticking a regulatory box; it may help reduce tangible risks and create opportunities for companies. As accurate and actionable ESG insights flow into the operation, new business models may emerge, which can lead to new revenue streams. Deloitte’s sustainability action report found that 62% of executives surveyed are prepared or currently undertaking extensive preparations for the expected increase in reporting requirements.²⁰

The solution that companies choose should provide hooks into back-end systems and automatic data collection, plus robust analytics for multiple data sources. Its report functionality should conform to CSRD, SEC, and other regulatory and voluntary frameworks—in addition to providing customizable reports that may be integrated to reflect business goals and processes. These solutions should also be configurable based on the size and industry of the company and its global presence, tailoring materiality guidance and ESG reports to the various regulatory requirements.

It’s important to strike a balance between ESG goals and competitive innovation, and the two aren’t mutually exclusive. A clear strategy and action plan that does not compromise stakeholder returns and puts long-term ESG goals in perspective is possible. With stakeholder engagement and a comprehensive view of operational and reputational risks, ESG tracking can create a competitive advantage. Government incentives and credits represent another critical path for mitigating short-term ESG expenses.²¹

Also, keep in mind that ESG tracking and reporting can help to attract investments and fuel growth. Nearly 80% of respondents to a Deutsche Bank survey said they want investments to be associated with companies that are positively impacting the world,²² and Deloitte estimates that by 2024, half of all professionally managed global investment assets—representing trillions of dollars—will have ESG metrics reported.²³ Investors seem to recognize the long-term value creation that ESG-focused companies bring in terms of risk mitigation, decarbonization, enhanced reputation, and stronger growth potential.

By **Michael Steinhart**
United States

Gillian Crossan
United States

Ariane Bucaille
France

Ankit Dhameja
India

Endnotes

1. Deloitte, “*ESG mega trends*,” October 26, 2022. See also - Zacharias Sautner, “*The effects of mandatory ESG disclosure around the world*,” Harvard Law School Forum on Corporate Governance, May 10, 2021.
2. Verdantix, “*Market size and forecast: ESG reporting software solutions 2021-2027 (Global)*,” January 13, 2023. See also - PitchBook market size estimate data, July 2023
3. One organization that validates ESG targets for large global companies is the *Science Based Targets initiative*. STBi participation is voluntary and requires consistent ESG tracking and reporting. Companies that fail to meet their commitments may experience negative brand impacts and reduced consumer confidence.
4. Soyoung Ho, “*Nearly all large global companies disclose ESG information*,” Thomson Reuters, March 1, 2023. See also - International Federation of Accountants, “*Momentum builds for corporate ESG disclosure and assurance, yet reporting inconsistencies linger, study finds*,” February 27, 2023.
5. Deloitte, “*2023 Gen Z and millennial survey*,” Deloitte, May 2023.
6. International Federation of Accountants, “*Momentum builds for corporate ESG disclosure and assurance, yet reporting inconsistencies linger, study finds*.”
7. Ethan Karp, “*Behind all the ESG virtue signaling, there’s a real opportunity for manufacturers*,” *Forbes*, March 31, 2023.
8. Capital Group, *ESG global study 2022*, May 2022.
9. Deloitte, “*ESG mega trends*.” See also - Deloitte, “*Sustainability regulation: A catalyst for transformation*,” accessed November 9, 2023.
10. Deloitte, “*Private companies investing in ESG reporting*,” 2023. See also - Workiva, “*The annual reporting barometer 2023: Facing up to the CSRD*,” June 2023. See also - Oliver Pike, “*What companies within and outside of the EU can expect of new European ESG regulations*,” Thomson Reuters, May 4, 2023.
11. Deloitte, “*The challenge of double materiality: Sustainability reporting at a crossroad*,” 2023.
12. Federal Register, “*Federal acquisition regulation: Disclosure of greenhouse gas emissions and climate-related financial risk*,” November 14, 2022.
13. Deloitte, “*Creating a climate of change digest*,” March 2023. See also - “*SB-253 Climate Corporate Data Accountability Act*,” California State Senate, October 9, 2023.
14. US Securities and Exchange Commission, “*Enhanced disclosures by certain investment advisers and investment companies about environmental, social, and governance investment practices*,” August 16, 2022.

15. Brightest, “*What your business needs to know about the EU Corporate Sustainability Reporting Directive (CSRD)*,” October 30, 2023.
16. Matt DiGuiseppe, “*The No. 1 ESG challenge organizations face: data*,” World Economic Forum, October 28, 2021.
17. Tani Colbert-Sangree, “*What are emission factors? And where can I find them?*” GHG Management Institute, October 31, 2022.
18. Pike, “*What companies within and outside of the EU can expect of new European ESG regulations.*”
19. Jessica Pransky, Alessandra Leggieri, and Kim Knickle, “*Green quadrant: Enterprise carbon management software 2022*,” Verdantix, August 2, 2022. See also - Report Linker, “*Global ESG Reporting Software Market Size*,” GlobalNewswire, January 25, 2023.
20. Deloitte, “*Sustainability action report: Survey findings on ESG disclosure and preparedness*,” December 2022.
21. John Labate, “*Assessing sustainability credits under the Inflation Reduction Act*,” Wall Street Journal, October 15, 2022.
22. Markus Müller, Daniel Sacco, Afif Chowdhury, and Arun Sudi, “*ESG survey 2022: Trends and concerns*,” Deutsche Bank, November 8, 2022.
23. Tania Lynn Taylor and Sean Collins, *Ingraining sustainability in the next era of ESG investing*, Deloitte Insights, April 5, 2022.

Acknowledgements

The authors would like to thank **Genevieve Halloran, Ayesha Iyer, John Mennel, Jeff Loucks, Duncan Stewart, Susanne Hupfer, Pankaj Bansal,** and **Payal Bansal** for their insights.

Cover image by: **Manya Kuzemchenko**

On solid ground: AgTech is driving sustainable farming and is expected to harvest US\$18 billion in 2024 revenues

Agriculture technology is designed to let the producers and farmers grow more food using less pesticides, energy, water, and resources, enhancing farm yields.

ARTICLE • 8 MINUTE READ

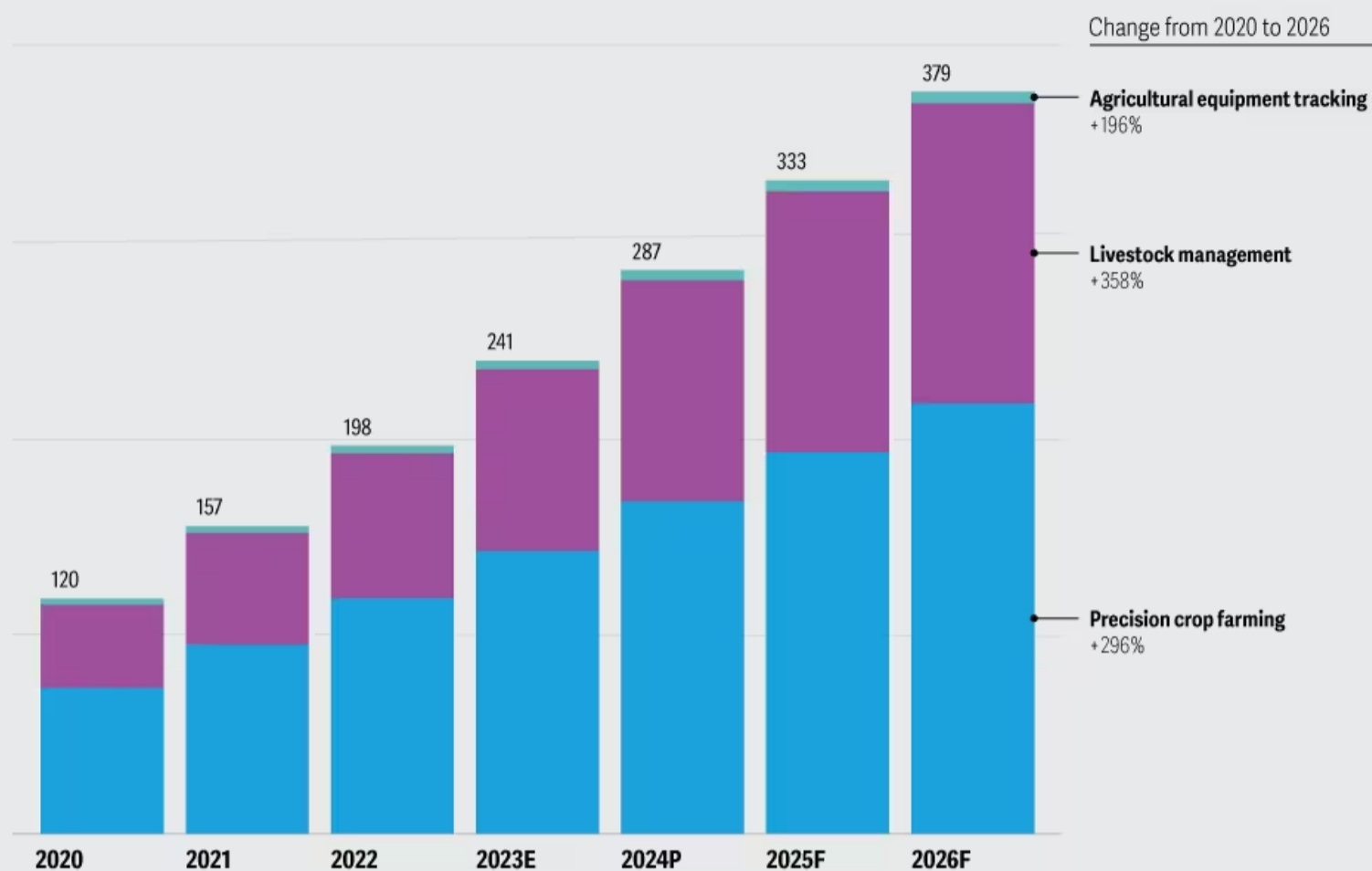
Deloitte predicts the installed base of Internet of Things end points for precision crop farming, livestock management, and agricultural equipment tracking will near 300 million by the end of 2024—a 50% growth over the 200 million installed base in 2022 (figure 1). Further, the overall agricultural technology (AgTech) revenue opportunity—including these IoT end points and connectivity devices—will be US\$18 billion globally in 2024, representing a 19% CAGR between 2020 and 2024.¹ Climate issues, geopolitical tensions, water and energy shortages, rising fertilizer costs, and inefficient production methods exert pressure on agricultural production. AgTech solutions can help improve crop yields, use equipment and livestock efficiently, plan harvests better, and adopt sustainable agrifood-production methods.

The use of these technologies has the potential to reduce emissions and save input costs for farmers. As a case in point, a 2022 Deloitte study (in collaboration with the Environmental Defense Fund) projected that precision agriculture tech solutions alone have the potential to abate 9.8 gigatons of carbon dioxide-equivalent (CO₂e) emissions between 2020 and 2050 and can save an estimated US\$40 billion to US\$100 billion in costs to farmers by 2030.²

Figure 1

The installed base of IoT endpoints for precision crop farming, livestock management, and agricultural equipment tracking is expected to near 300 million by the end of 2024

AgTech use case endpoint installed base, worldwide, 2020–2026 (unit shipments, in millions)



Notes: E denotes estimated values; P denotes predicted value; F denotes forecast numbers.

Sources: Graphic prepared by Deloitte based on data from Gartner®, *Forecast: Internet of Things, endpoints and communications, worldwide, 2021–2032, 2Q23 update, July 2023*; and Gartner®, *Forecast: Internet of Things, endpoints and communications, worldwide, 2020–2030, 4Q21 update, January 2022*.

Gartner is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the United States and internationally, and is used herein with permission. All rights reserved.

Deloitte Insights | deloitte.com/insights

Rising prices and supply-side issues escalate food-security concerns

Agriculture feeds billions of people, but it also feeds on natural resources and contributes to climate change. Today's global agriculture and food industry is responsible for more than 20% of global CO₂e emissions.³ Irrigation represents 40% of freshwater withdrawals. Meanwhile, the number of people facing food insecurity in 2023 is projected to be 345 million, up from 135 million pre pandemic.⁴ With global population expected to near 10 billion in 2050—2 billion higher than today⁵—the pressure on food demand is growing and inevitable. At the same time, the average age of farmers is rising while those taking up agriculture as a primary occupation is plummeting.⁶

There's an urgent need to address the looming food crisis,⁷ but scaling food production using current farming methods may be resource-intensive and inefficient. Commodity prices have gone up due to the increase in costs of labor, fertilizers, and equipment.⁸ The Russia-Ukraine war, trade sanctions, the pandemic's aftereffects, and other geopolitical issues have disrupted logistics and food supplies.⁹ Food security is a growing concern even in developed countries today.¹⁰

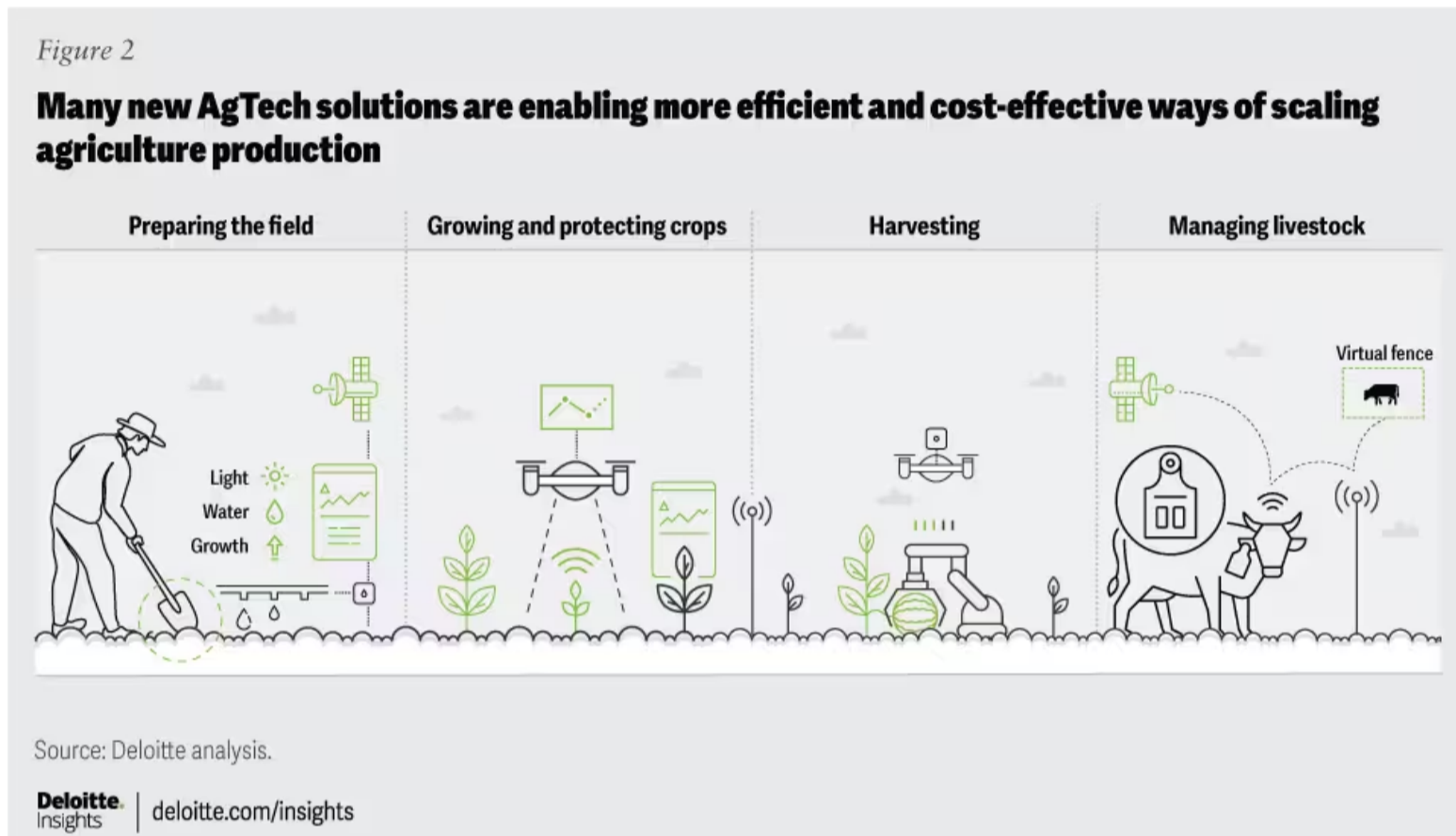
Nonetheless, AgTech could help solve many of these issues.

Technology unlocks value from farmlands and livestock to help augment food production

New AgTech solutions have the potential to scale agricultural production efficiently and in a cost-effective manner (figure 2). For example, hydroponics allows cultivating crops in nutrient-rich water instead of using soil and promise higher yields.¹¹

Moreover, at least 10 major technology and telecom companies are powering the agriculture industry through innovative solutions such as AI-based cultivation methods; farm and livestock data management platforms; and satellite, broadband, and IoT-enabled smart farming and vertical farming.¹² Besides, venture capitalists (VCs) continue to invest in AgTech startups despite the macroeconomic headwinds, and perhaps driven in part by the Russia-Ukraine war, given Ukraine's importance for global food supplies. They invested US\$10.6 billion in 2022,¹³ and in Q1 2023, VCs pumped in US\$1.9 billion into 172 AgTech startup deals.¹⁴

AgTech solutions are helping to revolutionize agriculture production, especially in the areas of preparing the field, growing and protecting crops, harvesting, and managing livestock (figure 2). Here's how:



Preparing the field

IoT devices and satellite connectivity pull critical farm data from multiple sources. For instance, farmers and agricultural advisers can gather data on soil type, moisture, and weather conditions and use handheld devices to record and upload their observations on digital farm management platforms. Further analytics can help estimate the quantity of water and fertilizer the plants require and determine crop protection needs, making precision agriculture a reality.¹⁵ Predictive planting solutions even analyze microclimate data such as soil moisture and expected rainfall to help discover plant-worthy and harvest-worthy areas in fields.¹⁶

Sensors attached to water sprinkler's arms or heads help regulate water flow and improve the precision level, sprinkling the exact quantity of water that is needed and analyzing the type of farmland and the crops planned for cultivation.¹⁷ Precision mobile irrigation systems can save 30% to 50% of water use compared to traditional irrigation methods.¹⁸ Moreover, drip and microsprinkler irrigation systems have proven to enhance water efficiency by up to 70% compared to less efficient irrigation methods.¹⁹

Growing and protecting crops

Farmers can also use AgTech to make real-time crop-placement decisions and monitor crop health. One such example is the use of agronomic intelligence in India.²⁰ A case in point is infrared mapping and surveying of farmlands: With a combination of spectral sensors and chips, cameras mounted on unmanned aerial vehicles (UAVs) or drones gather large volumes of data (e.g., soil moisture, plant health, etc.) that AI models analyze to share insights that assist farmers with targeted spraying operations.²¹ Early systems (developed in 2020-21) to fallow weeds showed the potential to save 97.5% chemical use. Aerial images captured by drones help locate weeds and upload images into a processing platform. Farmers can use those insights to then spray in places exactly where the weeds are.²² Recently, the use of AI has helped identify weeds with a 96% accuracy and spray the intended target with precision.²³

Autonomous weeders have the potential to eliminate 100,000 weeds every hour and cover more than 15 acres of onions in a day, versus a laborer who could weed one acre in the same duration.²⁴ Deloitte estimates agriculture-drone shipments to be in the range of 7–8 million globally in 2023.²⁵ At an average price of US\$500 to US\$700 per equipment, the drone market could be worth at least US\$4–5 billion.²⁶ Advanced and large-size drones cost upward of US\$20,000.²⁷ Assuming the drone market grows at 10% every year, we predict drone-driven revenue opportunity for semiconductor chips, sensors, and connectivity modules to be roughly US\$500 million (or 10% of the agriculture drone market) in 2024.

Harvesting

As growers in the United States and the United Kingdom deal with labor shortages for picking fruits and vegetables during the busy season,²⁸ agribots could address this issue. Soft fruits like fresh tomatoes and strawberries require a feather touch and have previously not been suitable for robotic picking. To address this unique need, AgTech startups are piloting agribots that biomimic human arms, use complex motion planning, and figure out the quality and degree of a fruit's ripeness.²⁹

A warehouse in Queensland (Australia) trialed using robots—equipped with computer vision, machine learning, and robotic grasping—to pack avocados, working alongside regular staff.³⁰ Similarly, harvesting robots use computer vision, AI-enabled ripeness detection, and robotic agility to pluck fruits softly from the vine. Sensors triangulate data to ensure the agribot distinguishes fruits from other objects (e.g., leaves, stems, etc.) during the harvesting process.³¹

Productivity and efficiency benefits coupled with mitigating labor shortage issues will drive demand for harvesting agribots. From an estimated value of US\$700 million in 2022, we anticipate the global harvesting robot revenue to reach US\$1 billion by 2025, growing 15% to 20% annually.³²

Managing livestock

Australia, the United States, and Latin America are experimenting with virtual fences, instead of traditional physical fences, to manage livestock.³³ With the help of GPS tracking, audio stimuli, and benign electric signals, cattle grazing can be contained within desired limits while farmers can manage grazing time and location, monitor cattle health, and improve their productivity too. For instance, farmers in Australia have used the motion sensor technology to analyze cattle movement and used that data to increase cattle productivity by about 20%.³⁴ Moreover, ranchers could change boundaries dynamically to, for example, keep the cattle away from less desired areas such as recently burned or affected grasslands.³⁵ Data from GPS and accelerometers could provide rich insights into animal health and social interactions among cattle.³⁶

At US\$50 per cattle collar device plus US\$12,500 per base station,³⁷ setting up a full-fledged, one-mile virtual fence for 100 cows could cost close to US\$20,000—compared with a traditional physical fence that ranges from US\$10,000 to US\$100,000 depending on type of fence and physical materials used.³⁸

The road to change

Despite the array of solutions and investment flows, the road to transforming agriculture through technology is not without bottlenecks.

Small-holder farmers produce much of the food output that the world consumes, but they face funding challenges to meet basic agribusiness needs.³⁹ Farmers may resist adopting technologies as they could view them as a risky and costly proposition, and in general, they may not be fully aware of IoT device connectivity options available for agricultural uses.

However, in 2023, agribusinesses are facing more pressure to decarbonize their operations as regulations tighten. And as farmers are dealing with margin pressures and high costs of resources, AgTech has become important to help accelerate the shift toward data-driven decision-making. Farmers, research labs, and agricultural advisers have started to work together to address the various roadblocks that they are dealing with, for instance, by building integrated data platforms to connect siloed data sets and solve interoperability issues.⁴⁰

The bottom line

AgTech solution providers may want to consider several actions when looking to move toward a more sustainable production and to enhance efficiencies:

Educate farmers on AgTech options: Farmers should be educated about the various types of network connectivity and IoT backhaul connection options. There's likely room for improvement if they work closely with the agriculture ecosystem players to help them discover connectivity needs based on specific use cases, such as using Wi-Fi or 2G/3G for crop-watering systems instead of the more advanced 4G/5G or satellite networks.⁴¹

Assist with tech implementations: Tech companies can support developing cost-benefit assessment tools to help farmers evaluate and identify trade-offs for non-AgTech versus AgTech-enabled farming methods. Besides, they can assist the agriculture ecosystem players to figure out what connectivity technology is needed for a specific issue, for example, using edge computing and 2G/3G cellular links to implement satellite-connected cattle collars in a livestock ranch to build virtual fencing.⁴² This might require assessing the nature, provenance, timing, and volume of data that would flow across the supply chain; and implementing a permissioned and trusted exchange of data from farm to plate.⁴³ Importantly, AgTech companies can be discerning in collecting the right amount of data and establishing data governance processes to address farmers' concerns about privacy and data usage.

Create an integrated view of data across the agriculture ecosystem: Blending granular data related to land, soil, climate, and water on a shared digital platform could help farmers and extended ecosystem participants glean insights about the most prominent levers to enhance productivity and quality. This would likely require integrating data from cloud, satellites, mobile devices, sensor networks, and agribots, and using AI to run analytics and deliver insights over a common data-sharing platform that farmers, scientists, researchers, and agriculture consultants and advisors can consume.⁴⁴

Enable sustainability and measure effectiveness: From a social accountability standpoint, farmers will likely be required to furnish impact data for nature, climate, and animal welfare. Novel options such as using low-methane producing supplements to contain livestock emissions when cattle belch,⁴⁵ and installing solar photovoltaic panels on farmlands to generate solar power (aka agrivoltaics)⁴⁶ are already being explored. Further, AgTech providers can develop technology to measure, report, and verify relevant metrics to help farmers demonstrate the efficacy of their sustainable farming practices. With emerging ESG regulations, technology that captures data to help comply with sustainability frameworks such as SBTi (science-based targets initiative) and TNFD (Taskforce on Nature-related Financial Disclosures) and tracks emissions information will become critical. For water usage, technology that monitors and optimizes water used for irrigation is expected to gain prominence, for instance, low-rank adaptation of large language models (LoRA) based analytics, coupled with satellite direct-to-device (D2D)⁴⁷ or mobile (4G/5G) or Wi-Fi-based sensor networks to track, schedule, and allocate precise amount of water for plants.⁴⁸

AgTech can play a much larger role to not only help address persistent challenges that have plagued the agricultural sector for decades, but even deliver tangible benefits to farmers and consumers alike—lowering costs and improving return on investment, driving sustainable growth by reducing the strain on resources, and making food more plentiful and affordable.

By **Karthik Ramachandran**
India

Ben van Delden
Australia

Ariane Bucaille
France

Gillian Crossan
United States

Duncan Stewart
Canada

Endnotes

1. Data and analysis based on information gathered from publicly available sources. Third-party sources pegged revenues from AgTech (including crop management solutions, GPS field mapping, supply chain management solutions, and software and analytics) at US\$9–10 billion in 2020. Based on conversations with AgTech experts and from our analysis of this market, we expect revenues to grow at 19% CAGR between 2020 and 25, based on which, we predict AgTech revenue to reach US\$18 billion in 2024 and cross US\$21 billion by 2025.
2. Environmental Defense Fund and Deloitte, *Pathways to net zero: Innovation imperative*, accessed November 21, 2023.
3. In 2019, 31% of global anthropogenic emissions measured in carbon-dioxide equivalent (CO₂e) were generated by agrifood systems. To read further, see: FAOSTAT Analytical Brief 31, *The share of agrifood systems in total greenhouse gas emissions: Global, regional and country trends, 1990–2019*, accessed November 21, 2023.
4. Based on data published in: World Food Programme, “A global food crisis,” accessed November 21, 2023; moreover, with rising food prices and supply-side issues, essential food items remained inaccessible to millions globally. In 2020, one in three people or approximately 2.4 billion—which was 320 million more than 2019—did not have access to nutritious food. To see further: United Nations Statistics Division, “*End hunger, achieve food security and improved nutrition and promote sustainable agriculture*,” accessed November 21, 2023.
5. United Nations, “*World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100*,” accessed November 21, 2023.
6. Based on *the World Bank data on employment in agriculture*, agricultural employment fell from 43% in 1991 to 26% in 2021. Also, the ratio of agricultural employment fell from 180,000 per million population in 2000 to 120,000 per million in 2019 globally; based on data from: Max Roser, “*Employment in agriculture*,” Our World In Data, accessed November 21, 2023.
7. WHO data on *malnutrition* (June 9, 2021) validates it’s a global challenge, and more so, for the developing world.
8. Talking about the labor issue in particular, immigration restrictions—several of which came into effect during the COVID-19 pandemic—have affected labor mobility in regions like Australia where there is a greater need for skilled labor in agriculture and rely on workforce migrating from other countries.
9. For example, 18 countries (as of June 30, 2022) had trade-related policy measures in place—affecting 5.5% calories traded in global market. To read further, see IMF, *Tackling the global food crisis: Impact, policy response, and the role of the IMF*, accessed November 21, 2023.
10. As a case in point: In the United States, 12.8% households were food-insecure at some point of time in 2022 (up from 10.2% in 2021). Of which, 5.1% (or 6.8 million households) experienced very low level of food security (up from 3.8% in 2021). To read further, see: Matthew Rabbitt, Laura Hales, Michael Burke, and Alisha Coleman-Jensen, *Household food security in the United States in 2022*, US Department of Agriculture, accessed November 21, 2023.

11. I. Habeeba, G. Vinothini, G. Rajasekar, “*Hydroponics—The future farming*,” *Just Agriculture* 3, no. 7 (2023).
12. Based on Deloitte’s research and analysis of select major global technology and telecom companies’ efforts and initiatives related to developing and launching AgTech solutions and offerings.
13. PitchBook, *2022 Agtech overview*, March 16, 2023.
14. PitchBook, *Q1 2023 Agtech report*, May 4, 2023.
15. As an example, agricultural researchers in the United States have already used a combination of drones, specialized cameras, and sensors to get a deeper understanding of crops, soil, and weather—to help farmers precisely know when and how much to irrigate. To read further, see: Dan Elliott, “*Farmers are using drones to help save an endangered US river*,” *USA Today*, August 14, 2019.
16. Based on insights from an expert interview. Also see: Leilei He et al., “*Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods*,” *Computers and Electronics in Agriculture* 195 (2022).
17. Drashti Bhavsar et al., “*A comprehensive and systematic study in smart drip and sprinkler irrigation systems*,” *Smart Agricultural Technology* 5 (2023).
18. GeoPard Agriculture, “*Role of precision irrigation methods in modern farming*,” accessed November 21, 2023.
19. AGRIVI, “*Water sustainability: How to grow your farm and take care of the environment*,” Agronomy Blog, accessed November 21, 2023.
20. In India, agronomic intelligence tech empowers farmers to make better decisions about crop selection and disease management depending on underlying ground conditions. There’s great scope to increase agricultural productivity by using technology in crop selection, package of practices and pest and disease management. Startups use agronomic intelligence to provide farmers with information that allows them to better tailor what they do to crops to conditions on the ground. To read further, see: B. K. Jha, “*Revolutionizing farming through agronomic intelligence*,” altLIPI, accessed November 21, 2023; Suparna Dutt D’Cunha, “*For India’s farmers it’s Agtech startups, not government, that is key*,” *Forbes*, January 8, 2018.
21. Cameras mounted on a tractor can analyze real-time field conditions using sensors, based on which, the devices can apply or spray the specific product that the crops need and in optimal quantity—be it nitrogen, plant-growth regulators, defoliant, fungicides, or herbicides. To read further, see: Dan Miller, “*Raven, Augmenta launch Sense and Act tech*,” *DTN Progressive Farmer*, April 30, 2023; Heather Hall, “*The smart future of agriculture*,” *Design World*, July 11, 2023.
22. Solar Impulse Foundation, “*Fallow weed detection*,” January 15, 2021.
23. Coco Liu, “*AI weed-jilling drones are coming for the mega farms*,” *Bloomberg*, April 19, 2023.

24. Kristin Houser, “[Farming robot kills 200,000 weeds per hour with lasers](#),” Freethink, October 19, 2022.
25. Deloitte analysis based on data gathered from publicly available sources including [Future farming](#) (2023) and [Counterpoint Research](#) (2023).
26. JOUAV, “[How much does a drone cost in 2023? Here’s a price breakdown](#),” September 15, 2023.
27. UAV Training Australia (UAV & Drone blog), “[Drones in Agriculture: Weighing up the costs involved](#),” March 15, 2022.
28. Stephen Gossett, “[How Root AI’s agricultural robots are powering the farmtech revolution](#),” Built In, May 19, 2020.
29. Ibid.
30. Australian Tree Crop, “[Trials underway for new packing and sorting technology at avocado packhouse](#),” October 22, 2020.
31. Gossett, “[How Root AI’s agricultural robots are powering the farmtech revolution](#).”
32. Astute Analytica report on [Harvesting robot market](#), February 21, 2023 (summary news release, accessed November 21, 2023).
33. WIPO, “[Chipsafer—The cattle tracking system that locates livestock](#),” accessed November 21, 2023.
34. Telstra, “[IoT in focus: Transforming the agriculture industry](#),” October 4, 2018.
35. USDA Northwest Climate Hub, “[Virtual fencing: A climate adaptation strategy](#),” accessed November 21, 2023.
36. Future Learn, “[Automated farming: Virtual fencing](#),” accessed November 21, 2023.
37. Association for Computing Machinery, “[Virtual fences for cattle find a home on the range](#),” *Wall Street Journal*, May 22, 2023.
38. Audrey Bruno, “[How much does farm fencing cost? \[2023 data\]](#),” Angi, May 27, 2021.
39. Based on interviews with Deloitte’s agribusiness experts in Australia and Europe, we note that most farmers are not only small-scale holders, but their agribusiness is family-owned, for instance, as seen in Australia and South Asia. And that presents challenges in raising funds and accessing credit.
40. For example, Data Farming launched cloud-based platform ([Digital Agronomist](#)), a crop-monitoring tool that provides integrated access to 10 million satellite maps, allowing users to map millions of hectares from across the world—designed for farmers and agronomists.
41. National Broadband Network and National Farmers Federation (Australia), [Connecting Australian Agriculture](#), November 2021.

42. A controlled experiment demonstrated how cows in virtual-fencing treatment pastures grazed only 4% of their time in burned areas vis-à-vis cows that didn't receive triggers from their virtual collars, and hence ended-up spending 40% time in burned areas. To read further, see: USDA Northwest Climate Hub, "[Virtual fencing excludes cattle from burned areas in sagebrush steppe](#)," accessed November 21, 2023.
43. Integrity Systems, "[Australian AgriFood Data Exchange](#)," accessed November 21, 2023.
44. For example, the [gaiasense](#) system combines diverse technologies including big data, cloud, IoT, AI/ML, and semantic web with interdisciplinary fields such as soil science, agricultural engineering, meteorology, agronomic and biological sciences, and environmental sciences. See: [gaiasense smart farming system](#).
45. Rod Nickel, "[Canadian dairy cows among first in world bred to belch less methane](#)," CBC, August 8, 2023.
46. *PV Magazine*, "[TSE and the agrivoltaic development in France](#)," June 8, 2023.
47. To read more on benefits and value of D2D technology, see our related *2024 TMT Predictions* chapter on "[direct-to-handset satellite](#)."
48. With 85% freshwater resources used for irrigation worldwide, current irrigation methods need a major upgrade and need to consider using AgTech-based methods. For instance, compared with traditional watering methods, LoRA (low-rank adaptation of large language models) based ML solution consumes less water and enables a 46% reduction in water usage, even as plants looked better than they would have with conventional watering. To read further, see: G. S. Prasanna Lakshmi et al., "[An intelligent IOT sensor coupled precision irrigation model for agriculture](#)," *Measurement: Sensors* 25 (2023).

Acknowledgements

The authors would like to thank **Dr. Daniel Terrill, Panos Kalogiorgas, Pete Edmunds, Negina Rood, Gautham Dutt, and Ankit Dhameja** for their contributions to this article.

Cover image by: **Manya Kuzemchenko**

Women's elite sports: Breaking the billion-dollar barrier

Interest from fans, broadcasters, and commercial partners is driving rapid growth in the revenues of women's elite sport. How can organizations continue to accelerate growth?

ARTICLE • 15 MINUTE READ

Deloitte predicts that in 2024, revenue generated by women's elite sports will surpass US\$1 billion for the first time, with US\$1.28 billion in total forecast revenues (figure 1). This total is projected to be at least 300% higher than when we last wrote our prediction on this topic in 2021. Our forecast is composed of commercial (US\$696 million, 55%), broadcast (US\$340 million, 27%), and matchday (US\$240 million, 18%) revenues.

A sharp increase in commercial revenue is the major growth driver, followed by income from broadcast and matchday sources. The largest geographical markets are forecast to be North America (US\$670 million, 52%) and Europe (US\$181 million, 14%) while the two most valuable sports are projected to be football/soccer (US\$555 million, 43%) and basketball (US\$354 million, 28%). Global competitions such as the FIFA Women's World Cup, Ladies Professional Golf Association (LPGA) tour, and Women's Tennis Association (WTA) tour, are expected to contribute US\$425 million (33%) of the forecast total.

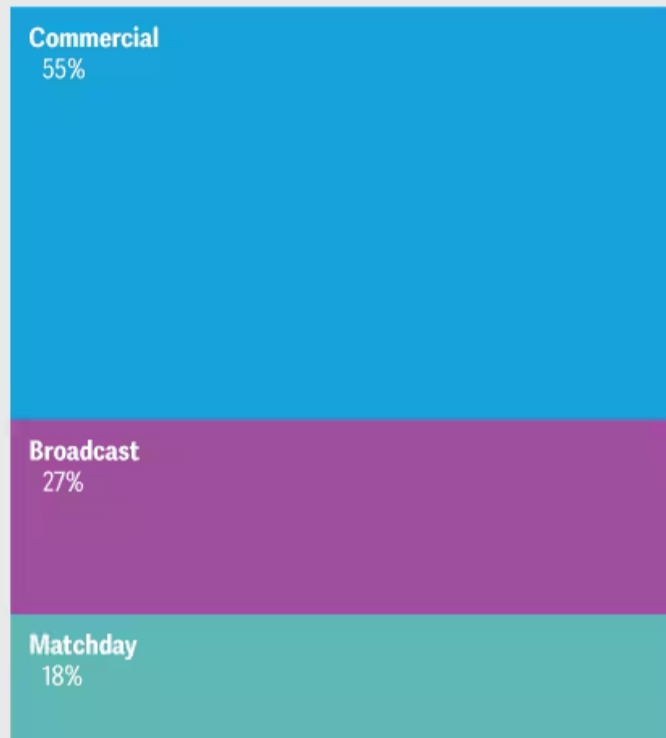
As an indicator of the potential in, and momentum behind, women's elite sports, Deloitte expects valuations for teams and leagues to continue to rise, with several teams' values predicted to exceed US\$100 million in 2024. The allure for investors could be the opportunity to buy into a growing sector for a modest outlay relative to other elite sports properties. Spending by new and established investors is likely to fund increased professionalization and commercialization across women's elite sports, which in turn could lead to more spectacular playing performances, increased spectator interest, and thus potentially higher returns on investment. The experience of the 2023 FIFA Women's World Cup was a testament to the continuing improvement in the technical capability of the players, coaches, and referees. As one example of the increasing prowess of female athletes, in the 2023 FIFA Women's World Cup, England's Chloe Kelly's penalty kick peaked at 111 km/h, a speed higher than every kick speed recorded in the 2022-2023 men's England Premier League season.¹

Surpassing the billion-dollar revenue milestone is expected to be significant, especially given the turbulence of recent years for elite sports. In 2024, women's elite sports should continue to bask in the multi-faceted success of the 2023 FIFA Women's World Cup which generated more than US\$570 million in revenue.² Women's elite sport is developing swiftly in stature, but is also still nascent, with many of the major leagues only established in the last decade, and many territories yet to establish a mainstream culture around women's elite sports.

Figure 1

Total revenues for women's elite sports are projected to pass the billion-dollar mark in 2024

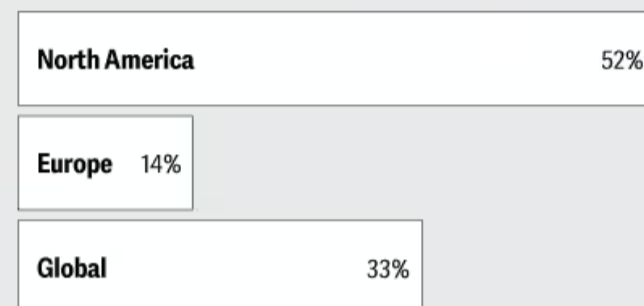
Value of global women's sport 2024 split by revenue type



Global women's sport revenue split by sport



Global women's sport revenue split by region



Source: Deloitte analysis, based on public data.

Deloitte Insights | deloitte.com/insights

Revenues: breaking down the billion dollars

The three main revenue categories for women's and men's elite sports are the same (matchday, broadcast, and commercial), but at this stage of the evolution of women's sports, it's commercial revenues that represent the largest share, according to Deloitte's analysis. In the men's game, revenue generated from the sale of broadcast rights is the primary contributor.³

Commercial: the major driver of revenue growth

Deloitte expects commercial income to have the biggest impact on revenues for women's elite sport in the near term. Commercial includes club sponsorship and partnership arrangements; central sponsorship revenue, retail, and merchandising sales; licensing; and all other non-matchday or broadcast revenues.

Sponsorships

The value of sponsorship in women's elite sport is likely to grow rapidly in 2024, encouraged by impressive recent evidence of returns on investment.

According to one study, for every dollar spent by a corporate sponsor in women's sports, more than seven dollars is generated in "customer value for that organization."⁴ Some sponsors of the Ladies Professional Golf Association (LPGA) reported up to 400% return in media value on their investment.⁵ Between 2019 and 2023, the LPGA *doubled* the number of active brands on its sponsorship roster, and the number of sponsors passed 1,000 for the first time in June 2023.

Sponsors of women's sports organizations and athletes are often attracted by the chance to associate their brands with themes of female strength, empowerment, and gender parity, among other factors. The UK's Women's Sport Trust found that 29% of consumers think more favorably of companies or brands that support women's sport through their sponsorship, 12% higher than for campaigns supporting men's sport.⁶

Sponsorship of leagues: renewals at higher values

In 2024, we expect league sponsorships to be renewed at higher values. This would follow the example of Barclays and the Football Association (FA) Women's Super League (WSL), the premier English women's football (soccer) league, which doubled to £30 million (US\$ 37 million) upon renewal for 2022–2025.⁷ The new sponsorship additionally included the Women's Championship (second-tier league) and support for the FA Girls' Football School Partnership.⁸

Club and team sponsorship: the era of unbundling

In 2024, we expect to see growth in standalone sponsorship deals for women's teams, especially in women's football (soccer). Historically, and more commonly in Europe, these were typically bundled with their affiliated men's team. This can make it difficult to quantify the specific financial value of the women's team.

The unbundling of rights will likely galvanize investment and brand activation into women's teams. As of 2023, just seven of the top-tier [Deloitte \(UK\) Football Money League](#) clubs negotiated the shirt sponsor of the women's side separately.⁹ The main sponsor for 38% of all global professional women's football clubs is for the women's team only, and a quarter of all kit sponsors are only for the women's team.¹⁰ We expect the proportion of women's only agreements to increase in 2024 and in the medium term. The optimal timing for unbundling varies by club, with key considerations such as club size, fan preferences, partner attributes and scale of commercial arrangement.

Opportunities for women's club partnerships are continuing to develop. In September 2023, Manchester City Women named baby gear brand, Joie, as the first stadium naming rights sponsor of a WSL club.¹¹ In October 2023, Arsenal FC Women debuted their away kit, in collaboration with Adidas and Stella McCartney, distinct from their men's counterparts.¹² As the sport professionalizes, this may create new and specific sponsorship opportunities unique to the women's game.

Competition sponsorship: over US\$300 million for the 2023 FIFA Women's World Cup and possibly more in 2027

We expect sponsorship revenues for the 2027 Women's World Cup cycle to outperform the 2023 event, as FIFA and global brands are likely to have more developed sponsorship strategies dedicated to women's football and new commercial partners looking to leverage the opportunity.

The 2023 FIFA Women's World Cup saw FIFA unbundle the commercial rights to the competition for the first time, netting US\$308m in sponsorship revenue and fulfilling its entire sponsorship inventory for the tournament.¹³ Under a new commercial partnership structure allowing brands to sign sponsorship deals specifically for women's football, FIFA secured 30 partners for the tournament, a 150% increase on the 2019 edition.¹⁴

Athlete sponsorship: Women athletes can command millions of dollars per year

Individual sponsorship in women's elite sport is likely to remain a key income source for athletes, with many likely to generate millions of dollars in sponsorship revenue in the lead-up to the Paris 2024 Olympic Games.

Alex Morgan, a US Women's National Team and National Women's Soccer League (NWSL) star, earned US\$6.3 million in off-field earnings in 2022.¹⁵ Her playing salary was US\$800,000.

Leah Williamson, captain of the England Lionesses during their European Championships win, became the first female football player to partner with Italian fashion house, Gucci.¹⁶

As the visibility of women's sports rises, sponsorships should become more valuable, driving further professionalization.

Shift in sports sponsorship spend to women's sports

We expect some companies to rebalance their sponsorship spending more towards women's sports. For example, in 2022, Ally Financial, a digital financial services company headquartered in the United States pledged to spend equally in paid media across men's and women's sports by 2027.

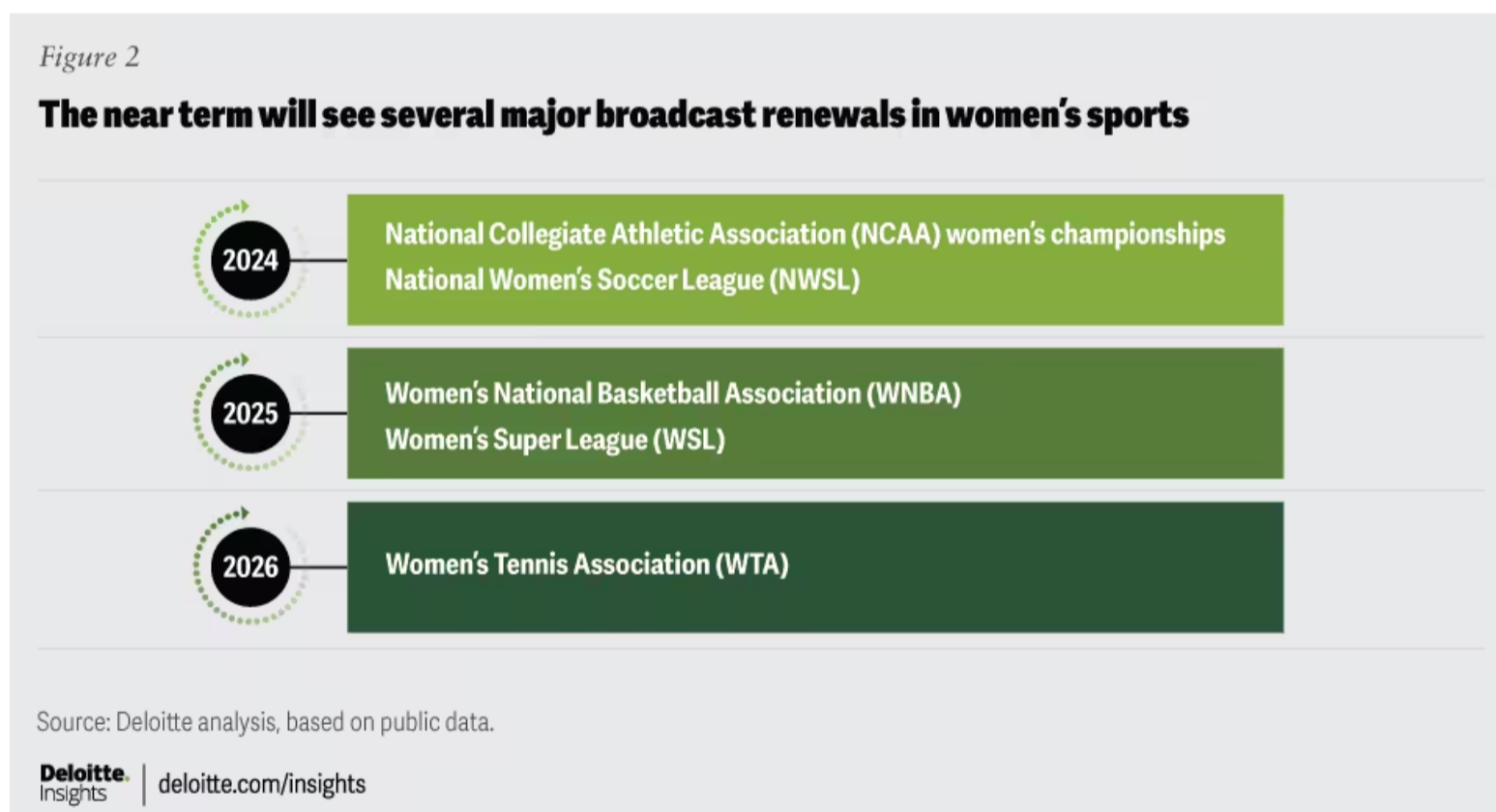
As the NWSL's title sponsor, Ally has collaborated with media companies to increase the league's visibility. Ally's increased advertising spend with TV network CBS was contingent on the 2022 NWSL championship game being shown in a prime-time broadcast slot for the first time.¹⁷ The Portland Thorns' win over the Kansas City Current generated a record audience of 915,000, a 71% year-over-year increase on the 2021 final.¹⁸

Broadcast: visibility, viability, and consistency

The number two revenue stream in women's elite sports in 2024 is forecast to be broadcast revenue.

Women's sport is currently balancing the benefits of free-to-air (FTA) transmission, which offers greater visibility, versus the higher fees, but lower addressable audience via pay TV.

In the 2023/2024 season, UEFA Women's Champions League matches will be mostly paywalled following two seasons of free-to-view games on DAZN's YouTube channel.¹⁹ Several major broadcast deals in major women's sports leagues will be renewed in the near term (figure 2), expected to deliver significant uplifts on the value of previous deals. The National Collegiate Athletic Association (NCAA) March Madness women's basketball tournament in the United States, will reportedly be worth a standalone value of over US\$100 million versus the current US\$34 million contract (expiring August 2024) which includes all 29 women's sports championships.²⁰



The NWSL announced in November 2023 a new four-year 'record-breaking' broadcast deal commencing in 2024 with coverage across four different platforms: CBS, ESPN, Scripps Sports, and Amazon.²¹ The deal is reported to be worth US\$240 million over the period, or \$US60 million per year, significantly higher than the prior deal.²²

The profile of the NCAA tournament can have a significant impact on both the perception of women's basketball and the valuation of the tournament's stars in the United States. Following the 2023 tournament, Caitlin Clark (Iowa) and Angel Reece (Louisiana State University) both agreed to major sponsorship deals.²³ The increased profile should bolster sponsorship values for female athletes and boost matchday revenues: Iowa women's basketball sold out the entire 2023/24 season with more than 15,000 tickets sold compared to only 6,500 tickets sold in the 2022/23 season.²⁴

Making viewing habitual: increasing consumption between the peaks

A focus in 2024 and the following years will likely be to make viewing habitual, rather than occasional. In recent years, women's elite sports have enjoyed ratings that rank among the most watched programs in some markets, but these are mostly for finals and semi-finals.

Deloitte expects that in 2024, women's elite sports will continue to be given more prime time slots, becoming easier to find for both casual and avid fans. Roku, the streaming company, introduced the Women's Sports Zone in 2023, a new centralized in-app location for US users to access live women's sports content.²⁵

Matchday: women's sports moving to the main stadiums

Matchday revenues are forecast to be the third-largest source of revenue in 2024.

The increasing use of major stadiums to host matches will allow more fans to experience women's elite sport firsthand, leading to increased attendance and in turn, greater matchday revenue. Higher ticket prices, from a low base, should also have a marked impact on income. In 2024, it's likely that very high attendance (more than 60,000) will be achieved for certain marquee matches, with records likely to be broken, as they were in both 2022 and 2023 (figure 3).

Figure 3

Attendance for women's sporting events continue to grow when given increased capacity and enhanced promotion

Date	Attendance	Event	Location	Sport
30th August 2023	92,003	Nebraska Cornhuskers v. Omaha Mavericks	Memorial Stadium, Lincoln, USA	Collegiate volleyball
22nd April 2022	91,648	Barcelona v. Wolfsburg (UWCL Semi-final)	Camp Nou, Barcelona, SPN	Soccer/football
30th March 2022	91,553	Barcelona v. Real Madrid (UWCL Quarter-final)	Camp Nou, Barcelona, SPN	Soccer/football
8th March 2020	86,174	Australia v. India (T20 World Cup Final)	Melbourne Cricket Ground, Melbourne, AUS	Cricket
10th July 1999	90,185	USA v. China (FWWC Final)	Rose Bowl, Pasadena, USA	Soccer/football

Source: Deloitte analysis, based on public data.

Deloitte Insights | deloitte.com/insights

In the NCAA, Nebraska's women's volleyball hosted a game at the school's American football stadium with 92,003 fans attending.²⁶ The August 2023 event shattered the previous NCAA attendance record for women's volleyball, also held by Nebraska, of about 18,000 attendees.²⁷

An American Football stadium was also the venue for the Iowa women's basketball team's record-breaking attendance for an exhibition game, in October 2023, with more than 55,000 tickets sold as of mid-October 2023, breaking the previous record of just under 30,000.²⁸

The confidence to host women's matches in major stadiums can be boosted by the regularity of large audiences across a range of sports. For the 2023 women's Ashes, a contest between England and Australia's women's cricket teams, ticket sales increased 450% from the 2019 series, with 110,000 fans in attendance. The newly established T20 Cricket Women's Premier League (WPL) in India, had an average per match attendance of up to 13,000 across its inaugural 20-match cricket season.²⁹

For women's football teams that are affiliated with men's clubs, we expect to see a greater share of the season's matches played at the main stadium. For example, in the 2023/24 season, Arsenal Women are scheduled to play five (up from three) WSL games at the Emirates Stadium, which has capacity for 60,000 fans,³⁰ thirteen times greater than the women's team's home stadium capacity of 4,500.³¹

In the prior season, Arsenal Women hosted eight games in total at Emirates: five Union of European Football Associations (UEFA) Women's Champions League matches as well as three WSL fixtures. The team's UEFA Champions League semi-final, second-leg matchup against Wolfsburg sold out with a crowd of 60,000.³²

In Mexico, clubs in Liga MX Femenil already play all matches in the same stadium as the affiliated men's team.³³ Larger stadiums help facilitate higher attendance, while also providing fans with familiar grounds and ease of access to attend games. It also means that women's teams across the league can benefit from the same field conditions as their male counterparts.

Differentiating the experience

Women's sport organizations are becoming more creative with their ticketing schemes and fan activations. For example, the AIG Women's Open, one of five major championships in women's professional golf, featured a fan village offering Q&As with women in sport, business and entertainment, big screens to watch the action, and live music.³⁴ Neither live music nor a festival atmosphere are traditional to the men's game.

Chelsea Women FC's 2023/24 matches at Stamford Bridge, the club's main stadium, will have options for live music and entertainment areas, family centers and supporter group sections.³⁵

Women's elite sports' modest ticket prices make it an affordable experience for families and widens the age mix at games. For the 2023-2024 season, ticket prices for leading teams such as Chelsea Women FC start at £15 (US\$19) per game;³⁶ for FC Barcelona Femení from about €14 (US\$15),³⁷ depending on the fixture; for the Chicago Sky, from \$15 per seat.³⁸ As organizations continue to grow and professionalize, the price of tickets will be a lever for increased matchday revenues, leading many to increase prices over the next few seasons.

In the upcoming seasons, we expect to see the demand for women's elite sports events grow, which will require careful planning and strategy from organizations. While attendance records for specific events could be broken, the next step on the path toward long-term success will be consistency in attendance for regular matches and creation of a distinctive experience at events.

As momentum builds, investment and valuations continue to grow

Deloitte expects women's elite sports team valuations to continue rising in 2024, with further clubs exceeding the US\$100 million threshold. Investment models are creating new channels for investment, including private equity funds which have supported the recent growth in values of men's elite sports teams. However, the club structures in Europe, with women's teams often attached to men's, may make independent valuations more complex in the short term.

Deloitte expects the most significant deals to be in the United States and Europe, as was the case in 2023.

The ownership group of the NBA's Golden State Warriors have reportedly agreed to pay a league-record \$50 million expansion fee to bring a WNBA team to California's Bay Area, the league announced in October 2023.³⁹ This is the first expansion of the league in more than a decade.⁴⁰

In April 2023, Sixth Street Partners committed US\$125 million to the new women's soccer club Bay FC.⁴¹ This included a US\$53 million expansion fee, paid to enlarge the size of the league. The expansion fee is ten times that paid in 2020 when Los Angeles and San Diego joined the league. Angel City FC, the Los Angeles franchise, has grown to be valued at more than US\$100 million in April 2021.⁴² With additional expansion teams on the horizon for the NWSL, the next cohort of clubs is likely to attract record-breaking valuations.

Investor appetite has also spread to Europe, where investment in women's elite football clubs is more complex due to the affiliation with men's teams. In May 2023, Michele Kang, who acquired the NWSL's Washington Spirit in 2022,⁴³ announced her intent to acquire a 52% stake in eight-time UEFA Women's Champions League winners Olympique Lyonnais Feminin.⁴⁴ This transaction created the first global, multi-club ownership model dedicated to women's soccer and fueled a discussion about valuing a women's team separately from their male counterparts, that will likely continue into 2024.

In 2023, the Women's Tennis Association (WTA) announced a strategic partnership with CVC Capital Partners (CVC) to accelerate the growth of women's tennis.⁴⁵ This agreement will include CVC acting as a catalyst to drive growth in the sport by providing increased access to the sport, the Tour's brands, and building athlete profiles.

There may be no major women's cricket deals in 2024, but in 2023, season one of the Women's Premier League (WPL) saw five teams sold for a cumulative US\$572 million.⁴⁶ Fifteen years ago, at the inaugural season of the men's Indian Premier League (IPL), eight franchises were bought for approximately US\$724 million.⁴⁷ The average team value for the WPL auction was higher than for the men's, at US\$114 million for WPL versus US\$90.5 million for IPL.⁴⁸

Celebrity ownership

A common theme across global sport, including the NWSL in the United States, is celebrity ownership groups. Celebrity ownership raises the profile of leagues and clubs, and in turn, increases the commercial appeal of these properties. Angel City FC's owners include entrepreneur Alexis Ohanian, actresses Jennifer Garner and Natalie Portman, and athletes Mia Hamm and Serena Williams. Kansas City Current's ownership group includes NFL Superbowl-winning quarterback Patrick Mahomes and his wife Brittany Mahomes. Former NBA All-Star Dwyane Wade joined the ownership group of the WNBA's Chicago Sky in July 2023.⁴⁹ Kevin Durant, a current NBA player, contributed to a US\$30 million capital raise for Athletes Unlimited, a women's sports league run by and for the athletes.⁵⁰ We expect this trend to continue in the near term, with additional celebrities in entertainment, sport, and fashion industries entering women's sports ownership conglomerates.

Women's sports focused investment

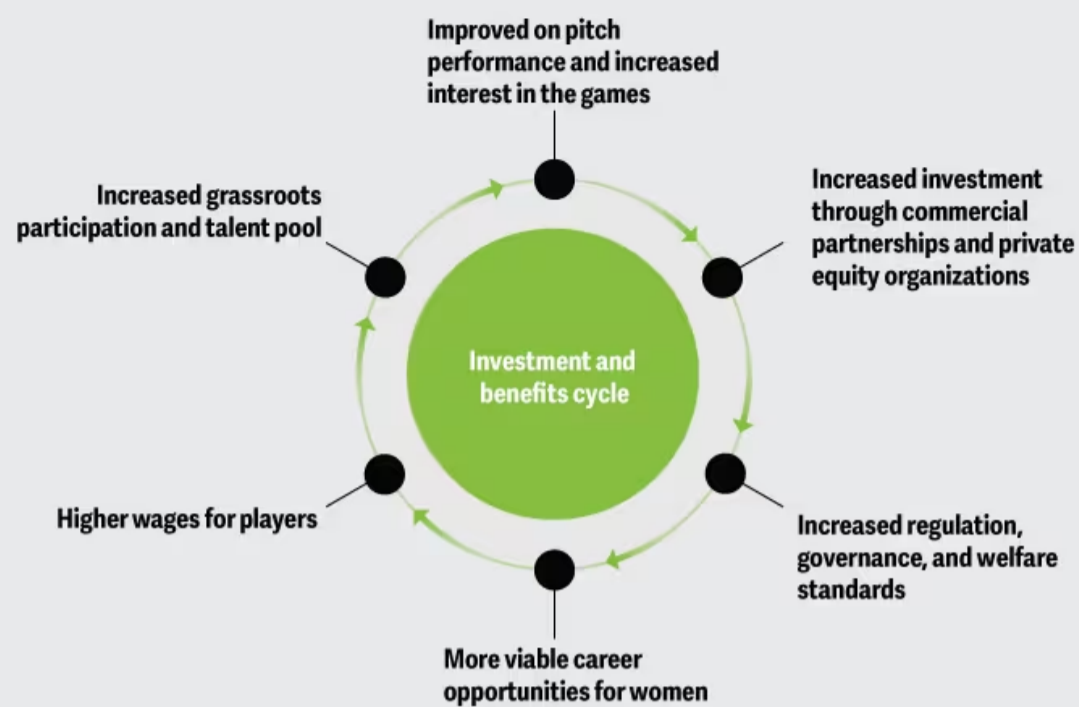
The introduction of dedicated investment funds for women's sports sets up further opportunities to grow women's elite sport. The Monarch Collective and Mercury 13 are both female-led, US\$100 million funds that will invest solely in women's sports properties. The Monarch Collective made its first investment in the NWSL's Boston expansion franchise in September 2023.⁵¹ Mercury 13 plans to make its first investment in Lewes FC, an English FA Women's Championship club known for its gender equality practices.⁵²

The virtuous circle of investment

More investment is likely to lead to better teams and a better product, which in turn often becomes more valuable (figure 4).

Figure 4

Investment in performance can mean stronger athletes and stronger earnings



Source: Deloitte analysis.

Deloitte
Insights | deloitte.com/insights

One benefit of investment is better facilities. The Las Vegas Aces in the WNBA recently unveiled a new practice facility and headquarters which, alongside multiple courts, features dedicated locker rooms, treatment areas, and cold tubs.⁵³ The increasing demands on athletes of women's elite sports will require investment in performance-focused facilities and staffing. The Kansas City Current in the NWSL is set to open the first stadium built specifically for a women's sports team, following the opening of their own US\$18 million training facility.⁵⁴

The bottom line

The impending billion-dollar milestone should be celebrated, but women's sports—at elite and grassroots levels—should be nurtured and invested in, both in 2024 and in the years to come. In the near term, our recommendations for women's elite sport to achieve its potential include the following considerations:

Women's elite sport should be developed as a product distinct from men's elite sport. Women's sports should not simply duplicate men's sports, but rather adapt and innovate to suit athletes and fans. In this nascent phase, women's sports can test new technologies, activations, and partnerships. It can forge its own path.

Develop the pipeline of talent. As women's elite sports become more visible, young girls will likely be inspired to participate. From school to the major stage, the industry should lower barriers to entry for young women. Increased visibility and investment in grassroots programs can provide girls with the opportunity to engage with sport, build confidence from a young age, and increase social connection by playing sport with others. Expanding the image of what female athletes look like, using inclusive imagery and storytelling to create a wide range of role models, could inspire higher participation rates by girls and women. One country that has committed to its pipeline is Australia, which used its cohosting of the 2023 FIFA Women's World Cup to announce a A\$200 million (US\$ 128 million) program to "promote equal access, build more suitable facilities, and support grassroots initiatives to get women and girls to engage, stay, and participate in sport throughout their lives."⁵⁵ Girls who grow up playing sports can become more engaged sports fans as an adult, creating an additional incentive for leagues, governing bodies, and countries to invest in the pipeline.

Protect athletes, reduce injuries. The rapid professionalization of women's sports has led to increased intensity, frequency of training, and reduced recovery time between each match. Since 2021, at least 87 players in eight of the world's top women's football (soccer) leagues have torn their anterior cruciate ligaments (ACL).⁵⁶ More investment in research and women-specific medical staff should be considered to understand female athletes' specific needs with regards to injury prevention. There's still much progress to be made: According to the third edition of FIFA's Women's Benchmarking Report, under half of the more than 200 football clubs surveyed employed a sports scientist, psychologist, nutritionist, and massage therapist.⁵⁷

Raise all profiles, nurture legends. To attract new fans and increase participation, organizations should focus on telling female athletes' stories outside of the pitch, court, and field of play. Digital platforms, particularly streaming sites and social media platforms, can play a role in engaging fans outside of major events. Increasing the awareness and viewing of women's sports depends on growing the media coverage of women's leagues, teams, and athletes. Online distribution is likely to be critical.

Grow the infrastructure for women's teams. Women's teams are often confined to minor stadiums that can be difficult to get to through public transit. They often share practice facilities and have antiquated medical rooms. Investment into modern facilities will likely bolster performance and product. As the infrastructure in women's elite sports improves, it is expected that so too will the on-field or on-court product, further incentivizing fan engagement.

Share the success. The top clubs in each league and the top leagues in each country are thriving. But this may put them out of reach of other leagues and teams. The top four revenue-generating clubs in the WSL (Arsenal, Chelsea, Manchester City, and Manchester United) accounted for 70% of the leagues' revenue, among a league with 12 clubs.⁵⁸ This polarization within the leagues often extends to the availability of support and medical staff, investment in facilities, and other important aspects that allow athletes to perform at the highest level.

At the 2023 FIFA Women's World Cup, there were signs of a diversification in success:

- Spain won for the first time, beating England's Lionesses who themselves had reached the finals for the first time.⁵⁹
- Colombia became the first South American country to qualify for the quarterfinals since 2011.⁶⁰
- Morocco's FIFA Women's World Cup debut was a first for any Arab country.⁶¹
- Jamaica's team, which crowdsourced to help fund its trip to the 2023 FIFA Women's World Cup⁶² reached the knockout stages for the first time and was the first Caribbean country to make it to the Round of 16. The team held traditional powerhouses France and Brazil to goalless draws.

Women leaders in sport. Women hold the minority of management, coaching, and ownership positions within the sports industry.⁶³ The proportion of women in executive, decision-making positions within sport lags progress made in other industries. In coaching and scouting, more opportunities for women to learn and participate should occur to ensure more diversity at the top levels of the sport. In July 2023, Washington Spirit announced its inaugural Coaching Mentorship program to help elevate and support female coaches in its region, the first program of its kind to create access for the next generation of female coaches.⁶⁴

Consistency is key. Record-breaking spikes in viewership and attendance, although impressive, are likely needed to prove reliable for major changes to be made in broadcast valuations and stadium considerations.

Forecast methodology

This publication contains a variety of information derived from publicly available or other direct sources. This includes analysis from various leagues and geographies but is not to be considered comprehensive of all leagues, clubs, and women's sports organizations.

Our projected results are based on a combination of upcoming figures known to us and other—in our view—reasonable assumptions. We have not performed any verification work or audited any of the financial information contained in any articles or references for the purpose of this publication.

In relation to estimates and projections, actual results are likely to be different from those projected because events and circumstances frequently do not occur as expected, and those differences may be material. Deloitte can give no assurance as to whether, or how closely, the actual results ultimately achieved will correspond to those projected and no reliance should be placed on such projections.

By

Jennifer Haskel
United Kingdom

Paul Lee
United Kingdom

Amy Clarke
United Kingdom

Pete Giorgio
United States

Endnotes

1. Nicola Davis, "*Speed, angle and confidence: Science behind Chloe Kelly's powerful penalty*," *Guardian*, August 11, 2023.
2. Gavin Hamilton and Ben Cronin, "*Infantino claims Women's World Cup has broken even, generating revenues of \$570m*," *SportBusiness*, August 18, 2023.
3. For example, for the 'big five' European football/soccer men's leagues, broadcasting revenue dominates and are typically over half the revenue. See – Deloitte, *A balancing act: Annual Review of Football Finance 2023*, June 2023.
4. Change our game, "*The value of you can be what you can see*," accessed November 14, 2023.
5. Sponsor United, *LPGA marketing partnerships report 2022-23*, accessed November 14, 2023.
6. The Women's Sport Trust, "*Women's Sport Trust produces comprehensive industry report into the positive impact of women's sport sponsorship on brands*," accessed November 14, 2023.
7. Barclays, "*Barclays doubles investment in women's and girls' football across the FA Women's Super League*," press release, December 15, 2021.
8. Sam Carp, "*The Women's Super League and Barclays: Four takeaways from four years of a landmark title sponsorship*," *SportsPro Media*, July 10, 2023.
9. Deloitte, "*Deloitte Football Money League 2023*," accessed on November 15, 2023.
10. FIFA, *Setting the pace*, accessed on November 15, 2023.
11. Josh Sim, "*Manchester City Women name Joie as stadium naming partner*," *SportsPro Media*, September 14, 2023.
12. Chris Wright and Toe Pok, "*Arsenal launch Stella McCartney away kit for 2023-24 season*," *ESPN*, September 28, 2023.
13. GlobalData, "*FIFA Women's World Cup 2023 to net \$307.92 million in annual value from main sponsorship inventory*," *GlobalData*, July 17, 2023.
14. FIFA, "*FIFA sells out all partnership packages for the FIFA Women's World Cup Australia & New Zealand*," July 20, 2023.
15. Brett Knight, "*The highest-paid players at the 2023 Women's World Cup*," *Forbes*, accessed November 15, 2023.
16. Chloe Street, "*England football captain Leah Williamson signs massive deal with Gucci*," *Yahoo Sports*, July 12, 2022.

17. Julia Herbst, "[This CMO uses her ad dollars to get CBS, ESPN, and others to pay more attention to women's sports](#)," Fast Company, July 6, 2023.
18. Rory Jones, "[NWSL championship game draws record 915k viewers on CBS](#)," SportsPro Media, November 2, 2022.
19. Tom Bassam, "[DAZN's free Women's Champions League coverage via YouTube sees 57% increase in fan interest](#)," SportsPro Media, May 23, 2023.
20. Amanda Christivich, "[The future of Women's March Madness media rights is bright](#)," Front Office Sports, April 7, 2023.
21. NWSL, "[NWSL announces landmark media partnerships with CBS Sports, ESPN, Prime Video and Scripps Sports](#)," November 9, 2023.
22. Cesar Hernandez, "[NWSL announces new 4-year rights deal with ESPN, CBS, Prime and Scripps](#)," ESPN, November 10, 2023; Steve McCaskill, "[NWSL agrees new domestic TV deals with CBS, ESPN, Amazon and Scripps](#)," SportsPro Media, October 9, 2023.
23. David Suggs, "[Angel Reese NIL deals: How LSU star developed lucrative partnerships, from Amazon to Coach](#)," The Sporting News, April 2, 2023; David Suggs, "[Caitlin Clark NIL deals: How Iowa star developed lucrative partnerships, from Nike to Goldman Sachs](#)," April 2, 2023.
24. John Steppe, "[Iowa women's basketball sells out 2023-24 season tickets for first time in program history](#)," Gazette, August 14, 2023.
25. Andrew Cohen, "[Roku meets surging women's sports demand with new women's sports zone](#)," Front Office Sports, May 17, 2023.
26. Sarah Kuta, "[Nebraska volleyball sets new world record for attendance at a women's sporting event](#)," Smithsonian, September 1, 2023.
27. Nebraska Huskers, "[Volleyball day in Nebraska tickets sold out](#)," April 27, 2023.
28. BVM Sports, "[Iowa women's basketball team poised to shatter all-time attendance record with the Crossover at Kinnick event](#)," August 17, 2023.
29. Annesha Ghosh, "[India's WPL has captured the imagination and this is just the start](#)," The Guardian, March 22, 2023.
30. Sky Sports, "[Arsenal Women to play more Women's Super League matches at the Emirates Stadium next season](#)," May 10, 2023.
31. Transfer Markt, "[Boreham Wood FC](#)," accessed November 15, 2023.
32. Arsenal, "[Arsenal FC](#)," April 30, 2023.

33. Nestor Juarez, "[The growth of Liga MX Femenil could help Mexico break the curse of 'El Quinto Partido'](#)," Global Sports Matters, December 8, 2022.
34. AIG Women's Open, "[Walton Heath - 2023 AIG Women's Open](#)," accessed November 15, 2023.
35. Chelsea FC, "[Bold ambition for new Chelsea FC Women season](#)," July 11, 2023.
36. Chelsea Football Club, "[Ticket prices - Chelsea FC Women](#)," accessed November 15, 2023.
37. FC Barcelona, "[FC Barcelona - SD Eibar](#)," accessed November 15, 2023.
38. Ibid.
39. Scott Soshnick, Eben Novy-Williams, "[Warriors to pay record \\$50m to secure WNBA expansion team](#)," Sportico, October 5, 2023.
40. National Basketball Association, "[WNBA expansion team awarded to Bay Area, will begin play in 2025](#)," October 12, 2023.
41. Sara Germano, "[Sixth Street commits \\$125mn to buy new US women's football club](#)," Financial Times, April 4, 2023.
42. Emma Hruby, "[NWSL expansion team Angel City FC valued at more than \\$100M](#)," Just Women's Sports, August 12, 2022.
43. Reuters, "[Washington Spirit and Lyon owners join to form global women's soccer group](#)," ESPN, May 17, 2023.
44. ED Dixon, "[Washington Spirit owner Michele Kang acquires Lyon Féminin to form multi-club soccer group](#)," SportsPro Media, May 17, 2023.
45. Women's Tennis Association, "[WTA announces strategic partnership with CVC to accelerate growth of women's sports](#)," press release, March 7, 2023.
46. Abigail Gentrup, "[Five women's premier league teams announced worth \\$573M](#)," Front Office Sports, January 25, 2023.
47. ESPN cricinfo, "[Big business and Bollywood grab stakes in IPL](#)," January 24, 2008.
48. Gentrup, "[Five women's premier league teams announced worth \\$573M](#)."
49. Emily Caron, "[Dwyane Wade invests in WNBA's Chicago sky](#)," WNBA, July 14, 2023.
50. Cassandra Negley, "[Kevin Durant again invests in women's sports as part of Athletes Unlimited's \\$30 million capital](#)," Yahoo News, September 29, 2022.
51. Monarch Coll, "[Announcing Monarch's first investment: NWSL Boston](#)," accessed November 14, 2023.

52. Josh Sim, "[Lewes set for investment as Mercury 13 announces plans for global women's multi-club group](#)," SportPro Media, August 24, 2023.
53. Mark Anderson, "[Aces enjoy state-of-the-art WNBA facility as they get set to host All-Star game](#)," AP News, July 10, 2023.
54. Andrew Coohen, "[Kansas City's \\$117M made-for-soccer stadium nearly complete](#)," Front Office Sports, June 22, 2023.
55. Prime Minister of Australia, "[Securing a sporting legacy for women and girls](#)," press release, August 19, 2023.
56. Ella Brockway, "[The growing women's soccer club that no player wants to join](#)," *Washington Post*, July 18, 2023.
57. FIFA, "[Setting the pace - FIFA Women's football benchmarking report](#)," August 19, 2023.
58. Deloitte, *A balancing act- Annual Review of Football Finance 2023*, June 2023.
59. Niall McVeigh, "[England's World Cup run ends in heartache as Spain claim their first title](#)," *Guardian*, August 20, 2023.
60. Asif Burhan, "[Colombia last remaining outsiders in quarter-finals of Women's World Cup](#)," *Forbes*, August 8, 2023.
61. Al Jazeera, "[Moroccan joy as national team makes history at Women's World Cup](#)," August 3, 2023.
62. GoFundMe, "[Reggae Girlz Ris](#)," accessed November 14, 2023.
63. Keyleigh N. Wallick, *Underrepresentation of women in sports leadership: Stereotypes, discrimination, and race*, Gettysburg University, 2018.
64. Washington Spirit, "[Mentorship program](#)," accessed November 14, 2023.

Acknowledgements

Authors would like to thank **Pete Giorgio, Tim Bridge, Lizzie Tantam, Dhruv Garg, Izzy Wray, Ben Stanton, Maxine Saunders, Eliza Pierce, Rupert Darbyshire, Tamara Gaynor, Chris Hanson.**

Cover image by: **Manya Kuzemchenko**

Driven to tiers: Streaming video services look to up their profitability game with viewers

In 2024, streamers are expected to charge more for premium content, fight churn with longer subscriptions, and satisfy bargain hunters with more pricing tiers.

ARTICLE • 8 MINUTE READ

Many media and entertainment (M&E) companies face a transition. Pay TV subscriptions have been declining, along with billions in revenue.¹ The streaming video services many M&E companies launched to replace pay TV have been, in most cases, unprofitable.² That could soon change. Whether consumers want no ads, the best sports leagues and latest blockbusters or prefer to pay less for ads, last week's highlights, and last year's TV episodes, streaming services are working to have more options to suit everyone's budget. There are differences in every market, but the behaviors of US streamers may foreshadow broader trends heading for more distant shores.

In 2024, Deloitte predicts that the number of streaming video-on-demand (SVOD) tiers offered by the top US providers will more than double from the 2023 average of four possible options (with some offering seven) to an average of eight tiers (with some offering many more).³ These are expected to cover combinations of ads, no ads, access to all content, access to limited content, monthly and yearly contracts, and bundles—and potentially new innovations, such as loyalty plans. Will there be the right option for everyone, will it help the streamers grow, and will it reduce churn? Or will it be like a restaurant with a 15-page menu...so many choices that consumers can't make up their minds?

Disruption changed the terms of the game

Streaming today isn't what it used to be. As the streaming model shifts from subscriber growth to profitability, more studios are now positioned to reshape the rules to their advantage. M&E companies with streaming video services are considering charging extra for premium content, reintroducing contracts, and delivering more ads to viewers—and more ad value to advertisers. Such changes could force a reshuffling of subscribers into more profitable and enduring relationships with streamers and studios.

Digital disruptors often succeed by doing three things simultaneously: 1) satisfying persistent, unmet market needs; 2) using technology to scale their services at low cost; 3) changing the economics of the industry to their advantage. The first SVOD disruptors did all three to unseat cable and satellite TV, a business that had been highly profitable for service providers, TV networks, and studios.⁴ In doing so, they didn't just change the distribution model to streaming, they disrupted the entire business model of TV and film.

When disruption happens, it can take time to understand the implications. When streaming services started, TV networks and studios thought they had yet another profitable distribution channel for shows and movies: network TV, theaters, pay TV, physical media like DVDs—and now internet streaming services. The high-profit media and entertainment engine seemed to be humming better than ever.

Then came the sting.

As the first wave of streamers grabbed greater market share, pay TV audiences were declining.⁵ This put pressure on studio business models that were dependent on pay TV revenues. As studios began launching their own streaming video offerings, they made some tactical assumptions:

- Subscriber growth was more important than initial revenues.
- The subscription-only model was the best, perhaps the only, option.
- If they spent enough on content, consumers would retain their subscription.
- Waning pay TV revenues would be enough to support investments required to ramp up streaming content, distribution technology, and marketing.
- With audiences suddenly spending more time at home during the COVID-19 pandemic, M&E companies needed to accelerate their streaming launches while the gold rush was on.

Most of these assumptions have not held up. Subscriber growth was critical, but it likely incentivized lower prices and higher content spending to acquire and retain customers who could cancel at any time.

Escalating content costs, high churn rates, low subscription prices, and the lack of ad revenue combined to create losses in the 25-30% range—the mirror image of pay TV profits of +25-35%.⁶ Investors gave M&E streaming services less time and money to turn things around. The result of the initial disruption, the pressures of COVID-19, and the strategic decisions made many streaming video services unprofitable.⁷

While media correspondents were reporting on the “Streaming Wars,” another shift was underway. Younger generations now give more of their entertainment time to competitors offering more social and interactive forms of entertainment. For millennials and Gen Z, TV and film are no longer the primary form of media. These cohorts are just as likely to engage with highly targeted, user-generated video content (UGC) and with interactive and immersive video games that often include social experiences.⁸ They still subscribe to streaming video but are more likely to churn when they run out of compelling content. This has driven up costs for streamers whose only lever for retention has been hit shows and desirable back catalogs.

Not only has the business model for film and TV been shattered, but the landscape of entertainment preferences is evolving to include more social, interactive, and immersive experiences.⁹ In the quest for profitability, M&E companies should look to find a way to make TV & film accessible to and affordable for as many people as possible, lest they take their engagement elsewhere.

Streamers are unlikely to return to the heady profits of pay TV, but they can improve

To paraphrase Machiavelli, sometimes the least bad option is the best one. M&E companies can't rebuild pay TV networks, but they can adapt their streaming models to boost profits. They've already taken some steps to provide differentiated streaming services based on viewers' content preferences and budgets. Now they can go further with tiered offerings that could enable more households to be subscribers, with premium content pricing and early access to new releases, and with contracts that could drive greater retention.

More subscription tiers can offer streamers more ways to win

Most well-known streaming services have two subscription tiers: an expensive premium subscription without ads, and an ad-supported version that costs viewers less.¹⁰ But there are more ways streaming services can appeal to viewers, based on their desire for premium content or, conversely, a cheaper, bare-bones service.

In Deloitte's conjoint analysis of US subscriber choices, over 40% of those surveyed are willing to pay more for premium services that include bundled content, such as live sports or video games, even when ads are included with the subscription.¹¹ This content can be costly to license and produce, so the ability to charge extra may be the difference between a profitable or losing average revenue per user (ARPU).

Second, Deloitte's survey found that subscribers are willing to trade small decreases in the subscription price for higher ad loads per hour. Subscribers might not be willing to watch 14 minutes of ads in practice, especially if competitors have similar offerings with lower ad loads, or they see the same three ads repeatedly. But the principle applies to consumers: subscribers care more about what they pay than they do about ad load.

While much has been made of targeted advertising (for good and ill), streamers should make their ad channel valuable to advertisers and agencies. But they should also be working to innovate on advertising content that is more engaging for consumers. Ads can be more creative, drafting on fast-moving trends and memes, and leveraging social media influencers and creators. Streamers can also experiment with prerolling ads before the start of shows. If streamers can find the right balance between these factors—and show stronger value to advertisers—they might unlock greater profitability.

Windowing and higher prices for the best content

Deloitte's analysis also shows the pricing power—or lack thereof—that streamers have in the market. A leading brand might charge US\$19.99 and gain about 15% of the market. To equal it, a weaker brand can only charge US\$8.99. Even at US\$5.99, the weakest brand can only capture about 13%. The lack of pricing power for less popular brands impacts ARPU and can make their path to profitability more difficult. Indeed, brands with low pricing power may need to find a way to make at least a few cents from each subscriber—or consider getting out of the direct-to-consumer (DTC) business.

Brand preference is often the result of tenure in the market and high-quality content. Leading brands' pricing power can enable them to experiment more with tiering, while potentially looking at the relative value of their best content. In times past, cable TV offered some of their high-performing content as pay-per-view options, such as charging an extra US\$65 to watch an exclusive boxing match. Live and appointment viewing could potentially drive more premiums for timely access. New movies and series could reintroduce windowing for different subscriber tiers, allowing premium members immediate access while delaying access to those keeping their costs down.

Deloitte further predicts that streamers will soon reserve premium content—shows, sports, movies, and games—for two kinds of viewers: Premium subscribers who opt for a high-cost subscription and mid-tier subscribers who sign up for a year. Minimalist subscribers and nonsubscribers will need to pay a premium for the best movies, TV shows and series, games, and sports—or wait 30 days until content is available on their plan.

Fighting churn with bundles and longer-term subscriptions

Easy cancellations can enable a certain percentage of consumers to “churn,” meaning that they will sign up for a service and cancel it later—perhaps sooner. Today, streamers allow viewers to watch premium content with a minimum investment of time and money. Subscribers can binge a show or two in a few days then cancel if they like. For most streamers, it can take several months of subscription revenues just to recoup their acquisition costs.¹²

In the year ahead, some streamers will likely introduce more friction to cancellations. They may offer bundles that combine multiple SVOD services or other streaming media services at a lower overall price—and with a six-month or year-long subscription. Some may pursue more collaborations with telecoms providing mobile and data plans. This could initially scare off some subscribers, such as those that churn frequently to chase content. But lower-priced tiers and bundles that spread costs out could make it easier for more people to maintain subscriptions—and for streamers to sustain revenues.

The bottom line

Just a few years into the streaming wars and many studios that leaned into DTC are seemingly realizing how much streaming disrupted their industry and how difficult it could be to be profitable as a streamer. It has been said that history doesn’t exactly repeat but it does often rhyme, and 2024 will likely see a return of some of the mechanisms and business models that helped media and entertainment companies become highly profitable before the streaming revolution. In this way, the industry is expected to try to rewrite the new rules of the game to be more favorable to their businesses.

Pursuit of subscriber growth has generated more costs than revenues, but now profitability is driving many streamers. Streamers are trying to learn how to make money off every viewer through a tiering strategy that offers options for premium subscribers and bargain hunters. For consumers who prefer to “hit and run,” streamers may need to charge a higher price for premium content. TV and film studios are primarily in the business of producing premium content, or at least more premium than social UGC. They’re likely in the position to demand higher prices for that premium.

More studios and streamers will likely consolidate or rebundle into offerings that make it easier for consumers to find and pay for access to all the content they’re looking for. Others will likely license more of their content to other channels—or fall back from DTC and become content dealers, solely producing content to sell into any channel willing to pay. Some content may include premium pricing, and more are expected to syndicate into other providers, retreating from the cost burdens of exclusive content.

Profits may also depend on trimming costs and cutting churn, but without charging a realistic price for their content, streamers will likely be challenged to find profitability. Charging more may cut streamers’ subscriber numbers initially but could net them more valuable subscribers. Either way, they should look to reach a reasonable and profitable ARPU or consider getting out of the DTC business. Ultimately, profitability is in service to the king: compelling content. Streamers may need more content, better advertising, and novel synergies with gaming and social media. But greater ARPUs and lower operational costs may help fill their coffers more than debt financing subscriber growth at all costs.

By **Jeff Loucks**
United States

Kevin Westcott
United States

Chris Arkenberg
United States

Jana Arbanas
United States

Endnotes

1. Brad Adgate, “*With cord-cutting, cable TV industry is facing financial challenges*,” *Forbes*, October 10, 2023.
2. Peter Kafka and Rani Molla, “*The streaming boom is over*,” *Vox*, January 5, 2023.
3. In October 2023, Deloitte evaluated the various subscription options available from seven of the largest US-based streaming video providers.
4. Martina Danesi, *Disruptive innovation in the video streaming industry: the case of Netflix*, Oxford Brookes University, October 12, 2021.
5. Brad Adgate, “*With cord-cutting, cable TV industry is facing financial challenges*.”
6. Doug Shapiro, “*One clear casualty of the streaming wars: Profit*,” *Medium*, October 27, 2020.
7. *SVOD Subscriptions and Business Models | Deloitte US*
8. Kevin Westcott et al, “*2023 Digital media trends: Immersed and connected*,” Deloitte Insights, April 14, 2023.
9. Jana Arbanas, Jeff Loucks, and Chris Arkenberg, “*Can SVOD survive the future of media?*” Deloitte, 2023.
10. Ibid.
11. In the fall of 2022, Deloitte surveyed 2,328 consumers in households with paid streaming video services. The consumers we surveyed either decided which services their households subscribe to, or strongly influenced the decision. We asked respondents questions about their household’s streaming services, and their preferences for streaming video content. We also asked them to select streaming services from among a series of options. The results were then analyzed using conjoint methodology.
12. Chris Arkenberg, Paul Lee, Andrew Evans, Kevin Westcott, *As the world churns: The streaming wars go global*, Deloitte Insights, December 1, 2021.

Acknowledgements

Cover image by: **Manya Kuzemchenko**

Cinematic and interactive universes: Games and studios come together to bring the biggest stories to life

Shifting demographics, innovative technologies, and the desire for high-performing IP are driving the evolution of storytelling.

ARTICLE • 7 MINUTE READ

In 2023, some of the largest and most successful film and TV franchises began as video games, and some of the biggest video games began as film and TV shows. Adapting games to films is not new but the quality and quantity of such adaptations—and revenues from them—have been growing in recent years, following the massive popularity of video games. Younger generations are especially engaged with games, and leading game franchises earn hundreds of millions of dollars from ardent players.¹ Now, Hollywood is looking to games for new intellectual property (IP) they can expand and monetize, and game companies are eyeing TV and film collaborations to help make their IP work harder and offset soaring game development costs.² Identifying, engaging, and honoring fandoms could be key to their success.

There are far more TV shows and movies based on video games slated for production in 2024 and beyond than any time before.³ Since the pandemic, people have been returning to theaters and US box office revenues had been climbing back toward prepandemic heights, though they're still about 25% off.⁴

Deloitte predicts that the share of theatrical box office revenues from video game IP will double by 2025. We further predict that most major video streamers will include shows based on games by 2025. Although the 2023 Hollywood strikes will influence the slate of movies and shows, by 2025 we predict the trend will be back on track.

The nearly US\$200 billion global games industry⁵ is exerting a growing influence on entertainment as storytelling in games becomes more sophisticated and fans give much of their entertainment time to game titles. Notably, several of the top 50 grossing films for 2023 also include successful video game offerings within the same cinematic universes.⁶ Whether they originate in games, TV, film, or elsewhere, compelling stories are expanding across media to engage fans, reach more audiences, and lift the broader value of franchises. Not only are these media converging more often but so are the production tools that enable them.⁷

Stories can often be bigger than any single media, and more people—especially younger generations—seem to want to be entertained by them and then break open the screen and walk inside to become the main star, the hero that saves the world. More studios will likely approach their IP with serials, films, and games in mind.

Why it works

In the summer of 2023, the Super Mario Brothers movie was the most successful video game film ever and added US\$1.3 billion to Nintendo's revenues.⁸ Before that, no video game film adaptation had ever crossed US\$500 million in box office earnings.⁹ Soon after, Electronic Arts cited its game *Star Wars Jedi: Survivor* as one of the two main drivers for its quarterly earnings.¹⁰ These successes follow Sony's *The Last of Us* series of games, the second of which won Game of the Year in 2020 before the story became a major serialized hit for Home Box Office (HBO) in 2023.¹¹ Prior to all of this, Netflix launched its mobile gaming offer, which has included video games based on movies and series from its streaming titles.¹²

The lines between gameworlds and cinematic universes are blurring, with more actors, storylines, and promotions moving between them.¹³ Beyond direct revenues, these kinds of cross-over successes can lift the broader value of the franchise beyond cinema. If successful, this can set up a virtuous feedback loop where game IP drives box office revenues, which could bring more people into the game, creating more fans for other adaptations. As studios look to better monetize their most valuable and successful content, they're being drawn into much larger cinematic and interactive universes—and the many challenges in developing and curating such expansive worlds.

Tapping into passionate fans and younger generations

Central to many of these successes are the fandoms that have grown around these stories, characters, and the worlds they exist within. Fans and fan communities make these stories more durable by sharing their passions with each other, welcoming new chapters and spin-offs, and following them across TV and film, gaming, and social media, and out into the physical world of merchandise and live experiences. Continuously renewing and stoking the passion of fans has helped successful franchises to span decades and generations, but it has often required innovation and a willingness to evolve and engage with old and new fans alike.

Indeed, the rising importance of video games is being driven by generational shifts in media and entertainment behaviors. Millennials and Generation Z are less enthralled by TV and film and are just as likely to give their entertainment time to social media and video games.¹⁴ They expect to engage with content, creators, celebrities, and stories across all these channels. Among them, gamers can be highly engaged, with many giving 60-100 hours or more to vast narrative games that can cost over US\$200 million to produce.¹⁵ Players can embark upon the hero's journey and work to build relationships with game characters, overcome challenges, and ultimately save the world. Done well, such gameworlds develop strong fandoms that feel like they're both a part of the story and the franchise. Similarly, live game services drive strong engagement and retention through social bonds, multiplayer competition, and in-game economies.¹⁶

A passionate fanbase can also make it easier for franchises to expand into new media—if franchises stay true to the core values and uniqueness that make a franchise compelling. Some metrics show that most core fans would rather engage with a new installment of their favorite franchise than something new and they are more likely to buy products related to the franchise.¹⁷ However, fan communities are often empowered to voice their opinions and expectations and will likely hold studios to account if they veer too far from the canon. At the same time, more decision makers in media have now grown up with popular gaming franchises as a part of their lives.¹⁸ They may be more attuned to the expectations of fans, with a better sense of where the franchise should honor its core value and where it can stretch and evolve without disrupting its value.

The bottom line

Games not only have IP that can help fuel TV and film but they can also offer strong fandoms that can be tapped into for more engagement and monetization. But studios should be careful to honor the core principles that make a story universe unique and compelling, not only to fans but to new audiences as well. Likewise, studios may be drawn to gaming as a way to expand but they should carefully consider the challenges of doing so, while honoring fans and building relationships with game studios. They should understand, however, that it may be harder to make games from movies than it is to make movies from games. Ultimately, adapting to the media behaviors of younger generations could mean that TV and film studios may have little choice but to enter gaming.

Critically, innovations in technology are likely making it easier to develop IP strategies that move across these media. As Deloitte noted in our *2023 TMT Predictions*, virtual production tools are bringing game engines into the center of production, able to generate and deliver assets directly to TV screens, cinemas, and gameworlds.¹⁹ This can not only make it easier to develop IP for both games and video but could also enable a new platform for cross-media storytelling. However, TV and film studios may face challenges in adopting virtual production tools and hiring the talent needed to make the tools perform, and in crossing the gap from making great TV and movies to developing compelling video games.

More relationships can help: Game companies are likely well-versed in developing 3D content and experiences and may already have talent working in the game engines that power virtual production. Successful game adaptations of cinematic universes may benefit from collaborations with independent game studios to develop them. Conversely, film and TV can offer game companies ways to expand their IP into additional revenue streams while potentially driving interest and engagement back to their games. Some of the most successful video properties based on games have made a point of speaking directly to core fans.²⁰

Film and TV studios offer expertise in storytelling, cinematic experiences, and reaching and engaging audiences. Such relationships could be key to driving down exorbitant production costs—the biggest summer theatrical blockbusters and the largest AAA games can cost US\$300 million or more to develop and market.²¹ More strategic coordination could, for example, centralize digital assets and production tools that can then make it easier to develop different stories and experiences within a story universe. Combining such production optimizations and strategic planning with strong audience data could further derisk decisions about content, casting, and distribution.

Relationships and production tools can make it easier to deploy stories to different media, but the unique value of the stories themselves and the fans that cherish them are often still the core strength of a studio. The fandoms that grow around a franchise and help sustain it are frequently drawn by something unique and meaningful about those stories, characters, and worlds. Studios across media should focus on and reinforce what makes them unique and meaningful, then clearly articulate this through all channels. Studios can more easily change the storylines and main characters, and move between games, TV, and film, if the core value is truly represented in each instance. This can enable flexibility to feed fans, reach new audiences, and deliver novel experiences to younger generations, while still staying true to what makes the franchise great.

By **Chris Arkenberg**
United States

Jeff Loucks
United States

Kevin Westcott
United States

Hanish Patel
United States

Endnotes

1. Jordan Sirani, "[The 10 best-selling video games of all time](#)," *IGN.com*, October 16, 2023.
2. Martin Berg, "[Gaming is Hollywood's next IP frontier](#)," *Variety*, June 28, 2023.
3. IGN, "[All upcoming video game movies and TV shows: 2023 release dates and beyond](#)," *IGN.com*, January 5, 2023.
4. Box Office Mojo, IMDBPro, "[Domestic yearly box office](#)," accessed on November 1, 2023.
5. Tom Wijman, "[Explore the global games market in 2023](#)," *Newzoo.com*, August 8, 2023.
6. Ibid.
7. Chris Arkenberg, Jeff Loucks, Kevin Westcott, and Gillian Crossan, "[Virtual production gets real: bringing real-time visual effects onto the set](#)," *Deloitte Insights*, November 30, 2022.
8. Marei Dealessandri, "[Tears of the Kingdom drives Nintendo sales up 50% year-on-year](#)," *Games Industry.biz*, August 3, 2023.
9. Chris Eggertson, "[From Super Mario Bros. to Sonic: a box office history of video games](#)," *Boxoffice Pro*, April 7, 2022.
10. Brendan Sinclair, "[FIFA, Jedi:Survivor drive EA first quarter growth](#)," *Games Industry.biz*, August 1, 2023.
11. Matthew Higgins, "[The Last of Us – a show that surprised and challenged audiences, even those who played the game](#)," *TheConversation.com*, March 17, 2023.
12. Mike Rougeau and Phillipe Thao, "[19 Netflix games to level up and play right now](#)," *Tudum by Netflix*, April 24, 2023.
13. Kenneth Shepard, "[Cyberpunk 2077 devs wrote Idris Elba's Phantom Liberty role just for him](#)," *Kotaku*, June 12, 2023.
14. Kevin Westcott, Jana Arbanas, Chris Arkenberg, and Brooke Auxier, "[2023 Digital media Trends: Immersed and connected](#)," *Deloitte Insights*, April 14, 2023.
15. Logan Sawyer, "[The open-world games that take the longest to beat](#)," *Gamerant*, October 2, 2023.
16. Stephen Gossett, "[The real social benefits of video games](#)," *Builtin.com*, January 23, 2023.
17. Todd Spangler, "[Star Wars, Marvel, One Piece, are top-ranked entertainment franchises of 2023 to date, Fandom study finds](#)," *Variety*, September 20, 2023.
18. Cade Onder, "[Zack Snyder announces Rebel Moon video game](#)," *Comicbook.com*, August 22, 2023.

19. Chris Arkenberg, Jeff Loucks, Kevin Westcott, and Gillian Crossan, *Virtual production gets real: bringing real-time visual effects onto the set.*
20. Ashley Amber, “*HBO’s ‘The Last of Us’: 10 small details only video game fans noticed,*” *Collider.com*, January 18, 2023.
21. Ethan Gach, “*Blockbuster game development costs are out of control,*” *Kotaku*, May 1, 2023; Hannah Ziegler, “*Why studios may need to rethink summer movie blockbusters,*” *Barron’s*, July 7, 2023.

Acknowledgements

The authors would like to thank **Wenny Katzenstein**, **Ben Smith-Hellman**, and **Richard Goldsmith** for their industry expertise.

Cover image by: **Manya Kuzemchenko**

Shuffle, subscribe, stream: Consumer audio market is expected to amass listeners in 2024, but revenues could remain modest

More consumers globally are tuning in to audio entertainment formats like podcasts, streaming audio services, audiobooks, and radio.

ARTICLE • 8 MINUTE READ

Audio entertainment formats—including podcasts, streaming music, radio, and audiobooks—are hitting high notes with consumers, but increasing profits remains challenging for some providers.

Deloitte predicts that more consumers worldwide will engage with audio entertainment overall in 2024—bringing the number of monthly average podcast listeners to over 1.7 billion, monthly average audiobook listeners to 270 million, monthly average streaming music subscribers to 750 million, and monthly average radio listeners to close to 4 billion—or roughly half of the world’s population (figure 1).

Annual revenues for these formats are also modestly on the rise for the most part. Adding up estimated annual global revenues for each of these formats—including podcasts (US\$3.5 billion), audiobooks (US\$7 billion), streaming music (US\$23 billion), and radio (US\$42 billion)—Deloitte predicts the global audio entertainment market will surpass US\$75 billion in revenue in 2024, a total year-over-year increase of around 7% across these four formats.

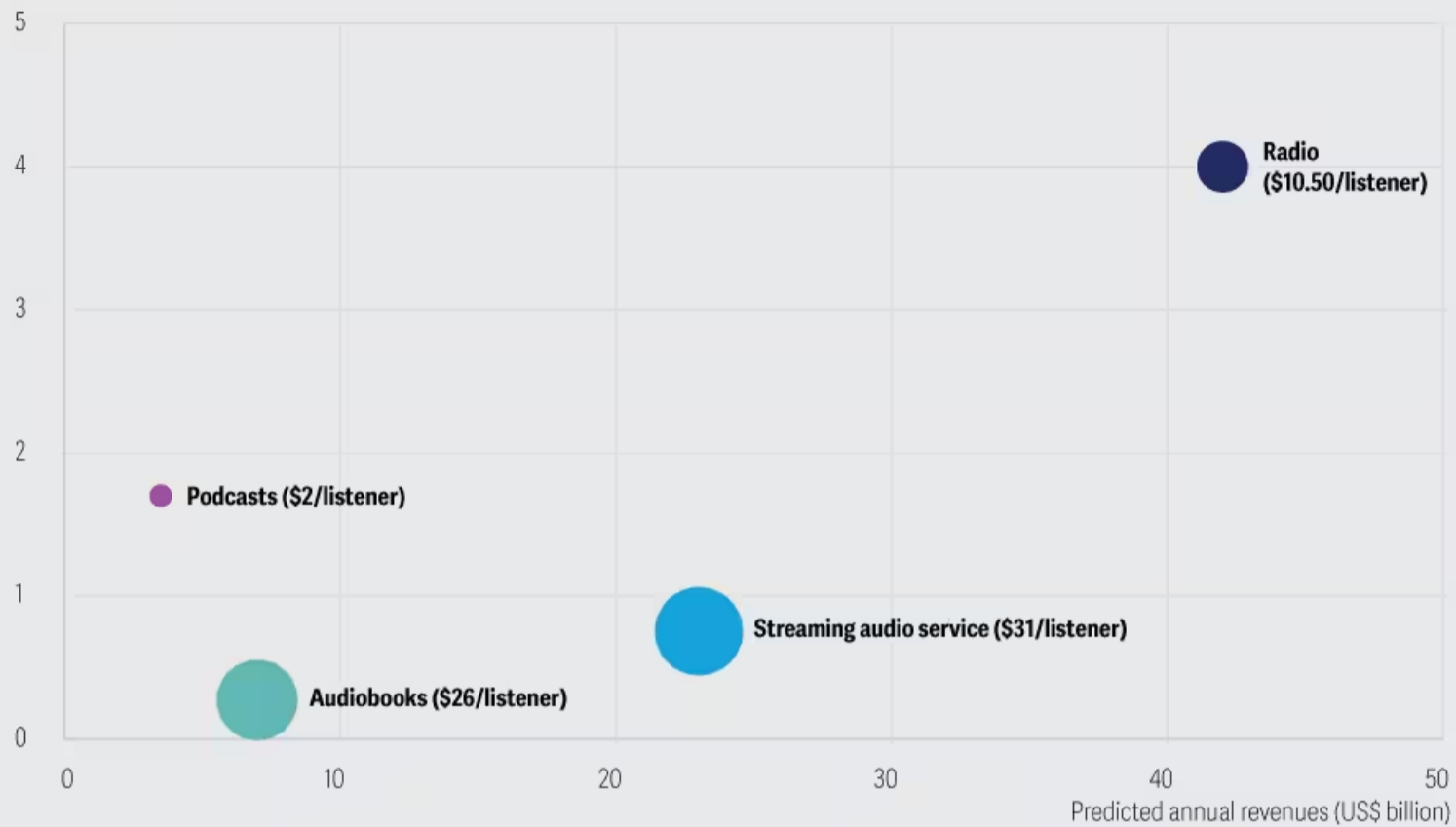
Audio has been surprisingly resilient, both during the pandemic and after.¹ But despite strong reach and listening hours, profit hasn’t always followed suit, which indicates monetization remains an open opportunity across the audio entertainment market.

Figure 1

The global audio entertainment market is expected to surpass US\$75 billion in annual revenues in 2024

● Audio medium (estimated annual revenues/listener)

Estimated number of monthly listeners (billions)



Note: All figures are projections.

Source: Deloitte analysis.

Deloitte Insights | deloitte.com/insights

Cost, content, and convenience are expected to continue to influence consumer audio market growth

The audio entertainment market will likely reach more people in 2024, largely because there is room for growth in the consumer market. For example, just a small share of global smartphone users (around 14%) pay for a streaming music service.² This represents a lucrative opportunity for platforms, content providers, and the audio industry to fill a gap and attract new audiences. The consumer audio market also provides access to the youngest audiences, namely Generation Zs and Millennials, who have largely grown up with streaming audio services and are most likely to adopt and engage with new technologies, platforms, and services.

More specifically, growth in the audio entertainment market will likely be bolstered by factors including cost, content, and convenience. Many audio formats are available for free, or at a low cost to consumers—which is notable as economic and inflationary pressures persist and people look to cut back on their entertainment spend.³ The content itself is also a big draw for audio consumers—there are songs, podcasts, books, and stations for many interests, genres, and niches—and there is a high supply of new content in the pipeline that is compelling, inclusive, and diverse.⁴ New content will be driven by independent creators entering the music and podcast spaces, and actors who might have been looking for new outlets during the Hollywood strike.⁵ Still, audio entertainment could continue to benefit from its convenience, portability, and the ease of integration into daily life, which is supported in many ways by continued consumer adoption of audio-enabling equipment such as smart speakers, headphones, and earphones.⁶

Integration and innovation will likely contribute to the podcast market

Podcasts are a growing segment of the audio entertainment market in terms of audience, but they have historically struggled to post revenues, let alone profits, due to high operating costs for companies and slow uptake by advertisers. Deloitte predicts that 1.7 billion people globally will listen to podcasts in an average month,⁷ driving the global podcast ad revenues to roughly US\$3.5 billion in 2024—a roughly 30% increase in revenue year over year.⁸ This could amount to US\$2 in annual revenue per listener. This growth in consumer engagement is expected to be supported by increased integration with other media entities, personalities and creators, and offerings. For example, there may be a trend towards popular TV shows and movies offering companion podcasts to supplement their on-screen offerings and online creators developing their own shows to extend their viral brand—both of which can provide greater diversity of content and draw in wider audiences from other platforms.⁹

Notably, however, could be the expected increased availability of podcasts and stronger show content. Some companies and platforms continue to invest in podcasts, making them more accessible to a larger cohort of people.¹⁰ And while the number of new podcast shows has declined since the pandemic, the number of new episodes has remained stable. This may mean fewer people are experimenting with (and then quickly abandoning) the medium.¹¹ But there is still room for innovation: Some podcast platforms now offer video podcasts,¹² which promises to appeal to those consumers who prefer visual entertainment formats. All these trends will likely draw more listeners and increase monetization opportunities for audio providers, creators, and advertisers alike, which may begin to offset the challenges that have halted profitability thus far.

Audiobooks get a boost from pandemic era reading habits and AI

Book sales and reading saw pandemic bumps, but as people resume their normal activities, *listening* to books remains a way to integrate reading habits into daily life. Growth in the audiobook market has been trending upwards (due in part to the increase in titles and easier access to audiobooks) and we expect 270 million people worldwide to listen to audiobooks in an average month in 2024—representing an increase in listenership of 15% year over year¹³ and US\$26 in annual revenue per listener. As audiobooks increase in popularity, Deloitte predicts that audiobooks will account for approximately six percent of total book sales worldwide, representing a 26% increase in sales year over year.¹⁴ This could bring global audiobook revenues—which are largely driven by subscriptions—to approximately US\$7 billion.¹⁵ In the US alone, which is the largest market for audiobooks, we predict that audiobook sales will surpass e-book sales and account for around 10% of book revenues in 2024.¹⁶

With popularity comes competition: With several streaming audio platforms having entered the audiobook market recently, competition is stiff.¹⁷ Platforms with a strong user base for other streaming audio services may be able to capture audiobook market share among their existing audience, especially if they can provide seamless interfaces that make it easy for listeners to consume the bulk of their audio entertainment content in one place.

Generative AI also represents a growth opportunity for the audiobook market with Apple Books and Google Play rolling out books read by AI-generated voices and Google Play offering autonarration for some selected titles.¹⁸ Although using AI to mimic the nuances of human speech is complex, these technologies could extend audio formats' accessibility to a wide range of books.

Growth in streaming audio will likely rely on consolidation and quality

Streaming audio services are popular among consumers, but there is still space in the market for growth and innovation. Deloitte predicts that more than 750 million people globally¹⁹ will subscribe to a streaming music service in an average month in 2024, and revenues will grow by roughly 15% year over year to reach US\$23 billion.²⁰ This could amount to US\$31 in annual revenue per listener, the most lucrative across these audio formats. It's worth noting that a segment of this audience, and of the revenues generated, comes from podcasts and audiobooks, both of which are available across several popular streaming audio services.

Subscriber growth on streaming audio services—and with it, revenues—are expected to be driven in part by consolidation and bundling within services, which could offer more value to subscribers and make subscriptions more appealing. For example, Spotify and iHeartRadio offer several audio formats within their apps—from music and audiobooks to podcasts and live events.²¹ Additionally, many services have expanded their subscription offerings and introduced ad-supported tiers, making subscriptions accessible to more listeners, a trend we expect to see adopted across other streaming audio services. Not to mention that resurgence of live concerts and major music tours²²—and the buzz they create—may be adding to streaming music consumption. Some streamers, such as Spotify, are innovating with AI technology to make the music listening experience more personalized to each user.²³ And to appeal to the most dedicated audiophiles who may be late adopters, many streaming services are investing in improving sound quality by rolling out lossless audio (where compression results in no loss of information) and high-fidelity audio features and tracks.²⁴ However, it's worth noting that lossless audio can't be accessed with most Bluetooth devices.

Cost and accessibility continue to make radio a popular medium, especially in developing markets

Radio's free audio entertainment, which is supported by ads, remains resilient amid an ecosystem of paid entertainment services. Inflationary pressures may contribute to its enduring popularity, particularly in developing markets. Deloitte predicts that around half of people globally (or close to four billion people) will listen to radio monthly in 2024, with radio ad revenues reaching US\$42 billion, indicating modest (approximately 1.5%) growth year over year.²⁵ Annual revenues per radio listener would amount to US\$10.50.

Notably, in the US, consumers across all age groups are likely to listen to radio monthly and young Americans (ages 18 to 34) are slightly more likely to be spending time with radio than live or time-shifted TV.²⁶ While radio listening time may be declining—or stabilizing—in some Western markets,²⁷ in other parts of the world, such as Africa, radio remains a widely used medium, especially for accessing news and information.²⁸

Still, radio broadcasters should consider the changing audio entertainment landscape, namely the shift towards digital offerings in many places and the potential for future FM switch offs.²⁹ If people are listening to music on a smart speaker or in the car, for example, what will make them choose to play a radio station versus a music streaming service? Broadcasters should also consider how generative AI could impact the future of radio: with live AI-powered deejays³⁰ or AI-created advertisements, which could help address a continued challenge for radio broadcasters, namely, generating advertising revenue. On the bright side, advertisers may find less competition in the radio space, as competitors migrate to other audio formats, which lowers cost per milles (CPMs) and offers a cost effective way to reach the masses.

The bottom line: the growth opportunities for audio entertainment

Audio entertainment formats are often overlooked because they are rarely a primary activity. People listen to radio while driving, audiobooks while doing the laundry, podcasts while walking the dog, and streaming music while working. However, being a secondary activity might be audio's greatest strength.

Though the consumer audio market (especially streaming audio and podcast formats) has traditionally struggled to generate revenue and profits, there are opportunities for providers, content creators, and advertisers to enter and innovate within the industry to attract new audiences, develop bespoke content, and spread brand awareness.

More audio platforms, if confident in their versatility, might consider consolidation and offering multiple audio formats—such as music, podcasts, audiobooks, and live events—in a single application or platform. This could allow platforms to charge more for their subscriptions, while attracting a wider pool of subscribers. In fact, nearly 70% of US consumers surveyed would like to have access to audiobooks, podcasts, and music all in one service.³¹ In addition, audio platforms should work to develop stronger content recommendation and curation systems to avoid consumer churn and keep consumers engaged. However, providers should also calculate what they can do well. For some, that might mean focusing on one audio format and fine-tuning their strategy around that deliverable, rather than trying to do it all.

Audio platforms and providers also have an opportunity to expand their revenue streams and increase profits. Streamers and other audio providers might consider investing in advancements in advertising technology to make audio entertainment more “shoppable”—enabling listeners to hear about a product and instantly buy it with minimal friction³²—therefore attracting sponsors and increasing cashflow.³³ For their part, brands should lean into these audio formats for their ad spots: Ads on podcasts are seen as unintrusive and can reach consumers who are well-aligned with brand interests and missions.³⁴ They could also work with well-known creators, or podcast hosts themselves, to read the ads or endorse their products.

Audio platforms and providers may also want to focus on securing relationships that can make their service easy for consumers to access, for example, having their app available or preset in new cars or integrating with gaming platforms which have proven to be popular for music discovery.

On the other hand, companies that have invested in podcasts to the tune of billions of dollars may want to consider their time horizons. The industry is growing, but not as quickly as many hoped in terms of revenues. Some might decide to exit in 2024. Equally, investing in podcasts may still make sense for some, but given monetization challenges, valuations should be a fraction of what we saw in 2020 through 2022 to generate plausible returns on investments.

Still, other media and entertainment companies could stand to benefit from this burgeoning audio market. Content creators and providers, such as movie and television producers, book publishers, and game developers, may want to think about their audio entertainment strategy early on. This could include developing a companion podcast show, having talent on podcast episodes as guests, producing an audiobook version of a popular print book, or commissioning artists or licensing music for video game soundtracks. This strategy can help to extend the life of IP, franchises, characters, and fictional worlds, reach new audiences, and keep dedicated fans engaged year round.³⁵

Thinking about audio as part of an overall media and entertainment strategy will likely be a *sound* business practice into 2024.

By **Brooke Auxier**
United States

Bree Matheson
United States

Duncan Stewart
Canada

Kevin Westcott
United States

Endnotes

1. Nielsen, “*Radio is ‘comfort food’ as media consumption rises amid COVID-19 pandemic*,” accessed on November 3, 2023; Sara Fischer, “*Audio takes off during the pandemic*,” Axios, February 15, 2021.
2. Batya Levi, John Hodulik, and Ryan Gravett, “Warner Music Group Corp - Initiate with a neutral: Music on hold,” UBS Global Research and Evidence Lab, June 29, 2023.
3. Kevin Westcott, Jana Arbanas, Chris Arkenberg, Brooke Auxier, *2023 digital media trends: Immersed and connected*, Deloitte Insights, April 14, 2023.
4. Market.Us, “*Podcasting is becoming popular because of its accessibility, flexibility, and wide range of content, which caters to diverse listener interests*,” Yahoo! Finance, May 15, 2023.
5. Duncan Cooper and Zing Tsjeng, “*The Hollywood strike will affect way more than movies and TV*,” VICE, July 14, 2023.
6. Magna Intelligence, “Global Earphones and Headphones Market Databank,” September 16, 2021; Research and Markets, “*Worldwide smart speaker industry to 2028*,” GlobeNewswire, December 1, 2022.
7. Deloitte analysis based on: Georgina Howes, “*Podcast market forecast: 2022-2027*,” Omdia, October 27, 2022.
8. Deloitte analysis based on: Georgina Howes, “*Podcast market forecast: 2022-2027*,” Omdia, October 27, 2022 and Aron Kressner, “*Media 2023: Podcasting after the hyper, by the numbers*,” Forbes, February 2, 2023.
9. Brooke Auxier, *Podcasts and their hosts create an environment ripe for influence and discovery*, Deloitte Insights, 2023.
10. Peter Suci, “*Podcasts added to YouTube Music—Will it expand the audience*,” Forbes, May 8, 2023.
11. James Cridland, “*Have the number of new podcasts really ‘plummeted?’*” PodNews, January 25, 2023.
12. Ariel Shapiro, “*Spotify is going big on video podcasts*,” The Verge, March 8, 2023.
13. Deloitte analysis based on: Georgina Howes, “*Digital audio forecast report 2022-2027*,” Omdia, December 8, 2023; Reuters, “*Spotify takes on Amazon’s Audible, launches audiobook service for U.S. users*,” September 20, 2022; Ariel Shapiro, “*Spotify comes for audiobooks*,” The Verge, June 9, 2022.
14. Dimitrije Curcic, “*Audiobook statistics*,” Wordsrated.com, June 7, 2023.
15. Deloitte analysis based on: Georgina Howes, “*Digital audio forecast report 2022-2027*.”
16. Porter Anderson, “*US book publishing revenues in 2022 were \$28 billion: AAP StatShot*,” Publishing Perspectives, May 31, 2023.

17. Reuters, "[Spotify takes on Amazon's Audible, launches audiobook service for U.S. users](#)," September 20, 2022.
18. Jordan Hart, "[AI-generated narrators threaten voice actors hoping for audiobook gigs](#)," *Business Insider*, June 17, 2023.
19. Deloitte analysis based on: Mark Mulligan, "[Music subscriber market shares 2022](#)," Midia, December 7, 2022; Doug Anmuth, US Internet Guidebooks, March 2023, JPMorgan Research, March 8, 2023.
20. Deloitte analysis based on:IFPI, [The Global Music Report 2023](#), 2023.
21. Spotify, "[Spotify reveals more opportunities and features for creators during stream on](#)," press release, March 8, 2023; Victoria Song, "[Spotify appears to be planning free audiobook trials for subscribers](#)," *The Verge*, September 8, 2023.
22. Ben Sisario, "[Live music is roaring back. But fans are reeling from sticker shock](#)," *The New York Times*, April 7, 2023.
23. Spotify, "[Spotify debuts a new AI DJ, right in your pocket](#)," press release, February 22, 2023.
24. Tucker Bowe, "[What exactly is 'lossless' audio?](#)" *Gear Patrol*, June 23, 2023.
25. Deloitte analysis based on: Duncan Stewart, [TMT Predictions 2019, Radio: Revenue, reach and resilience](#), Deloitte Insights, December 11, 2018; Deloitte analysis of audiences share by individual market.
26. Nielsen, [Audio today 2023: How Americans listen](#), June 2023.
27. DigitalRadioUK, [Radio and audio review: International market report](#), September 21, 2021; Gill Hind and Bella Monkcom, [Where is tomorrow's listener? UK radio in 2023](#), Enders, September 14, 2023.
28. UN News, "[Radio: The universal medium that leaves no one behind](#)," February 11, 2022; Farai Mutsaka, "[Many rely on radio broadcasts in Zimbabwe and across Africa](#)," The Associated Press, March 27, 2023.
29. Roy Martin, "[No FM switch off in the UK until at last 2030 says DCMS](#)," *RadioToday*, October 21, 2021.
30. Maria Diaz, "[Would you listen to AI-run radio? This station tested it out on listeners](#)," ZDNet, May 2, 2023.
31. Kevin Westcott, Jana Arbanas, Chris Arkenberg, Brooke Auxier, [2023 Digital media trends: Immersed and connected](#), Deloitte Insights, April 14, 2023.
32. Brooke Auxier, [Podcasts and their hosts create an environment ripe for influence and discovery](#); Brooke Auxier, [The future of shoppable media can build on the success of social shopping](#), Deloitte Insights, 2023.

33. Gillian Follett, “*Podcast ads universally boost brand metrics—and longer ads work better*,” *AdAge*, October 26, 2022.
34. The Infinite Dial, *The podcast consumer 2023*, Edison Research, 2023.
35. Brooke Auxier, *Podcasts and their hosts create an environment ripe for influence and discovery*.

Acknowledgements

The authors would like to thank **Akash Rawat** for his contributions to the development of this piece. They would also like to thank **Jana Arbanas, David Cutbill, Alan Shaub, Paul Lee,** and **Matt Varraveto** for their thoughtful reviews.

Cover image by: **Manya Kuzemchenko**

Will endless low cost content do to gaming what it did to TV and film?

UGC gaming platforms that incentivize user-generated content are expected to make big payouts to creators in 2024. As this practice grows, however, it could put pressure on top-tier games and services.

ARTICLE • 7 MINUTE READ

Two popular video games have enabled hundreds of millions of players to be 3D creators. Now, a third heavyweight has launched new tools and incentives to stoke the fires of 3D user-generated content (3D UGC) on their own platform.¹ More people have been creating 3D content, and top 3D UGC platforms have been paying out more money to them. Deloitte predicts that these platforms will pay out almost US\$1.5 billion to their content developers in 2024. Deloitte further predicts that the number of paid independent developers on 3D UGC gaming platforms will exceed 10 million.

This prediction is buoyed by the success and inertia of 3D UGC platforms for players, creators, and brands; the arrival of new tools and incentives to make 3D content easier to create and monetize; and the strength of the UGC gaming model for video and social media.

Expanding the creator economy for games could drive greater innovation in game experiences and digital goods, stronger engagement for 3D UGC platforms, and more favorable economics for both platforms and creators. These platforms likely see an opportunity to democratize gaming and interactive experiences. But by unleashing endless cheap 3D content, will they disrupt the industry?

What is 3D UGC?

Like user-generated video platforms and social media, 3D UGC democratizes content creation for games. On 3D UGC platforms, users can create, share, and monetize digital goods, like skins and cosmetics (things an avatar might wear), and equipment (gear that might offer some utility to the player). Platforms have developed robust marketplaces for creators to trade in such digital goods.² Users can also create entire games and immersive 3D experiences, then leverage discovery tools provided by the platform. Like social media, these user-generated games and their creators can become very popular, drawing millions of players and fans each, and generating valuable brand relationships.³

Although it's gotten much easier and more popular, there have always been some players—and some game-makers—that have seen the appeal of creating in games. Modders, as they're called, have brought fan fiction into games, enabled personalization with digital goods and cosmetics, and added features that players were clamoring for, but game companies were unable to prioritize.⁴ Some mods have evolved into hit games of their own,⁵ while others have directly influenced how game companies have grown their franchises.⁶

Leveraging a distinct blockiness that may have made content creation easier by presenting a LEGO-like building experience, Roblox launched in 2005 as a platform for users to develop and share their own games and digital goods. In 2009, Minecraft offered its version of 3D in-game construction adding resource management and survival mechanics. Both have worked to lower the barriers to 3D modeling and game creation, spawning creators, companies, marketplaces, and other games offering 3D creator capabilities. In 2023, Epic Games announced new creator capabilities for its popular multiplayer franchise, Fortnite, on its Unreal Engine development platform, diverging from the blockiness of its competitors with more realism. To incentivize game creators on their platform, they also announced revenue sharing opportunities based on engagement.⁷

User generated content could shift the economics of gaming

Developing blockbuster story-driven games has become very expensive and complex, and such games compete for engagement with top live game services. An explosion of cheap or free 3D user-generated content could draw more people away from top-tier games and game services and into simpler, more commoditized and niche games. For the leading games studios and publishers that tend to dominate engagement and revenues, 3D UGC gaming platforms could become an alternative destination for immersive and interactive experiences.⁸

For game studios, building and delivering great games has become an increasingly challenging and expensive endeavor. Making a hit game that might sell for US\$65 can take years to develop and cost upwards of US\$300 million to produce and market. Games that run as live services shifted the economics of gaming from making money on the launch of a new game to recurring in-game revenues from personalization with digital goods, new content experiences within the game service, and access to special events. Such games often thrive on the social nature of their services but also should continue to deliver new experiences to keep players engaged.

Platforms for 3D UGC lean into this by deputizing and incentivizing players to keep adding new content to their services. The platform itself can radically lower the cost of game development and marketing for creators while enabling them to monetize their passions and build fandoms. Like user-generated video services and social media, this can lead to highly diverse content catering to the most popular and the smallest niches.

How platforms are making 3D UGC easier to create

One challenge for 3D UGC is that it is typically more difficult to produce than video content. This difficulty of building in 3D modeling tools has historically kept it a narrow pursuit, but some game companies have steadily been making it easier.⁹ Power users can attain greater capabilities while nontechnical creators can develop compelling games. Platform marketplaces enable creators to trade in texture packs and shaders, 3D objects to populate the game, and game mechanics that package interactions like opening a door or driving a vehicle.¹⁰

With the rise of generative AI, providers are actively working to make 3D content creation easier.¹¹ These tools are emerging to automate some parts of coding and testing for more involved game elements; they enable faster development of dialog and localization into different languages, as well as visual generation of character and set design; they support early experiments in non-player characters that can engage in conversations rather than reciting a few canned lines; and at the edges of innovation, they are working to enable natural language prompts that can generate games on the fly.¹²

Making UGC easy to create and share, and enabling creators to monetize their work, has driven a near-infinite tide of content for social media. This has put more pressure on premium content, drawn audiences away from TV and film, and changed the advertising landscape. If 3D UGC platforms are successful, they could bring about a similar sea change for gaming.

The bottom line

For the games industry, 3D UGC could be a double-edged sword. Enabling and incentivizing much greater 3D UGC could amplify innovation in games, digital goods, and monetization while fostering more interest in and engagement with a rich diversity of interactive experiences. Such innovation might not only drive the expansion of an independent 3D creator economy, but some of those creators may also move on to establish successful game studios of their own, bringing yet more diversity and competition to the industry. Additionally, some of those games and gaming innovations could matriculate into stronger game studios, further stoking their own competitive capabilities.

At the same time, however, history has not been kind to incumbents when their control over content gets disintermediated and democratized. Some of the challenges confronting film and TV—high costs of content development, disruption of distribution channels and lucrative theatrical windows, and the shifting behaviors of audiences—have arguably been exacerbated by the rise of endless free content, lofted on the winds of social networks and the preferences of younger generations.

Leading platforms for 3D UGC have been around for a while without disrupting the games industry, but as their games and experiences become richer, more innovative, and easier to develop and monetize, they could attract more attention and engagement. Greatly amplifying 3D UGC could create more fragmentation of gaming content and creators while consolidating them onto a handful of platforms. This could shift the economics towards gaming and interactive experiences that are often cheaper to develop.

At nearly US\$250 billion annually, game studios may be in a stronger position than TV and film studios have been. Younger generations are flocking to gaming in its many forms, and more game IP is producing hit movies and streaming TV.¹³ Still, video game companies may face a similar shifting landscape between premium and commodity content. Developing premium games has become costly, and top game companies could face greater competition and cost pressures from 3D UGC, drawing more players to a seemingly infinite abundance of cheap or free interactive experiences. At the same time, 3D UGC platforms could enjoy an abundance of 3D content that could cost them far less, but they may wrestle with managing, moderating, and monetizing so much commodity content, as well as with the challenges of copyright infringement, harmful content, and looming regulations. In the middle, companies that are neither commodity 3D UGC nor premium studio-led may need to pick a side to survive.

On the way to their metaverse visions, platforms for 3D UGC may succeed in democratizing game development for the masses. This could lead to an explosion of creativity and innovation for gameplay, and also new forms of immersive and interactive content and experiences. But in doing so, 3D UGC platforms could disrupt their own industry.

By **Chris Arkenberg**
United States

Jeff Loucks
United States

Kevin Westcott
United States

Endnotes

1. Jay Peters, “*Epic now rewards Fortnite creators for how long people play their experiences*,” *The Verge*, June 1, 2023.
2. Jay Peters, “*Epic is merging its digital asset stores into one huge marketplace*,” *The Verge*, March 22, 2023; Brendan Sinclair, “*Lessons from Minecraft Marketplace*,” *Games Industry.biz*, June 6, 2022.
3. ActivePlayer, “*Roblox live player count and statistics*,” accessed November 7, 2023; Melody Day, “*Most popular brand collaborations with Minecraft*,” *Net Influencer*, June 18, 2023.
4. Charles Burgar, “*Skyrim: 10 features you always wished it had (that are added by mods)*,” *Game Rant*, October 20, 2020.
5. Joakim Henningson, “*The history of PlayerUnknown’s Battlegrounds*,” *Red Bull*, August 27, 2020.
6. Scott Reismanis, “*10 influential mods that transformed the landscape of gaming*,” *Medium*, June 16, 2021.
7. Fortnite team, “*Introducing Creator Economy 2.0*,” *Fortnite creator portal*, March 22, 2023.
8. Newzoo, “*Top public games companies by revenues*,” accessed November 7, 2023.
9. Rebekah Valentine, “*Turns out the hardest part of making a game...everything*,” *IGN*, August 19, 2021.
10. Forte, “*Game marketplaces survey*,” *Medium*, April 1, 2021.
11. Jay Peters, “*Roblox’s new AI chatbot will help you build virtual worlds*,” *The Verge*, September 8, 2023.
12. Joshua Lu and Robin Guo, “*The generative AI revolution will enable anyone to create games*,” *Andreessen Horowitz*, March 17, 2023.
13. *IGN*, “*All upcoming video game movies and TV shows: 2023 release dates and beyond*,” January 6, 2023.

Acknowledgements

The authors would like to thank **Ankit Dhameja** for supporting research and **Richard Goldsmith** for his industry expertise.

Cover image by: **Manya Kuzemchenko**

Smartphone authentication: The killer app that can augment the smartphone's utility

Smartphones can help you securely log in, tap to pay and enter cars, buildings, and airports. As security concerns grow, the need for authentication will likely consolidate the smartphone's status as the ultimate goldilocks device: the right size, power, connectivity and trust.

ARTICLE • 8 MINUTE READ

In 2024, the smartphone is expected to have one of its most successful years ever, despite a historically modest sales forecast of 1.26 billion units,¹ several hundred million units short of its all-time peak of 1.57 billion.² The smartphone's success will be due in part to the ever-rising value of its five-billion user ecosystem—with authentication becoming an increasingly important addition to the device's value.

Smartphones are expected to be used to authenticate processes across an ever-widening range of actions, such as accessing websites, making payments, unlocking cars, and controlling entry to physical buildings trillions of times in 2024. And in the medium term, the volume could swell to tens of trillions in a year with additional applications including proving identity, becoming mainstream. In 2024, usage of the smartphone as an authenticator will still likely be a fraction of its long-term potential impact, as illustrated in Figure 1.

Figure 1

The smartphone’s anticipated usage as an authenticator in 2024 is significant and also modest relative to potential usage

Application	Current smartphone usage for authentication	Total addressable global market
Access to online accounts (email, social media, e-commerce work applications)	1.3 trillion passcodes sent to smartphones in 2023 ^a	4.3 billion email users, 2.6 billion e-commerce customers in 2023 ^b
Access and use of cars	The majority of the 60 million cars sold annually in 2024 offer smartphone ^c authentication as an alternative to physical keys	1.5 billion cars in use ^d
Access to planes	Primarily based on mobile boarding passes; default for one major airline ^e	There were 4.5 billion airline passengers in 2019 (pre-Covid) ^f
Access to houses	Primarily based on physical keys	There were about 198 million households in the EU in 2022 ^g and 127 million households in the US in 2020 ^h
Airline travel	Primarily based on physical documentation	4.5 billion airline passengers in 2019 ⁱ
Office entry	Primarily based on physical passes	> 1 billion office workers ^j
Access to public transport	14 countries globally supporting phone or watch-based entry ^k	239 billion journeys per year in 39 countries alone ^l
In-store payments	84% of adults use mobile phone wallets in China; ^m in the United States, around 6% of in-store spend was via mobile wallet in Q2 2022 ^o	US\$1.5 trillion retail spend excluding e-commerce, Q2 2023, in the United States ⁿ

Source: Deloitte, 2023, based on publicly available third-party sources including: Juniper Research, Adobe Experience Cloud, Kelley Blue Book, Hedges & Company, Emirates, the International Civil Aviation Organization, Eurostat, Pew Research Center, The World Bank, Apple, International Association of Public Transport, Electronic Payments International, PYMENTS.com, and U.S. Census Bureau News. *See “Figure 1 endnotes” for more detail.

Deloitte Insights | deloitte.com/insights

Authenticating online access: The smartphone is a common link between two-factor authentication and passkeys

The smartphone is likely to play an increasing role in managing fraudulent access to online accounts. In 2024, it is expected to predominantly be used for delivery of passcodes as part of the two-factor authentication (TFA) process, whereby a one-time password (OTP) is sent to a phone, often within a text message.³ In 2023, there were an estimated 1.3 trillion such messages sent via telecom networks, generating an estimated US\$26 billion from network traffic alone.⁴

The smartphone may also be used increasingly to generate passkeys—likely the medium-term replacement for passwords. Passkeys authenticate access to online accounts without passwords.⁵ With this approach, a pair of keys is generated for every account, one public and one private key stored on a phone. Users who want to access an account check that the keys match. The private, device-based key is released once the user has been validated using the same process that would normally be used to unlock the phone, which could be biometric (face or fingerprint) or via a password or pattern. As of 2024, usage of passkeys could be modest; by 2030 usage may become higher as it could supplant TFA.

One driver for smartphone-based authentication, based on either technology, is the growing average number of online accounts and associated volume of breaches. The cost of attacks, which are predicated on vulnerabilities that exist via passwords to authenticate access to a growing number of online accounts, is likely unsustainable. Users are asked to create a unique and, ideally strong, password for each account with some enterprises requiring that workers change their password quarterly. The relatively static and limited human ability for recall cannot cope with the combination of the growing number of accounts held, and the ask to memorize rising numbers of “strong” passwords.⁶

The outcome is an abundance of weak passwords, with the most popular still being “123456” and “password.”⁷ Repositories of passwords paired with user IDs are often targeted and there were an estimated 24 billion passwords—one for every three people on earth—exposed by hackers in 2022.⁸ The annual cost of data breaches is forecast at more than US\$5 trillion in 2024.⁹ Furthermore, passwords are often repeated: One analysis found 64% of people used the same password across multiple accounts, incrementing the impact of a breach, as a single user ID and password combination may unlock multiple accounts.¹⁰ Password users can also be vulnerable to phishing attacks, designed to trick users into sharing credentials with malign entities. An estimated 3.4 billion malicious emails are sent daily.¹¹

Two-factor authentication and passkeys can provide another level of security, as additional information beyond the pair of password and user ID is required. Either approach can be effective at minimizing the impact of breaches. It can repel almost all automated bot attacks and bulk phishing attacks.¹² A very high proportion of compromised accounts did not use multi-factor authentication.¹³

TFA may incur a charge for the delivery of the one-time password, while pass keys do not (aside from bandwidth usage);¹⁴ the cost of each TFA may limit how frequently smartphone-based authentication is triggered. One deployment of passkeys found that using smartphone-based biometric authentication enabled a two-thirds reduction in the number of OTPs per user, saving 1.9 pence (US 2.4 cents) per message.¹⁵ The momentum behind passkey is likely to grow further to the commitment by Apple, Microsoft, and Google in May 2022 to support the same passkey standard.¹⁶ Apple launched support for passkeys with iOS 16 in September 2022,¹⁷ and Google supports these for all operating systems from Android 9.0.¹⁸ A growing, but still selective number of companies supported passkeys as of September 2023.¹⁹

Authenticating commerce

The smartphone is also likely to play a growing role in authenticating transactions, both online and in stores. This is expected to be based on a range of technologies, including biometric identification capabilities. Two- or multi-factor authentication and passkeys will likely also play an important role.

Mobile already represents a significant share of all e-commerce, but the majority of purchases made are still offline. In the United States, almost half of sales (47%, equivalent to US\$99.5 billion) were via mobile in the 2022 holiday season, up from 43% the prior year.²⁰ But as of Q2 2023, e-commerce was still 15.4% of all payments, up one percentage point year-over-year.²¹ E-commerce’s share of all sales has tended to increase steadily since the mid-90s, except for the anomalous period of 2020–2021.²² If this trend continues, the role of smartphones for online transactions should continue to increase.

In-store, the impact of payments via smartphone apps is smaller. According to one analysis, only three cents of every dollar spent in-store in Q2 2022 in the United States was paid via a smartphone app.²³

Authenticating physical access

The smartphone is also likely to be used increasingly to validate physical access into buildings. Buildings that use card readers to permit or prohibit access often use Near-field communications (NFC) to exchange information between the gate and the card. The first smartphones to incorporate NFC were launched in 2011 and as of 2024,²⁴ this capability is likely to be ubiquitous. As such, smartphones could substitute for access cards. Alternatively, Bluetooth could be used to communicate with the reader. With hardware and software upgrades (of a complexity and cost that varies by company), existing gates could be ready to work with smartphones.

The proportion of businesses whose premises are mobile-ready has been growing steadily. According to one survey of businesses in North America, EMEA, and Asia Pacific, 24% were mobile-capable in 2022, up from 16% in 2020, and a further 42% were planning to upgrade.²⁵

This migration to smartphones could save on operational costs, address risks as well as reduce environmental impacts. Smartphone-based entry passes can be distributed via app downloads. They can also be canceled remotely via over-the-air instruction. Businesses around the world are likely to have teams dedicated to the allocation of physical cards to staff and visitors and temporary cards to replace forgotten cards in 2024. A phone-based approach would still require oversight and could enable some staff to be re-assigned from their current repetitive role of handing out entry cards.

There will likely be debates on the risks of migrating to smartphone-based entry this year. One benefit of the ID card in most buildings is visible identity; however, in some offices ID cards are often pocketed and identity checks are not often commonplace. ID cards may also be stolen, enabling bad actors to gain access to a building if there is lax security. By contrast, a smartphone's biometric authentication could be used to provide further validation prior to tapping the phone on the reader, like the process for making a payment or entering the subway. Additionally, individuals may leave their ID card at home or elsewhere (including public venues) but may be more vigilant about their smartphones because of their utility.

The sustainability dividend from migrating to mobile may be significant. The traditional, legacy method of validating access to buildings is via photo identity cards housed in lanyards. There are 3.4 billion people in the global workforce.²⁶ If only half of these are issued a lanyard, that implies almost two billion lanyards, some of which may end up landfill. There are also multiple temporary ID cards issued for events. The Fira Barcelona hosts 2.5 million visitors each year;²⁷ the Las Vegas Convention Center has two million visitors.²⁸ Some shows are already migrating to smartphone-based digital access passes, including the Mobile World Congress at the Fira Barcelona, saving on the need to manufacture and subsequently dispose of physical passes and lanyards, and on the need to dedicate staff to issue them.²⁹

Smartphone access passes could also be used for other functions, such as payment at vending machines, access to printing machines, and checking in to events, such as university lectures, or conference sessions. In the United States, there were 53 universities that had adopted smartphone passes as of September 2022.³⁰

As well as authenticating access to commercial premises, over time smartphones may also be used more commonly to permit entry into private homes. One benefit of this would be the ability to send keys to guests on a time-limited basis.³¹

Authenticating travel

Prior to the pandemic, there were 4.5 billion airline passengers per year in 2019.³² Prior to boarding, passengers need to show their boarding pass and may also need to show identification. A boarding pass can be within a mobile app, particularly for regular travelers. This can save on printing and reduces the likelihood of loss. Some baggage receipts are also moving to apps.³³

However, proof of identity at a caliber sufficient for travel is slowly moving online. One of the most advanced countries is Ukraine, which launched an app in 2020 that hosts multiple documents including a national identity card. As of December 2022, almost 18.5 million Ukrainians (more than 40%) had downloaded the app.³⁴ In the United States, three states have launched support for digital driving licenses: Arizona, Georgia, and Maryland.³⁵ In 2024 and 2025 a European National Identity initiative co-funded by the EU plans to trial smartphones for applications including mobile driving licenses.³⁶ In the UK, the government is targeting availability of digital driving licenses by 2024:³⁷ Development programs have been underway since 2016.³⁸

The smartphone may also be used for elements of pre-authorization for travel, for example for the submission of fingerprints required for entry visas. Over the coming years, smartphones could be used to capture this biometric data in lieu of specialized machines. The United Kingdom's government has been evaluating smartphones to capture fingerprints and face data.³⁹

A migration to smartphone-based national or regional identity is unlikely in the near term but could be likely in the medium to long-term: As authentication for a widening range of high-value processes, including accessing US\$100,000 cars, US\$1,000,000 homes, and US\$10 million office buildings becomes more common, trust and familiarity in creating smartphone-based identity could increase. A large proportion of smartphone owners may be ready to add identity to their array of smartphone applications. According to Deloitte UK's Digital Consumer Trends, about a quarter of respondents in developed markets would like to use their phone as their driving license or passport.⁴⁰

Bottom line

In 2024 and over the coming decades, smartphones could replicate and exceed the functionality of tens of billions of physical authentication tools in use today, including keys, passwords, driving licenses, passports, credit cards, and cash. The smartphone's success is not limited to just unit sales: The value of its multiplier is becoming more significant.

The addition of authentication to the smartphone's utility could be analogous to its assimilation of the functionalities of multiple form factors, including compact cameras, MP3 players, alarm clocks, handheld GPS navigation, office desk phones, and tourist guidebooks.

However, authentication may be more valuable capability than playing music, snapping selfies, or setting an alarm. Smartphone-based verification of identity can accelerate, enhance, and reduce the cost of processes that are often fundamental to commerce, enterprise security, and border control.

Modern society often requires technologies such as keys, passports, and means of payment. However, these tools do not always need to be physical – they can exist as software capabilities within smartphones and can be better keys, passports, and payment tools as a result.

As society migrates to smartphone-based authentication, it will be important to help ensure that no users are left stranded: change is often challenging, depending on the individual.

Given the widening future scope of the smartphone, it is likely to cement its position as a successful device. This may dampen (but not eradicate) discussions of when it might be toppled by another form factor.

Figure 1 sources

- A. Rosie O'Connor, "[Mobile authentication market: 2023-2028](#)," Juniper Research, October 23, 2023.
- B. Adobe Experience Cloud, "[Top ecommerce statistics for 2023](#)," May 12, 2023.
- C. Chantel Wakefield, "[Cars that use digital keys in 2023](#)," Kelley Blue Book, June 14, 2023.
- D. Hedges & Company, "[How many cars are there in the world in 2023?](#)", accessed November 14, 2023.
- E. Emirates, "[Emirates goes digital, phases out paper boarding passes for flights departing Dubai](#)," May 12, 2023.
- F. ICAO, "[The world of air transport in 2019](#)," accessed November 16, 2023.
- G. Eurostat, "[Household composition statistics](#)," accessed November 16, 2023.
- H. Richard Fry, Jeffrey S. Passel And D'Vera Cohn, "[U.S. household growth over last decade was the lowest ever recorded](#)," Pew Research Center, October 12, 2021.
- I. ICAO, "[The world of air transport in 2019](#)."
- J. The World Bank, "[Labor force, total](#)," accessed November 16, 2023.
- K. Apple, "[Countries and regions that support Apple Pay](#)," August 10, 2023.
- L. UITP, "[Data: Public transport & urban mobility data](#)," accessed November 16, 2023.
- M. Sorin-Andrei Dojan, "[Mobile wallets, most popular payment method in China: GlobalData](#)," Electronic Payments International, July 6, 2023.
- N. PYMNTS, "[Mobile wallet adoption](#)," August 2022.
- O. US Census Bureau News, "[Quarterly retail e-commerce sales 2nd quarter 2023](#)," press release, August 17, 2023.

By

Paul Lee
United Kingdom

Ben Stanton
United Kingdom

Kevin Westcott
United States

Ariane Bucaille
France

Endnotes

1. Needham Mass., “*Global smartphone shipments expected to decline 1.1% in 2023 as recovery is pushed forward into 2024 amidst weak demand, according to IDC tracker*,” IDC, March 1, 2023.
2. Counterpoint, “*2023 global smartphone shipments to hit decade low as Apple inches closer to top spot*,” press release, August 17, 2023.
3. Jack Flynn, “*17 essential multi-factor authentication (MFA) statistics [2023]*,” Zippia, February 6, 2023.
4. O'Connor, “*Mobile authentication market: 2023-2028*.”
5. Thorin Klosowski, “*RIP, Passwords. Here’s what’s coming next*,” Wirecutter, January 11, 2023; Apple Support, “*Use passkeys to sign in to apps and websites on iPhone*,” accessed November 16, 2023.
6. Denise Raghetti Pilar, Antonio Jaeger, Carlos F. A. Gomes, and Lilian Milnitsky Stein, “*Passwords usage and human memory limitations: A survey across age and educational background*,” *PLoS One*. 7, no. 12 (2012).
7. Patricija Cerniauskaite, “*Are we still lazy with our passwords? The 2021 top 200 most common passwords list is here*,” NordPass, November 23, 2021.
8. Clare Stouffer, “*139 password statistics to help you stay safe in 2023*,” Norton, June 26, 2023.
9. United Nations, “*As Internet user numbers swell due to pandemic, UN Forum discusses measures to improve safety of cyberspace*,” accessed November 16, 2023.
10. SpyCloud, *Annual Identity Exposure Report 2022*, accessed November 16, 2023.
11. Valimail, “*Email fraud landscape spring 2021*,” April 16, 2021.
12. Josephine Wolff, “*Is multifactor authentication less effective than it used to be?*,” Slate, February 22, 2022.
13. Catalin Cimpanu, “*Microsoft: 99.9% of compromised accounts did not use multi-factor authentication*,” ZDNET, March 5, 2020.
14. Rubion, “*What is SMS 2FA? Text message authentication explained*,” April 20, 2022.
15. FIDO Alliance, “*National health service uses FIDO authentication for enhanced login*,” February 24, 2021.
16. FIDO Alliance, “*Apple, Google and Microsoft commit to expanded support for FIDO standard to accelerate availability of passwordless sign-ins*,” May 5, 2022.
17. Apple Support, “*Use passkeys to sign in to apps and websites on iPhone*”; Apple Support, “*Sign in to an account on your Mac with a passkey*,” accessed November 16, 2023.

18. Google Chrome Help, "[Manage passkeys in Chrome](#)," accessed November 16, 2023.
19. [Passkeys Directory](#).
20. Adobe, "[Adobe: Holiday shopping season drove a record \\$211.7 billion for e-commerce](#)," January 11, 2023.
21. US Census Bureau News, "[Quarterly retail e-commerce sales 2nd quarter 2023](#)."
22. Benedict Evans, "[Back to the trend line?](#)," July 28, 2022.
23. PYMNTS, "[Apple Pay has 48% share of mobile wallets yet only tiny sliver of total retail payments](#)," August 15, 2022.
24. GSM Arena, "[Nokia 6131 NFC](#)," accessed November 16, 2023.
25. IFSEC Insider, "[A guide to mobile access control systems](#)," August 23, 2023.
26. The World Bank, "[Labor force, total](#)."
27. Fira de Barcelona, "[Key facts and figures](#)," accessed November 16, 2023.
28. Vegas Means Business, "[Las Vegas Convention Center](#)," accessed November 16, 2023.
29. MWC Barcelona, "[Digital badge](#)," accessed November 16, 2023.
30. Wikipedia, "[List of campus identifications in mobile wallets](#)," accessed November 21, 2023.
31. Nuki, "[Say hello to the smartest Nuki door lock ever](#)," accessed November 16, 2023.
32. ICAO, "[The world of air transport in 2019](#)."
33. Rachel Chang, "[This airline is phasing out paper boarding passes](#)," Condé Nast Traveler, May 15, 2023.
34. Ukraine Now, "[Digital country](#)," accessed November 16, 2023.
35. Apple, "[Apple announces first states signed up to adopt driver's licenses and state IDs in Apple Wallet](#)," press release, September 1, 2021; Umar Shakir, "[Apple's digital state ID cards are now available for Maryland residents](#)," The Verge, May 26, 2022.
36. Potential, "[Building the future of digital identity in Europe](#)," accessed November 16, 2023.
37. RAC, "[Digital driving licences will arrive before 2024](#)," September 20, 2021.
38. BBC, "[UK developing digital driving licence](#)," May 16, 2016.

39. UK Government, “ *Biometric self-enrolment feasibility trials*,” July 4, 2022.

40. Deloitte, “*Digital Consumer Trends 2023*,” accessed November 16, 2023.

Acknowledgements

Cover image by: **Manya Kuzemchenko**

Signals from space: Direct-to-device satellite phone connectivity boosts coverage

Integrating satellite and terrestrial mobile networks could unlock new revenue for the satellite, semiconductor, and telecom industries.

ARTICLE • 7 MINUTE READ

Is completely liberating connectivity from geography possible? Providing uninterrupted coverage to mobile devices (like smartphones and IoT devices) anywhere on Earth without specialized equipment for users—that is the goal of a growing ecosystem of satellite and mobile network operators, handset manufacturers, semiconductor companies, and global regulators. Deloitte predicts that more than 200 million smartphones that can connect with satellite services will be sold in 2024. These phones are expected to contain about US\$2 billion of special chips.¹ In addition, spending on satellite construction and launches could bring the total technology investment to help enable this market to more than US\$3 billion in 2024.

To be clear, this technology does not compete with terrestrial cellular services from mobile network operators. It provides limited connectivity in areas where there is no terrestrial cellular coverage. But it cannot compete with terrestrial cellular networks for providing high speed, low-latency connectivity for the foreseeable future. Not only does it not compete, but it also allows telcos the opportunity to provide text and possibly voice service in areas where they have no current coverage via partnering, and to charge for it.

If you build it, will they come?

There is a race to build a new set of satellite-enabled global telecommunications services for consumers, enterprises, and governments. Basic services for emergency communication, simple text messages, and IoT monitoring have already started. The new emergency services have enabled rescues and saved the lives of people in car crashes and natural disasters.²

To help make this a reality, capital is being raised, satellites are being launched, chips have been developed, regulatory frameworks established, and agreements with mobile network operators are being secured. Some are very bullish on the potential size of the market (reaching tens of billions of dollars in the next 10 years), while others are expecting it to take a long time to mature and generate sufficient revenue.³ In either case, this could represent an opportunity to significantly increase the size of the global satellite communications market.

Being able to make a voice call through a satellite from a handset is not new. Iridium had its first call in 1999 and is providing service today to nearly 400,000 commercial voice and data subscribers as of the end of 2022.⁴ But those handsets are expensive, and they look and function nothing like the five billion smartphones that consumers carry. What is new is that there are space companies making and using satellites that can talk directly to existing smartphones, and there are smartphone companies that are adding relatively inexpensive chips to models that can talk directly with satellites.

Smartphone makers are hoping that adding satellite connectivity will be seen as an attractive feature—encouraging existing users to upgrade to new phones and helping makers gain share against competing phones that don't offer the capability. Smartphone shipments have languished recently, a predicted 1.15 billion in 2023, down from around 1.4 to 1.5 billion per year between 2015 and 2018, and manufacturers are eager to grow the market and gain share.⁵ Mobile network operators are looking for new revenue from more subscribers and add-on services. They're also looking to expand their geographic footprint without building expensive new ground-based infrastructure. Satellite operators are hoping to capture a growing piece of the global telecommunications market. They want to get access to a large number of new customers and fresh investment from their partners.

There are two different types of consumers who could help drive the growth of this market. The GSMA estimates that 95% of the world's population is covered by a mobile broadband network.⁶ That leaves about 400 million people who don't have access and could benefit from satellite-based mobile services. However, many in this group may not be able to afford the handsets or services needed. The much more significant market is made up of the ~5.6 billion current mobile subscribers around the world.⁷ Integrating satellite and terrestrial mobile networks on their handsets could help fill in service gaps, maximizing geographic, as well as population, coverage. If a moderate number of current subscribers are willing to pay an incremental amount in addition to their normal monthly bills, the market could be quite robust.

Although the consumer market is getting the bulk of the attention, commercial and government services could also help influence the market. Communications services for public safety and national security are natural use cases for direct-to-device (D2D). In addition, IoT applications that only require periodic connectivity and transmit limited amounts of data could see strong adoption—especially in the health care, industrial, and transportation industries. The market for satellite-based IoT services is already robust and growing—Globalstar and Iridium both reported gains in their commercial IoT subscriber base in the second quarter of 2023.⁸ More advanced direct-to-device capabilities could broaden two-way communication and accelerate the number of connected devices.

Constructing a winning team: The who's who of D2D

There are many companies trying to develop this market, each with slightly different approaches and technologies—some are using existing infrastructure, others are building their own, some are taking an open approach and others are being strictly proprietary. Some of the larger, more advanced efforts include:

- **Apple and Globalstar:** Apple has invested US\$450 million in Globalstar to upgrade their ground stations, develop a new generation of satellites, and secure 85% of their constellation's capacity in LEO to provide emergency communication services to iPhone 14 and 15 users.⁹ The "Emergency SOS via satellite" function currently enables users in 14 countries (with 740 million in total population) to share critical information with emergency services while outdoors with no wifi or cellular service.¹⁰ Additionally, Apple has said they are partnering with AAA on a "Roadside Assistance via satellite" feature for US iPhone 14 and 15 owners.¹¹ Both of these services are free for the first two years of phone ownership. It's unclear how users will be charged after the initial two-year period.
- **SpaceX and T-Mobile:** T-Mobile is working with SpaceX to use their Starlink LEO satellite constellation to provide text messaging services for customers in the United States (with plans to move to voice and data in the future).¹² Their aim is to provide full global texting coverage by the end of 2024.¹³ To enable these services, Starlink's second generation satellites will need to be used, which have more advanced antennas.¹⁴ This generation of satellites began launching in early 2023. SpaceX is also working with telecom operators Salt in Switzerland and One New Zealand to eventually provide similar services.¹⁵

- **Lynk Global:** This startup is looking to construct its own satellite constellation to provide direct-to-device services. Currently, it only has three satellites in LEO, but its goal is to launch 1,000 by 2025 (with potentially thousands more).¹⁶ Lynk touts that its system can work (in varying capacities) with any cell phone without using a specialized chip.¹⁷ They're working directly with mobile network operators in over 40 countries. Their first test case will be with the Palau National Communications Corporation.¹⁸
- **AST SpaceMobile:** This company has only just started to build out its satellite constellation. It deployed a test satellite with the largest commercial antenna ever in LEO at the end of 2022.¹⁹ AST SpaceMobile has also demonstrated its capabilities with a two-way test call.²⁰ The initial launches for its constellation are currently slated for 2024—it needs 110 satellites for sufficient global coverage.²¹ AT&T is leasing part of its wireless spectrum to AST SpaceMobile for their use.²² AST SpaceMobile is also working with other mobile network operators, including Telefónica, Vodafone, and Rakuten.²³
- **MediaTek and Bullitt:** MediaTek has developed a GEO-satellite compatible chipset (MediaTek MT6825) to be used in Android devices.²⁴ This is being used, along with the Bullitt Satellite Connect platform, in devices like ruggedized Android smartphones, in addition to a stand-alone accessory that can be linked to any Android or iOS device.²⁵ Since the service uses GEO-based satellites (from Inmarsat and EchoStar), messages take longer to send and receive, but don't require aiming the smartphone.²⁶ Plans range from \$5-25/month, depending on the number of messages used.
- **Huawei and China Telecom:** Huawei's Mate 50 series of smartphones have the capability to send texts for emergency services through China's BeiDou Navigation Satellite System.²⁷ Their newest phone, the Mate 60 Pro, has more advanced satellite-based calling and messaging features through China's three Tiantong-1 satellites in GEO.²⁸ China aspires to deploy one or more LEO satellite broadband constellations in the coming years that may expand D2D services.²⁹

Bottom line

This market is in its early stages and very dynamic, with many companies around the world seeking a stake. However, just because something is possible and there is significant activity, doesn't mean that it will be successful. There are several challenges to the development of this market.

Economic: Will the overall demand be there for the capacity that's being built? Will consumers be willing to pay for these services? If so, what's the best way for services to be monetized (for example, subscription, per use, etc.)? Can the necessary fundraising for launching, operating, and maintaining satellite constellations be sustained? Will semiconductor and handset manufacturers see a spike in demand?

Technical: In addition to the intrinsic technical challenges with deploying and maintaining large satellite constellations, such as successful launches and operating in a crowded orbital environment, there are also specific direct-to-device issues. Coordinating spectrum use and managing potential interference between operators will likely be an ongoing challenge.³⁰ Global standards should be developed for things like cybersecurity and interoperability. The 3rd Generation Partnership Project Release 17 is a start, covering standards for the role of satellites in global 5G networks.³¹ What technical approach will predominate—one where chipsets in smartphones power satellite communication or where satellites act more as space-based cell towers enabled by network-on-the-edge architecture? In either case, advancement in both satellite and smartphone technology will likely be necessary to enable the full potential of D2D.

Regulatory: There are currently few specific regulations for the direct-to-device market. Just like other initiatives, satellite, and mobile network operators should work through the appropriate agencies in the countries they are operating in for approvals—for things like spectrum allocation, orbital slots, and landing rights (permission for satellite services in a country). In the United States, the FCC has worked to enable satellite and mobile network operators to collaborate easier. They recently proposed a regulatory framework to allow satellite operators to apply to access spectrum controlled by mobile network operators.³²

Blending these different markets together to make connectivity like air is ambitious and potentially lucrative. There are many benefits—connecting the unconnected, improving safety and emergency response, expanding IoT applications, and increasing the resiliency and reliability of the global telecommunications infrastructure. The question is, how fast could a fully realized, integrated satellite and terrestrial network contribute to new innovations and behaviors that can be widely adopted and monetized?

By

David Jarvis
United States

Duncan Stewart
Canada

Raghavan Alevor
United States

Kevin Westcott
United States

Endnotes

1. Deloitte analysis.
2. Andrew Orr, "[Life-saving Apple emergency SOS summons help for 10 stranded hikers](#)," Appleinsider, May 17, 2023; Corinne Reichert, "[iPhone 14 emergency SOS facilitates car cliff crash rescue](#)," CNET, July 24, 2023; Florence Ion, "[Family escapes Maui fires using Apple's emergency SOS](#)," GIZMODO, August 10, 2023.
3. David Hodes, "[Fine-tuning the work needed to adapt commercial 5G to military communications](#)," ViaSatellite, March 14, 2023; Jason Rainbow, "[The promise of direct-to-device](#)," SpaceNews, July 14, 2023; Jason Rainbow, "[Satellite executives spar over direct-to-device evolution](#)," SpaceNews, March 13, 2023.
4. Rachel Jewett, "[Iridium posts double-digit revenue increase in 2022, surpasses 2M subscribers](#)," ViaSatellite, February 16, 2023.
5. Arjun Kharpal, "[Global smartphone market to hit decade low in 2023 but Apple could take top spot, research shows](#)," CNBC, August 17, 2023; Dan Gallagher, "[How much is too much for a smartphone?](#)," *Wall Street Journal*, September 7, 2023.
6. Anne Delaporte and Kalvin Bahia, *The state of mobile internet connectivity 2022*, GSMA, accessed November 2, 2023.
7. GSMA, "[Data](#)," accessed November 2, 2023.
8. IoT.Business.News, "[The global satellite IoT subscriber base to reach 21.2 million by 2026](#)," September 01, 2022; Globalstar, "[Globalstar announces second quarter 2023 results](#)," press release, August 3, 2023; Iridium, "[Iridium announces second-quarter 2023 results; reports record operational EBITDA](#)," press release, July 25, 2023.
9. Apple, "[Emergency SOS via satellite on iPhone 14 and iPhone 14 Pro lineups made possible by \\$450 million Apple investment in US infrastructure](#)," November 10, 2022; Jason Rainbow, "[Apple lends Globalstar \\$252 million for satellite-enabled iPhones](#)," SpaceNews, February 28, 2023.
10. Apple Support, "[Use emergency SOS via satellite on your iPhone](#)," accessed November 2, 2023.
11. Jennifer Pattison Tuohy, "[Apple is adding roadside assistance via satellite to the iPhone](#)," Verge, September 13, 2023.
12. T-Mobile, "[T-Mobile takes coverage above and beyond with SpaceX](#)," press release, August 25, 2022.
13. Starlink, "[Starlink direct to call](#)," accessed November 2, 2023.
14. William M. Wiltshire, Paul Caritj, David Goldman, and Jameson Dempsey, *Consolidated opposition to petitions and response to comment of space exploration holdings, LLC*, Federal Communications Commission, May 30, 2023.

15. Lucy Craymer, "[One New Zealand to use SpaceX satellites to improve coverage](#)," Reuters, April 3, 2023; Jason Rainbow, "[FCC unveils proposed satellite direct-to-device regulatory framework](#)," SpaceNews, March 17, 2023.
16. Jason Rainbow, "[Lynk Global on verge of initial commercial direct-to-device services](#)," SpaceNews, June 23, 2023.
17. Lynk, "[What we do: Connecting everyone, everywhere](#)," accessed November 2, 2023.
18. Lynk Global, "[Lynk and Palau National Communications Corporation \(PNCC\) begin initial Sat2Phone service to pncc mobile subscribers](#)," June 21, 2023.
19. AST SpaceMobile, "[BlueWalker 3](#)," accessed November 2, 2023.
20. Jason Rainbow, "[AST SpaceMobile conducts first direct-to-device voice test](#)," SpaceNews, April 25, 2023.
21. Jason Rainbow, "[Operational AST SpaceMobile satellites could proceed without prototype](#)," SpaceNews, May 17, 2023; Jason Rainbow, "[AST SpaceMobile discloses further satellite delays and cost increases](#)," SpaceNews, March 31, 2023.
22. Jason Rainbow, "[AT&T and AST SpaceMobile seek permission for spectrum leasing deal](#)," SpaceNews, May 11, 2023.
23. AST SpaceMobile, "[Strategic partners](#)," accessed November 2, 2023.
24. Mediatek, "[5G satellite \(NTN\): Leading the industry in 5G satellite connectivity](#)," accessed November 2, 2023.
25. David Lumb, "[Bullitt to launch its iPhone-rivaling satellite smartphone in March](#)," CNET, February 24, 2023.
26. Jason Rainbow, "[Inmarsat and MediaTek expand direct-to-device partnership](#)," SpaceNews, April 12, 2023; Umar Shakir, "[This bluetooth fob turns iPhones or android devices into two-way satellite messengers](#)," Verge, February 25, 2023.
27. Allison Johnson, "[The first phone maker to add satellite texting to its devices is... Huawei](#)," Verge, September 7, 2023.
28. Huawei, "[Huawei Mate60 Pro](#)," accessed November 2, 2023.
29. *Wall Street Journal*, "[China plans to build a satellite network to rival Elon Musk's Starlink](#)," July 18, 2023; Andrew Jones, "[China to begin constructing its own megaconstellation later this year](#)," SpaceNews, March 28, 2023.
30. Jon Brodtkin, "[AT&T tries to block Starlink/T-Mobile plan for satellite-to-phone service](#)," ArsTechnica, May 20, 2023.

31. Munira Jaffar and Nicolas Chuberre, “*NTN & satellite in Rel-17 & 18*,” 3GPP, July 1, 2022.
32. Federal Communications Commission, “*FCC proposes framework to facilitate supplemental coverage from space*,” March 17, 2023.

Acknowledgements

The authors would like to sincerely thank **Karthik Ramachandran** for his contributions to this article. They would also like to thank **Luke Sienkiewicz** and **Adam Routh** for sharing their expertise on the topic.

Cover image by: **Manya Kuzemchenko**

No bump to bitrates for digital apps in the near term: Is a period of enough fixed broadband connectivity approaching?

In some parts of the world, some consumers may have all the bandwidth they need in 2024. They are still expected to want ever better internet connections, just not necessarily faster ones.

ARTICLE • 9 MINUTE READ

The demand for ever higher internet speeds over fixed networks has been a constant for most of the history of the internet. However, recent shifts in technology and consumer behavior suggest that this trend may be slowing or stopping.¹ The rise of online video streaming has, historically, significantly driven bandwidth demand, but even this shift has its limits. Deloitte predicts that in 2024, over 90% of the most used online applications on fixed broadband networks in developed markets will have the same vendor-recommended bitrate as in 2023. The most commonly used applications include streaming video and audio, video calling applications, online games and home security: Most of these applications state recommended minimum bandwidth on their websites. We further predict that at least 80% of these applications may see no increase to officially advised bitrate for the two-year period between 2023 and 2025. Over time, the bitrate for some applications might reduce due to better compression.²

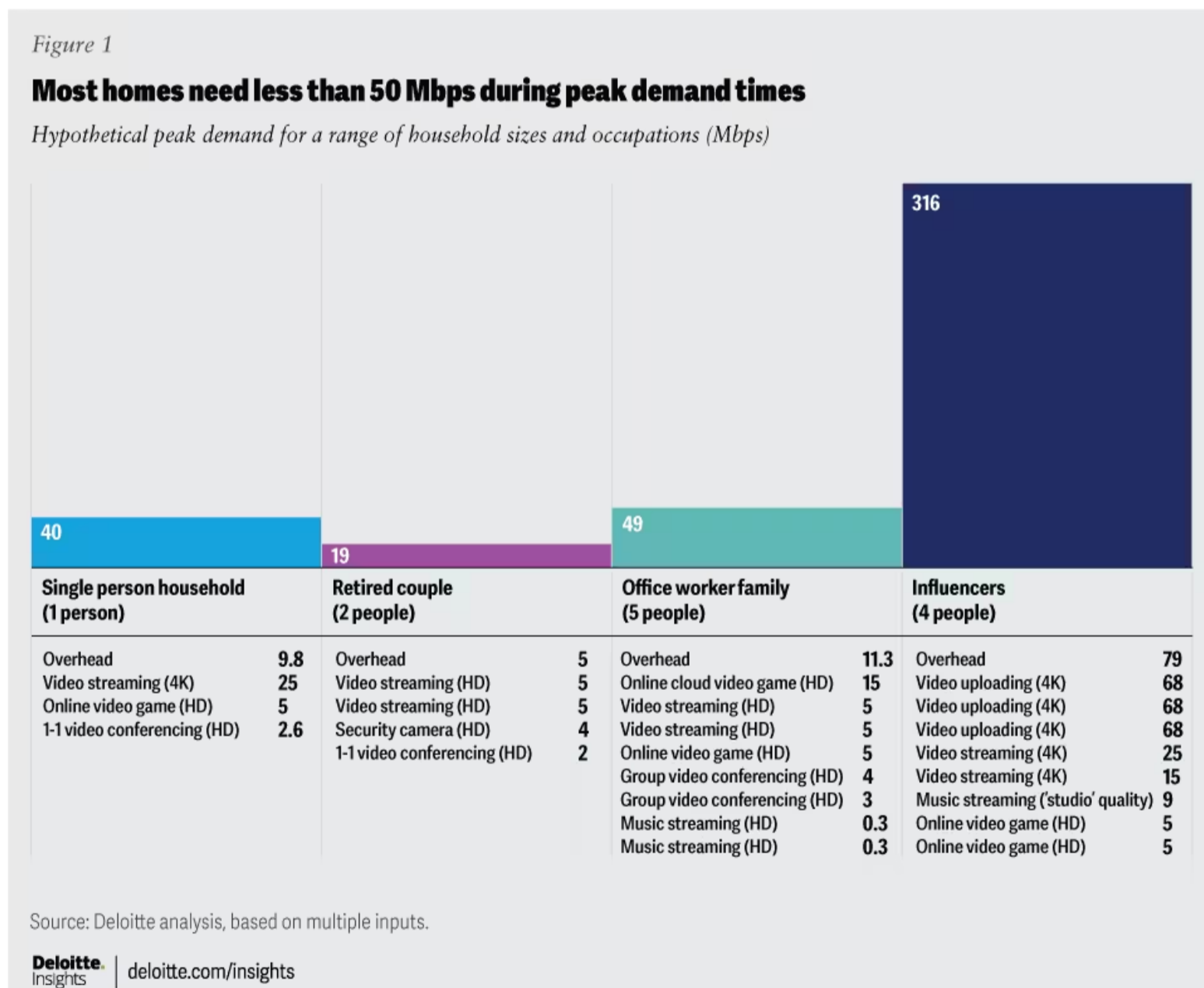
For many households, aggregate peak-hour fixed broadband requirements may increase only moderately in the near-term future. As such, there may be less incentive for subscribers to upgrade proactively to a significantly higher speed package.³ Consumers are likely to be less focused on speed and more on other capabilities, such as reliability, bundles with media content, indoor range, or outright value.

That said, consumer demand for higher speeds is only part of what drives network infrastructure needs for broadband providers. Globally, there are governmental incentive programs and regulatory requirements aimed at eliminating the digital divide with minimum speed requirements. These range from as little as 10 megabits per second (Mbps) to 1 gigabit per second download speeds, so providers should build networks with these speeds to receive funds or comply with mandates.⁴ Next, there are competitive pressures in the market: If one provider is advertising ultrafast speeds, others may need to match or at least come close. Further, building network infrastructure is often an investment with a 20 year plus horizon, so there is an element of futureproofing. Finally, newer technologies can offer more than just speed; they are likely more reliable, more sustainable, easier to provision, cheaper to operate, and have lower latency. Putting that all together, broadband providers are expected to migrate away from copper digital subscriber line technologies and older cable solutions, toward Data Over Cable Service Interface Specification 4.0 and fiber to the home, even though consumers may not yet “need” those technologies in 2024 and 2025.

When considering bandwidth demand per household, trends in the composition of homes should be noted. In developed markets, household size is generally in long-term decline. In the European Union, as of 2022, average household size was 2.1 people; over three quarters of households had no children. While a home with three or more kids and two parents may have high aggregate bandwidth demands, less than 5% of households have three children or more;⁵ single or double adult occupancy predominates.⁶ Indeed, by 2025, single adult households may be the most common type.⁷ The United States shows similar trends. In 2022, average size for family households (a third of the total) was 2.5.⁸ Of the remainder, average size was 1.25, with 37.9 million households, that is 29% of the total, being single-person occupied.⁹

For a household with 2 people, aggregate peak-time usage for the most used applications is likely to be well under 100Mbps in 2024,¹⁰ even allowing for overhead. A reliable 100Mbps connection would likely be more than sufficient to deliver two high-definition (HD) video streams to a television. If someone was also listening to streamed music, this would—per the websites for some of the most popular applications—add under 1Mbps per “high quality” stream.¹¹ For online multiplayer PC or console video games, 5Mbps–10Mbps per session is the most likely recommended bitrate.¹² Video calls require up to 4Mbps downstream and upstream.¹³ There may be other applications, such as security cameras, that would use additional capacity, but these devices typically use 1–3Mbps per device.¹⁴

A household with four influencers—an atypical example—could require over 300Mbps, if three were uploading 4K video at the same time while downloading other content (figure 1). However, less than 1% of households in most markets would likely have this composition of individuals.¹⁵ The more common composition—adults with no children—will very likely have far more modest needs, in 2024 and in the medium term.



Video on demand’s sweet spot is likely to remain 5Mbps

The most ubiquitous data consumptive application in most developed market households in 2024 is likely to be video-on-demand (VOD). In most developed markets, about three quarters of households have access to VOD.¹⁶ In many developed markets, consumption of all forms of on-demand video is likely to be multiple hours per day, and in many homes, VOD is likely to drive the majority of household network traffic.¹⁷

The most common recommended bitrate for VOD in 2024 will be 5Mbps or lower, which will be for HD viewing.¹⁸ Deloitte predicts about a fifth of VOD will be streamed in 4K with the recommended bitrate varying between 15Mbps and 25Mbps.¹⁹ Sports streamed in 4K, which if not readily compressed, will have recommended bitrates starting from 30Mbps.²⁰ Video watched on smaller screens, such as smartphones, PCs, or tablets would likely require a lower bitrate at each resolution level.

Over the last 12 years, required bitrates for streaming VOD have tended to decline as compression has improved.²¹ Over the next five years, further advances in compression are likely.

VOD platforms are unlikely to increase recommended bitrates in 2024. Increasing the bitrate to offer a higher-quality image may only make sense if consumers can perceive this improvement. However, many viewers might remain satisfied with HD streams in 2024, and over the medium term. One illustration of the public's acceptance of current streamed HD and 4K quality is the contrasting bitrate for optical discs, which were subscription VOD's predecessor. The bitrate for HD disks is up to 40Mbps and for 4K it's up to 128Mbps.²² This means that the bitrate delivered from an optical HD disk may be double that of a 4K stream, in other words a Blu-ray HD may deliver higher quality as measured by bitrate than a 4K one.

Deloitte does not expect a marked shift from HD to 4K viewing. There are at least three constraints. One is biological: Many viewers may simply not be able to differentiate and so are unlikely to pay a premium for 4K.²³ A second is historical: Some popular library content from the 1990s and early 2000s is only in HD. A further constraint is that most TVs in developed markets are HD. In 2019, HD sales were 52% of all sets sold globally,²⁴ and it may take many years for HD sets to be replaced by 4K.²⁵ In the United States market, it was only in 2021 that most homes (51%) self-reported owning at least one 4K set.²⁶

Average actual VOD bitrates may even decline for some households. VOD platforms are constantly optimizing content encoding. Netflix has reduced the bitrate for 4K to as low as 1.8Mbps for segments of some content.²⁷ Furthermore, some households may downsize from 4K service to HD to trim cost. As from November 2023, 4K content on Disney+ will only be available on a premium tier in some markets.²⁸ Finally, any crackdowns on sharing, which may ramp up in 2024, may result in less streamed content being watched in some homes, reducing bandwidth demand.

This section has focused on streaming to TVs. Arguably streaming to virtual reality or augmented reality glasses could require much higher bitrates, as could holography. But we are not predicting a notable rollout of such capabilities in 2024 or 2025. The active installed base of televisions in 2024 is likely to be in the billions, which is higher than the expected installed base of virtual reality devices (in the low tens of millions) or augmented reality headsets (in the hundreds of thousands), meaning that this is unlikely to require higher broadband speeds.²⁹

Audio: Most applications are under 1Mbps

Streamed music continues to rise in popularity. For 2024's sixteen-year olds, subscription music services will have been available for the entirety of their lives.³⁰ Deloitte predicts that in 2024, the median streaming speed for music will be about 300 kilobits per second (Kbps) for a service that is marketed as high quality.³¹

There will be significantly higher fidelity services available with much higher bitrates of up to 10Mbps, for Tidal's "Master" service.³² But these may represent only a small proportion of subscribers and usage, partly because of the higher prices associated with audiophile streaming, and also because Bluetooth, which is the most common form of connecting wireless headphones, does not support bitrates above 1Mbps.³³

Speech services such as podcasts and radio are also likely to be under 0.5Mbps.³⁴ Podcasts are about 100Kbps–150 kbit/s.³⁵

Console and gaming PC video games: Most usage will be under 10Mbps

The most data-consumptive games in 2024 are likely to be those played on consoles and PCs. The most popular applications are likely to require in the region of 5Mbps to 10Mbps for online play, with the game content having been downloaded previously.³⁶ Cloud gaming requires up to 45Mbps but is niche, with under 5% of households likely to use such services at least monthly in 2024.³⁷

Connectivity is required for occasional file downloads, but this can be done in the background or during quiet periods. For online game players, downstream and upstream speeds are just two of many parameters that matter; another critical factor could be the ping rate, which measures network latency.

Usage is expected to be significant but much smaller relative to mobile games, which are mostly played on smartphones. The active installed base of devices capable of supporting the latest generation of PC and console games is in the hundreds of millions; this compares to five billion smartphones.³⁸ At the start of 2022, the latest generation of Sony's and Microsoft's console had sold about 50 million units,³⁹ and in July 2023, Sony announced it had sold 40 million units.⁴⁰ The installed base of gaming PCs is expected to be in the low hundreds of millions in 2024, and annual sales were sized at 45 million in 2022.⁴¹ About a third of all consoles and gaming PCs are expected to be used daily in 2024.⁴² By contrast, more than 90% of smartphones and 80% of smart TVs will likely be used daily.⁴³

In the future, cloud gaming—whereby the game is hosted remotely with video images streamed to any screen—could become popular. Cloud gaming services stream video of the game from a data center to the player's screen, enabling high quality games to be played without a gaming PC or console. An HD quality game may need between 6Mbps and 25Mbps,⁴⁴ and a 4K game may need up to 45Mbps.⁴⁵ The exact bitrate will depend on frame rate (measured in frames per second) and color range.

Video calling's sweet spot may be under 5Mbps

Another common application in people's homes is likely to be video calling. About 60% of populations in developed markets are in the workforce, and of these, about half may need to do video calls occasionally or more frequently from home.⁴⁶

For HD connections, most services require 4Mbps for group video calls and less for one-to-one calls.⁴⁷ Standard-definition video calls require speed between 1Mbps and 2Mbps. Video calls can take place at any time of the day, but work-related calls are likely to take place during the daytime, and depending on the size and age of the household and the season, calls may not compete with other applications for bandwidth. For those that prefer to have audio only calls, bandwidth demand is trivial at under 0.1Mbps.⁴⁸

For some types of calls, the uplink speed may be challenging. Broadband connections are typically weighted to the downstream connection, that is, most of the bandwidth is allocated for the downlink.⁴⁹ For most homes, and for many years, this has often been a rational apportionment, with the uplink rarely being used. However, for video calls, a need for 5Mbps uplink can be problematic for some connections during busy periods in a day.

For a small number of users, applications may be streamed to the user's device, possibly for security reasons or because they are using a thin client (a simple computer that derives its primary processing power over a network connection from a remote server) or network-based device. For the most used office applications, the maximum bandwidth demand would be about 2Mbps.⁵⁰

Smart home, under 5Mbps per device

Better connectivity has enabled networked security cameras to become relatively affordable. Economies of scale—and the sheer volume of cameras manufactured for devices every year—have helped the price of security devices to decline. Smart cameras remain relatively niche. Deloitte's surveys indicate that about one in five homes have at least one device, and footage may be stored locally rather than uploaded.⁵¹

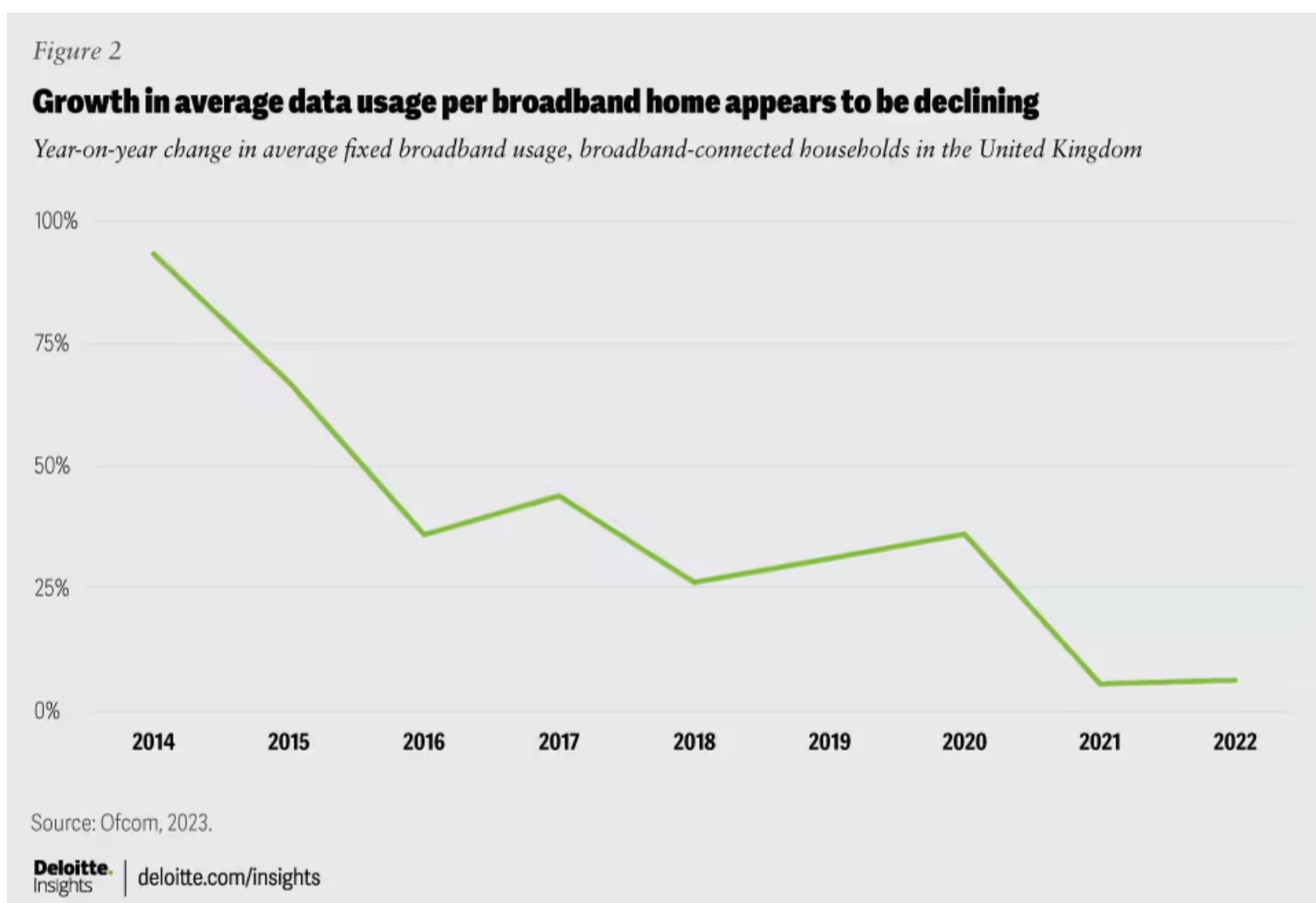
Connectivity per HD device is up to 4Mbps upstream,⁵² but 2Mbps is a common recommended speed.⁵³ For some models, bitrates increase when motion is detected.

The bottom line

In the 2010s, the shift from broadcast video, digital video recorders, and DVDs to online video delivered a multiyear change in aggregate bandwidth demand. Historical growth in bandwidth demand, however, may no longer be a reliable indicator of future demand. The era of enough connectivity may be here.

Shifts tend to be finite. Some can take decades to complete; others may take years. In recent years, adoption of VOD has slowed and in the absence of any new applications or devices that could require significantly more bandwidth, household demand for bandwidth might be reaching its natural ceiling, at about 100Mbps.

A metric that may signal a steady decline in the need for higher broadband speeds is the growth in gigabytes transferred per month (figure 2).



The implications are likely multiple.

First, in some geographies, telcos may benefit from focusing less on outright speed and more on other attributes, including reliability. There are also considerations on the positioning of speed tiers: Offering 200Mbps as an entry level tier may encourage the majority of an operator's subscribers to opt for the lowest and cheapest tier. Conversely, starting at 50Mbps may nudge subscribers into opting for a mid-range tier.

Second, given the variation in uptime between different technologies, the case for migrating to fiber to the home may remain strong relative to legacy technologies with copper terminations.

Third, if aggregate required bitrates are slowing down, this may also mean that other network technologies including 5G or 4G fixed wireless access (FWA) may be viable. In the US market, as of 2022, the fifth fastest network was an FWA network, with median speeds of 24Mbps.⁵⁴

The era of rapidly increasing bandwidth demand may be coming to an end in some markets (or at least pausing), as shifts in technology and consumer behavior suggest a near-term plateau. Online video streaming, historically a major driver of bandwidth demand, is becoming less demanding. If consumers no longer need faster speeds, it's likely a matter of time before this ceases being a meaningful differentiator.

By **Paul Lee**
United Kingdom

Eytan Hallside
United Kingdom

Dieter Trimmel
Germany

Endnotes

1. Seventy-six percent of UK consumers are satisfied with their internet connection. Read Deloitte, *Digital Consumer Trends 2023*, August 2023.
2. Historically, new compression algorithms have been introduced every ten years or so, namely MPEG-2 in 1995, AVC in 2005, and HEVC in 2015. The next generation of compression algorithms are likely to be enabled via AI, which should enable better video quality at a given bitrate. Read Jean-Louis Diascorn, “*How AI technology is dramatically improving video compression for broadcast and OTT content delivery*,” SMPTE, October 21–24, 2019.
3. We note that some households may be opted into a higher speed package for other reasons, such as performance bumps included as part of a retention package from the same operator. Further, some subscribers may decide to switch to a fibre to the home (FTTH) primarily for lower tariffs, higher reliability, or lower latency. For more information on performance enhancements from FTTH, refer to FTTH Council Europe, “*Everything you need to know about FTTH*,” accessed November 2023. Also, a small proportion of customers may always opt for the highest tier available, regardless of actual need.
4. Local Government Association, “*Tackling the digital divide—House of Commons, 4 November 2021*,” November 3, 2021.
5. Eurostat Statistics Explained, “*Household composition statistics*,” August 9, 2023.
6. Ibid.
7. Ibid.
8. United States Census Bureau, “*Table A1 of America’s families and living arrangements: 2022*,” November 17, 2022.
9. In 1960, only 13% of homes were occupied by a single individual (United States Census Bureau, “*Census bureau releases new estimates on America’s families and living arrangements*,” press release, November 17, 2022).
10. In Europe and North America, the average size of a household is under three people (United Nations Department of Economic and Social Affairs, “*Population facts: Household size and composition around the world*,” October 2017).
11. This is based on the bitrates for the most commonly used services, such as Amazon Prime Music (Amazon United Kingdom, “*Amazon Music Unlimited in HD FAQ*,” accessed November 2023).
12. Harry Guinness, “*Four ways fiber helps you win in “Fortnite”*,” Frontier Communications Parent, Inc., May 16, 2023.
13. For example, Zoom requires up to 4Mbps downstream, when in gallery mode with up to 50 participants (Zoom support, “*Zoom system requirements: Windows, macOS, Linux*,” September 11, 2023).

14. Google Nest help, “[Internet bandwidth and speed requirements for Nest cameras](#),” accessed November 2023; Blink, “[Camera video quality](#),” accessed November 2023.
15. According to one study, about 1.6% of people are influencers with at least 1,000 followers in Europe; there are relatively few households with four people in the European Union. There are relatively few households composed of unrelated people, which could include influencers who were not part of a family. In the United Kingdom, there are under 10% of households in 2022 that included two or more unrelated adults. See Influency, “[The largest influencer study in Europe in 2023](#),” accessed November 2023; Office for National Statistics, “[Families and households in the UK: 2022](#),” statistical bulletin, May 18, 2023.
16. The most common types of video-on-demand (VOD) are video sharing platforms, subscription video-on-demand (SVOD), and broadcaster video-on-demand (BVOD). As an example of a developed market, 73% of UK consumers have access to SVOD (see Deloitte, [Digital consumer trends 2023](#), August 2023).
17. UK citizens consumed 39 minutes of SVOD/advertising VOD per day, 50 minutes of video-sharing per day, and 157 minutes of broadcaster content (including BVOD) per day in October 2023 (Barb Audiences Ltd, “[Total identified viewing summary](#),” October 2023).
18. Netflix, “[Internet connection speed recommendations](#),” accessed November 2023; Prime Video, “[Prime Video support](#),” accessed November 2023; Disney Plus, “[Disney+ Hotstar helpdesk](#),” accessed November 2023.
19. Under one-fifth of Netflix’s subscribers were on the premium tier (which is required for viewing in 4K) between 2017 and 2019 (Tom Harrington, Abi Watson, and James Barford, [Netflix: Looking towards 2025](#), Enders Analysis, June 3, 2020); as of November 1, 2023, 4K was only available on the premium tier for Disney+ Hotstar (Disney+ Hotstar, “[Homepage](#),” accessed November 2023).
20. However, in most markets, most hours of live sport watched in 2024 is likely to be watched via broadcast. Over time more content will likely be streamed, Sky Sports in the UK recommends “at least 25Mbps to stream in UHD (4K) (Sky Sports, “[Sky Sport—How to enjoy UHD?](#),” accessed November 2023).
21. Ioannis Katsavounidis, “[Dynamic optimizer—a perceptual video encoding optimization framework](#),” Netflix technology blog, Medium, March 5, 2018; Netflix technology blog, “[Per-title encode optimization](#),” Medium, December 14, 2015; also see Netflix ISP index where in the United Kingdom, for example, most ISPs were clocked at 3.2Mbps in September 2023 (Netflix, “[ISP speed index United Kingdom leaderboard](#),” accessed November 2023).
22. Kavon Zamanian, “[Everything you ever wanted to know about bitrates](#),” Premium Beat, July 5, 2016.
23. UHD Alliance, “[Can the human eye see the difference between 1080p and 4K?](#),” December 17, 2018.
24. Growth from Knowledge, “[4K is becoming standard for today’s TVs](#),” September 4, 2019.
25. In 2021, 8K TV sets were 0.15 percent of all TVs sold (Deloitte, “[TV’s New Year’s resolution: The start of the 8K wave](#),” December 6, 2020); Tom Bailey, “[8K TVs still struggling to sell](#),” What Hi-Fi, April 29, 2022.

26. Steve Koenig, “*Consumer technology ownership hits new highs*,” Consumer Technology Association, July 23, 2021.
27. Aditya Mavlankar, Liwei Guo, Anush Moorthy, and Anne Aaron, “*Optimized shot-based encodes for 4K: Now streaming!*,” Netflix technology blog, Medium, August 29, 2020.
28. Tom Hedley, “*You’ll soon have to pay extra to watch Disney+ in 4K*,” Disney Plus Informer, August 10, 2023.
29. Statista, “*Number of TV households worldwide from 2010 to 2027 (in billions)*,” October 2021; IDC, “*Slower growth for AR/VR headset shipments in 2023 but strong growth forecast through 2027, according to IDC*,” press release, March 21, 2023.
30. Spotify, “*About Spotify*,” accessed November 2023.
31. Spotify, “*Audio quality*,” accessed November 2023.
32. Tidal, “*Hear more from your music*,” accessed November 2023.
33. Audiophile Heaven, “*Bluetooth sound quality guide—what are the codecs, and how do they work*,” accessed November 2023.
34. BBC’s streamed radio services, including podcasts, are delivered at up to 320Kbps, and as low as 48Kbps (BBC Sounds, “*What are the codecs, bitrates, and protocols used for BBC radio online?*,” accessed November 2023).
35. Spotify, “*Audio quality*.”
36. Fortnite can be played with a “3Mb line” as a minimum. Read: Luke Edwards, “*What broadband speed do I need for Fortnite?*,” Tech Radar, May 22, 2020.
37. Recommended bitrate for Google Stadia is up to 35Mbps (Jarred Walton, “*Stadia 4K streaming will use up 1TB of data in 65 hours*,” PC Gamer, June 6, 2019).
38. Oberlo, “*How many people have smartphones in 2023?*,” accessed November 2023.
39. Piers Harding-Rolls, “*Console market 2022 review: Hampered by lack of hardware availability*,” Ampere Analysis, February 28, 2023.
40. Jim Ryan, “*PlayStation 5 surpasses 40 million in sales*,” Sony Interactive Entertainment, July 27, 2023.
41. IDC, “*Growth in gaming PCs and gaming monitors to lag in 2023, according to IDC*,” press release, March 21, 2023.
42. *Digital Consumer Trends* research in 17 markets in Europe, Asia, and Latin America, May–September, 2023. For example, see: Deloitte, *Digital consumer trends 2023*.

43. Ibid.
44. Blacknut Cloud Gaming, “[Network connection configuration and optimization](#),” accessed November 2023; Sony Interactive Entertainment, “[Getting started with PlayStation Plus](#),” accessed November 2023; Nvidia, “[GeForce Now system requirements](#),” accessed November 2023.
45. Nvidia, “[GeForce Now system requirements](#),” accessed November 2023.
46. Pallavi Rao, “[Visualizing the American workforce as 100 people](#),” Visual Capitalist, May 15, 2023.
47. Zoom support, “[Zoom system requirements](#)”; Google Workspace admin help, “[Google Meet hardware requirements](#),” accessed November 2023.
48. Zoom support, “[Zoom system requirements](#).”
49. Andrew Braun, “[Why are upload speeds so much slower than download speeds?](#),” Make Tech Easier, August 15, 2018.
50. Contributors, “[Remote desktop protocol bandwidth requirements](#),” Microsoft, May 26, 2022.
51. Deloitte, [Digital consumer trends 2023](#).
52. Google Nest help, “[Internet bandwidth and speed requirements for Nest cameras](#).”
53. Ring LLC, “[Support center—general information](#),” accessed November, 2023; Blink, “[Camera video quality](#).”
54. Mike Dano, “[Here’s what’s next for T-Mobile’s fixed wireless access business](#),” Light Reading, September 22, 2022.

Acknowledgements

The authors would like to thank **Jan-Piet Nelissen, Matt Roberts, Vincent Fosty, Charlotte Perfect, George Budden, Alexander Mogg, Sam Breakwell, Duncan Stewart, Paul Franek, Ben Stanton, Eliza Pearce, and Peter Corbett.**

Cover image by: **Manya Kuzemchenko**

Keeping it local: Cloud sovereignty a major focus of the future

More data, increased cybersecurity threats, and geopolitical tensions are expected to increase demand for cloud solutions that can operate locally. Meeting this need can protect a company's reputation, operations, and bottom line.

ARTICLE • 12 MINUTE READ

In 2024, the world is forecast to generate 149 zettabytes of data.¹ Cloud makes this scale of data creation and processing possible. A zettabyte, for reference, is a thousand, thousand, thousand, thousand gigabytes. If each byte were a grain of sand, there would be enough to fill every beach on earth almost 20,000 times.²

Cloud is now fundamental and a major industry, forecast at almost US\$600 billion in 2023.³ It's an engine for transformative change and has contributed to improved service delivery, helped enable workforce mobility, and created new frontiers in analytics and artificial intelligence. Cloud computing was founded on the idea that data location doesn't matter—at least, not as much as economies of scale, or the agility to quickly spin up computing resources on the fly. As the volume, value, and sensitivity of data stored has surged, “cloud sovereignty,” the principal that data stored in the cloud should be subject to the laws of the country in which it physically resides, has become a major focus for policymakers.

In 2024, Deloitte predicts that the national focus on cloud sovereignty will intensify in all developed markets. Therefore, government cloud, a subset of solutions designed to meet the stringent compliance requirements of government agencies, is forecast to surpass US\$41 billion in 2024, up 16% over 2023.⁴ And distributed cloud, which can be a solution to comply with data residency restrictions, is forecast to grow to a US\$7 billion market, up from a little over US\$4 billion in 2022.⁵

A brief history of cloud computing

The concept of cloud computing dates to the early days of the internet, with time-sharing in the 1960s⁶ through to telecom virtual private networks (VPNs) in the 1990s.⁷ But the mid-2000s marked a paradigm shift with the launch of Amazon Web Services (AWS),⁸ which brought scalable, on-demand compute to the masses. Google, Microsoft, and others soon followed suit, launching their cloud platforms to cement cloud's foundational role in modern digital infrastructure.

In the past two decades, companies, governments, institutions, and citizens have gradually moved their data and workloads from private infrastructure (think server racks in an office cupboard) to monolithic cloud data centers. In that time, many also adopted a new breed of “cloud native” applications. As vast quantities of data were stored and transferred across global networks, some governments and companies became concerned about its jurisdiction, governance, and ownership, and the term “cloud sovereignty” started gaining traction.

Legal contradictions cause international frictions

Data localization laws may invariably create operational complexity for global businesses. Regulation tends to change frequently, as countries wrestle with national security, data protection, and new technologies. Companies are expected to be increasingly reliant on data, automation, and AI. They should consider how regulation might change, be invalidated, or added to. Compliance today is essential, as is operational agility. Being able to quickly adapt to changes in the regulatory framework is crucial, as hundreds of countries develop their regulatory positions, each with their own nuances, some of which might be inconsistent with one another.

So far, several attempts at regulating transatlantic data flow between the European Union (EU) and United States have been invalidated by the European Court of Justice (figure 1). In the last decade, there has also been a flurry of sovereignty-linked regulation, such as EU GDPR (General Data Protection Regulation), the US CLOUD Act (Clarifying Legal Overseas Use of Data), and some state-level initiatives like the California Privacy Rights Act (CPRA).

Global businesses, as mentioned, need to navigate variation across localized laws. The CLOUD Act, for example, empowers US authorities to access data stored abroad for select law enforcement purposes, but under GDPR, personal data can only be transferred outside the European Economic Area if given an “adequate” level of data protection.⁹ If a company releases data in accordance with the CLOUD Act but violates GDPR, it could face hefty fines.¹⁰ In some cases, governments can reach bilateral agreements such as one between the United Kingdom and United States,¹¹ but these can take time. Instead, companies often resort to end-to-end encryption, where only the sender and receiver can decrypt the data. This means that while service providers might hand over data to law enforcement, the data would be unintelligible without the decryption keys.

Figure 1

Notable developments in EU and US sovereignty regulation: An evolving policy landscape

Between the European Union and United States alone, there have been several notable milestones in the regulation of transatlantic data flows:

2000	Safe Harbor Agreement	A framework ensuring that US companies could transfer personal data from the European Union to the United States while maintaining EU data protection standards. ^A
2013	Snowden Revelations	Revelations of extensive unveiled surveillance by the US National Security Agency ignited global concerns over data privacy and governmental surveillance and reignited discussions on data sovereignty.
2015	Safe Harbor Invalidated (a.k.a. Schrems I)	The European Court of Justice invalidated the Safe Harbor Agreement due to concerns that US surveillance laws undermined European privacy rights. ^B
2016	EU-US Privacy Shield	The European Union and United States negotiated a new framework, known as the EU-US Privacy Shield, to address the shortcomings of the Safe Harbor Agreement and provide a new mechanism for transatlantic data transfers. ^C
2016	EU GDPR	The European Union introduced the General Data Protection Regulation (GDPR), which placed stringent regulations on data transfer outside the European Union, mandating companies to ensure EU personal data is processed according to appropriate measures of protection. ^D
2018	US CLOUD Act	The US introduced the Clarifying Lawful Overseas Use of Data Act (CLOUD Act), allowing US law enforcement agencies to request personal data stored by tech companies for select law enforcement purposes, irrespective of where this data is located. ^E
2020	EU-US Privacy Shield Invalidated (a.k.a. Schrems II)	The European Court of Justice invalidated the EU-US Privacy Shield, again due to concerns about US surveillance laws and the insufficiency of US protections for European citizens' data. ^F
2023	EU-US Data Privacy Framework	US and European leaders announced a framework to replace the EU-US Privacy Shield, including a new redress mechanism and the Data Protection Review Court, a kind of tribunal for transatlantic data flow. ^G

Source: Deloitte, 2023, based on publicly available third-party sources including: International Association of Privacy Professionals, Court Justice of the European Union, European Commission, European Union's General Data Protection Regulation, United States' CLOUD Act, and the European Parliamentary Research Service. *See "Figure 1 endnotes" for more detail.

Deloitte Insights | deloitte.com/insights

To navigate the complex regulatory landscape, data management is crucial. Companies should understand the variety of data types they have, and which classification of data sits in each system (for example, personal information, payments data, regulated financial information). Companies should also consider their "exit strategy" from any current cloud provider, in case of future sovereignty infringements or a changing regulatory landscape. However, the length of cloud contracts, which can be five or more years, and high egress charges (the price to shift data away from a cloud provider) may inhibit customers from migrating between cloud providers.

Localization is a global challenge

The relationship between the European Union and United States is a high-profile example of how data sovereignty frameworks develop, but it's not the only one. Around the world, many other nations have developed a stance on localization (this is not an exhaustive list):

Russia: One of the earliest and most rigorous enforcers of data localization, Russia's Federal Law No. 242-FZ mandates that Russian citizens' personal data be stored within its borders.¹² This has had far-reaching consequences for global companies and brands. For example, LinkedIn has seen access restricted in Russia since 2016, as a consequence of its localization laws.¹³

China: The Chinese Cybersecurity Law (2017) stipulates that "critical information infrastructure operators" store personal data and important business data gathered or generated in China within Chinese territory.¹⁴ China has since consolidated its position on localization with the Data Security Law and Personal Information Protection Law (2021), which enforces classification of data and governs international transfers depending on the classification.¹⁵

Saudi Arabia: Saudi Arabia introduced its Cloud Computing Regulatory Framework (CCRF) in 2018,¹⁶ which specifies data sovereignty conditions on cloud service providers and enforces data residency for certain types of data. In 2021, it introduced its Personal Data Protection Law, under which citizen personal data can only be transferred overseas under a limited set of conditions.¹⁷

Sovereignty concerns are likely to escalate

Data sovereignty is already a prominent theme in international discussions, and its significance is likely to grow. These factors are likely to become more prominent in 2024 and beyond, from tensions between companies and nations regarding personal data, to geopolitics like the Russia-Ukraine war, to cloud complexity, with a blend of hybrid and multi-cloud structures, to data protection and cybersecurity issues. Reliance on data is immense, and cloud will likely be the de facto solution for storing, managing, and analyzing it.

Company vs. state: There is a balance between national security and individual privacy, and high-profile cases have often compelled governments to set assertive jurisdictional boundaries over data. For example, the widely publicized case in 2016 in which Apple refused to bypass the encryption on an iPhone, related to an FBI terror investigation.¹⁸ While this case did not cross jurisdictions, it did underscore the complexities of pitting the interests of state security agencies against data privacy rights and the responsibilities of tech companies to protect user data. Similar cases have occurred across jurisdictions, however, such as the 2013 dispute between Microsoft and the US government over data stored in Ireland that was related to a narcotics investigation.¹⁹

Geopolitics: Global tensions between major powers have resulted in technology and data becoming arenas for diplomatic and trade skirmishes.²⁰ The Russia-Ukraine war has brought data jurisdiction issues to the fore, with Ukraine quickly migrating critical data like its population register, land and property ownership, tax payment records, and education records to cloud.²¹ On the other hand, US-based cloud service providers suspended sales in Russia shortly after the invasion.²²

Cloud complexity: Initially, cloud was about shifting on-premises operations to a single cloud provider. However, today's enterprises often leverage multiple cloud platforms simultaneously—an approach known as multi-cloud—to benefit from features of each provider, optimizing costs, performance, and scalability. Another common approach is hybrid cloud—a combination of private (on-premises) and public cloud solutions—to allow data-sensitive applications to remain in-house while other workloads can benefit from the expansive resources of the public cloud. While multi-cloud and hybrid cloud strategies offer flexibility and optimization, they also introduce challenges. Data is dispersed across various environments, potentially across different jurisdictions. Each cloud provider may have data centers in numerous countries, each with its unique data protection regulations. This makes the management and governance of data sovereignty more intricate than ever. It's no longer about navigating the rules of one country but understanding a complex tapestry of global regulations.

Data protection and cybersecurity: Generative AI, machine learning, and automation (which are often enabled by cloud computing) are likely to become pivotal for business operation, prompting governments to be even more scrupulous when it comes to data jurisdiction. Data breaches have become common, and data localization mitigates risks in some instances. The 2020 SolarWinds hack, which impacted multiple US government agencies, highlighted the vulnerabilities of centralized cloud systems.²³ The attack, in which hackers inserted malicious code into a routine software update, showed that security and sovereignty should be addressed across entire supply chains, not just at the individual level. Companies should understand the software-as-a-service and third-party providers their organization uses: which cloud platform they run on, what sort of data they process, what their encryption levels are, and what the risks are. Such incidents make a compelling case for countries to have more control and oversight over data concerning their citizens. Cybersecurity challenges are set to exacerbate: The cost of cybercrime is predicted to be US\$14.6 trillion in 2024, more than double the US\$6 trillion cost a few years ago in 2021.²⁴

Sovereignty extends beyond geography. It also includes operation and governance. One example is Amsterdam Trade Bank, which was sanctioned by the US government in 2022 over its Russian ownership.²⁵ Its data might have resided in Europe, but the cloud service provider with operational control was US-based and was still able to revoke access to company email accounts and related data. Cloud providers have partnered with local operators to overcome this type of risk, with some governments and other entities demanding that it's not just enough for the data to reside in a geography; the operator of that cloud infrastructure must also be local.²⁶

As the world becomes more digitally connected, the need to delineate boundaries, ensure security, and protect citizen rights is becoming ever more critical. Emphasis on data and cloud sovereignty is expected to intensify in the coming years, driven by geopolitics, security concerns, and the protection of individual rights.

Sovereignty solutions are an opportunity but also a challenge for cloud providers

Cloud service providers, recognizing the growing importance of data sovereignty, have rolled out various products, services, and features. Government cloud solutions are one example—specialized cloud computing environments tailored to meet the strict regulatory and compliance needs of governmental agencies.

Cloud providers have extended cloud services to the enterprise edge. One example includes a fully managed service that deploys the infrastructure (and therefore, cloud services) to clients' on-premises location.²⁷ Another allows enterprises to run cloud services from their own data center, ensuring data remains on-premises or within a particular jurisdiction. Cloud service providers tend to offer a portfolio of solutions, to cater to businesses that need to keep data within certain regions due to regulatory requirements.²⁸

Such products are aligned with “distributed cloud,” though all may not be strictly labeled as such. Distributed cloud refers to the distribution of public cloud services to different physical locations, while the operation, governance, updates, and evolution remain the responsibility of the originating public cloud provider. In simpler terms, it's about bringing the cloud closer to where data is generated and consumed.

While these services offer a range of benefits, including latency, they also come with some drawbacks compared to traditional cloud services, such as:

- **Cost:** Distributed cloud solutions often require upfront investments in hardware and infrastructure, as opposed to traditional cloud services, which are pay-as-you-go. Also, even though the main cloud provider manages the software stack, the onsite hardware can lead to additional maintenance costs. Finally, IT teams might need training to manage and operate these new distributed cloud environments efficiently.
- **Complexity:** Integrating distributed cloud services with existing on-premises systems can be complex. Operating in a hybrid or multi-cloud mode likely means managing workloads across different environments.
- **Limited services:** Distributed cloud offerings might not have the full suite of features available in the central public cloud. Features and updates available in the central cloud might take time to become available on distributed cloud platforms.
- **Constrained scalability:** While traditional public cloud services offer virtually limitless scalability, distributed cloud solutions might be limited by the local infrastructure's capacity. Increasing capacity might necessitate additional hardware investments, whereas in a traditional cloud, it's often a matter of provisioning more resources through software.
- **Vendor lock-in:** Relying on a particular cloud provider's distributed solution can lead to vendor lock-in, making it challenging to switch providers or use multi-cloud strategies without significant efforts and costs.
- **Performance:** The performance of distributed cloud hardware on-premises might not always match the performance of the infrastructure in the cloud provider's data center. Even with on-premises or edge

deployments, there might be scenarios where data needs to traverse to the central cloud, leading to potential network bottlenecks.

For cloud providers, growing demand for sovereign cloud might create an opportunity to sell more high-value services, but on balance, it could chip away at profitability. For them, the economically optimal outcome is to sell hyperscale public cloud in every jurisdiction without restriction. But the fragmentation of global cloud infrastructure, with bespoke architectures tailored to stringent compliance mandates, can lead to higher operational costs, and squeeze margins, even if such services are sold at a higher price. That said, it's an opportunity for local service providers, as well as traditional hardware vendors, especially as customers increasingly rely on hybrid cloud (and hence need a lot of infrastructure).

The bottom line: Companies should act, not react

For companies operating in today's global digital economy, adhering to regulations about how data is stored, managed, and processed is critical not only to avoid significant legal repercussions and fines, but also to maintain trust with customers and partners. As geopolitical landscapes shift, regulations may be updated, and concerns over data privacy intensify, a company's ability to navigate data and cloud sovereignty issues can directly impact its market reputation, operations, and bottom line. Companies should act to best position themselves.

Firstly, they should conduct a comprehensive data audit. This should include identifying data sources, and classifying data based on sensitivity. For instance, personal user data might be treated differently from anonymized analytics data or metadata. Companies should also consider a data residency strategy if they don't have one. This involves deciding where data will reside based on technical performance needs (like latency) and regulatory requirements, and might mean using local data centers, distributed cloud, or cloud regions. And finally, companies should review their data storage and transfer policies, and ensure data is encrypted, at rest and in transit. If data does cross borders, encryption can offer an added layer of protection against unauthorized access.

Leading practice would also involve investing to understand local regulations, which might involve engaging local experts, and also training staff across multiple departments (e.g., IT, legal, operational), particularly as regulations change and evolve. Beyond that, companies should be as transparent as possible with partners, clearly communicating to customers and supply chains how and where their data is being stored and processed. For supply chains in particular, companies should ensure they understand how and where suppliers store and process data. Finally, companies also need to develop a strategy in case of data repatriation for scenarios where data needs to be moved back from a cloud or foreign server to a local server. If possible, they should ensure any contracts with cloud providers have provisions that allow for changes.

Sovereignty is a journey that should be embedded in current cloud strategies, and all cloud users may need to design and architect for sustainable and sovereign platforms. That journey should include three stages:

1. **Advise**, in which a company should define its sovereignty posture, design its sovereignty strategy (which would include data and workload categorization), and develop proofs of concept to prepare for sovereignty.
2. **Implement**, in which a company should architect sovereignty and implement data controls.
3. **Operate**, in which a company should manage its sovereignty ecosystem, build methods to improve observability and risk monitoring, and consider automation and cost optimization.²⁹

Cloud sovereignty is a major strategic issue for multinational companies. Getting things right can instill greater customer confidence, reduce the risk of legal repercussions, and secure a company's data assets. And the regulatory landscape is unlikely to stand still. If bytes of data are like grains of sand on a beach, then regulatory changes are the tide that can disrupt, reshape, and wash it away. Companies should ensure compliance and foster trust. One way to do that is to remain constantly vigilant, always learning, and prepared to adapt as the regulatory tide changes.

Figure 1 sources

- A. Ernst-Oliver Wilhelm, "[A brief history of Safe Harbor \(2000–2016\)](#)," International Association of Privacy Professionals (IAPP), accessed November 20, 2023.
- B. Court of Justice of the European Union, "[The Court of Justice declares that the Commission's US Safe Harbour decision is invalid](#)," press release, October 6, 2015.
- C. European Commission (EC), "[EU Commission and United States agree on new framework for transatlantic data flows: EU-US Privacy Shield](#)," press release, February 2, 2016.
- D. EU, [Directive 95/46/EC](#) (General Data Protection Regulation).
- E. US Congress, [Clarifying Lawful Overseas Use of Data Act](#) or the CLOUD Act.
- F. Hendrik Mildebrath, "[The CJEU judgment in the Schrems II case](#)," European Parliamentary Research Service (EPRS), September 2020.
- G. EC, "[Adequacy decision for the EU-US Data Privacy Framework](#)," July 10, 2023.

By

Ben Stanton
United Kingdom

Alfons Buxo Ferrer
Spain

Gillian Crossan
United States

Paul Lee
United Kingdom

Adam Gogarty
United Kingdom

Kevin Westcott
United States

Endnotes

1. IDC and Statista, “*Data volume creation and consumption in the future*,” 2020.
2. All the Trivia, “*How many grains of sand are on Earth?*,” April 25, 2022; Per C., “*Zettabyte*,” TechTerms.com, last updated December 15, 2012.
3. Gartner, “*Gartner forecasts worldwide public cloud end-user spending to reach nearly \$600 billion in 2023*,” press release, April 19, 2023.
4. Deloitte estimate, based on primary research, and factoring in industry research by IMARC Group, Mordor Intelligence, Straits Research, and Data Bridge market research.
5. Markets and Markets, *Distributed cloud market*, August 2022.
6. By the early 1960s many people can share a single computer, using terminals . . . these are the first common multi-user systems,” from “*Timesharing – the first online communities*,” Computer History Museum, accessed November 20, 2023.
7. Septimiu-Vlad Mocan, “*VPN history – Everything you need to know about VPN development over the past 25 years (and a quick glimpse of the future)*,” TechNadu, February 4, 2020.
8. Amazon, “*Amazon.com launches Web Services, developers can now incorporate Amazon.com content and features into their own web sites; extends ‘welcome mat’ for developers*,” press release, July 15, 2002.
9. US Congress, *Clarifying Lawful Overseas Use of Data Act or the CLOUD Act*, H.R. 4943, 115th Cong. (2017–2018); European Union (EU), *Directive 95/46/EC (General Data Protection Regulation)*, April 27, 2016.
10. Klaus Foitzick, “*U.S. CLOUD Act vs. GDPR*,” activeMind.legal, February 29, 2020.
11. UK.gov, “*UK/USA: Agreement on Access to Electronic Data for the Purpose of Countering Serious Crime [CS USA No.6/2019]*,” October 7, 2019.
12. Stanford, “*Amending Certain Legislative Acts of the Russian Federation as to the Clarification of the Processing of Personal Data in Information and Telecommunications Networks*” July 21, 2014.
13. BBC News, “*LinkedIn blocked by Russian authorities*,” November 17, 2016.
14. Jack Wagner, “*China’s Cybersecurity Law: What you need to know*,” *The Diplomat*, June 1, 2017.
15. Ryan D. Junck et al., “*China’s new Data Security and Personal Information Protection Laws: What they mean for multinational companies*,” Skadden, November 3, 2021.
16. Communications, Space & Technology Commission (CST), “*Cloud computing*,” accessed November 20, 2023.

17. Abdulaziz Al-Bosaily, Masha Ooijsaar, and Dino Wilkinson, “*Saudi Arabia issues Personal Data Protection Law*,” Clyde & Co., September 26, 2021.
18. Tim Bradshaw, “*FBI ends stand-off with Apple over iPhone*,” *Financial Times*, April 22, 2016.
19. Leo Kelion, “*Microsoft battles US over warrant for drugs case emails*,” BBC News, September 9, 2015.
20. Jean Gil Barroca, Alfons Buxo, and Bruno Silva Batista, “*Cloud sovereignty: Three imperatives for the European public sector*,” Deloitte Insights, 2023.
21. Tim Anderson, “*‘Russian missiles can’t destroy the cloud’: Ukraine leader describes emergency migration*,” *The Register*, November 30, 2022.
22. Ron Miller, “*Amazon, Microsoft and Google have suspended cloud sales in Russia*,” *TechCrunch+*, March 10, 2022.
23. Saheed Oladimeji and Sean Michael Kerner, “*SolarWinds hack explained: Everything you need to know*,” *TechTarget*, November 3, 2023.
24. Anna Fleck, “*Cybercrime expected to skyrocket in coming years*,” *Statista Technology Market Outlook*; National Cyber Security Organizations; FBI; IMF.
25. Jacob Atkins, “*Solvent but bankrupt: How sanctions felled Amsterdam Trade Bank*,” *Global Trade Review*, May 31, 2022.
26. For example, see Matt Small, “*Hyperscaler and VMware sovereign cloud solutions indicate that local partnerships are key to the offering*,” *Analysys Mason*, September 4, 2023.
27. AWS, “*AWS Outposts family*,” accessed November 20, 2023.
28. Google Cloud, “*Anthos*,” accessed November 20, 2023.
29. Deloitte, *Cloud sovereignty: Unleashing the potential of sovereign cloud: A gateway to resilience and adaptability*, 2023.

Acknowledgements

The authors would like to thank **Jean Gil Barroca**, **Robert MacDougall**, **Lucia Lucchini**, and **Leslie Wolf** of Deloitte LLP, and **Vipul Mehta** and **Kirti Khattri** of Deloitte SVCS India Pvt L for their contributions to this article.

Cover image by: **Manya Kuzemchenko**

Life after debt: Venture debt funding could grow again in 2024

After a rough 2023, tech companies should expect to see a rebound in technology venture debt funding, fueling innovation.

ARTICLE • 7 MINUTE READ

After four straight years of \$30+ billion in US venture debt activity, followed by a plunge in 2023 to an estimated US\$12 billion, a partial bounce-back may be next (figure 1). In 2024, US tech venture debt could rise to US\$14–16 billion, up 25% from 2023 levels.

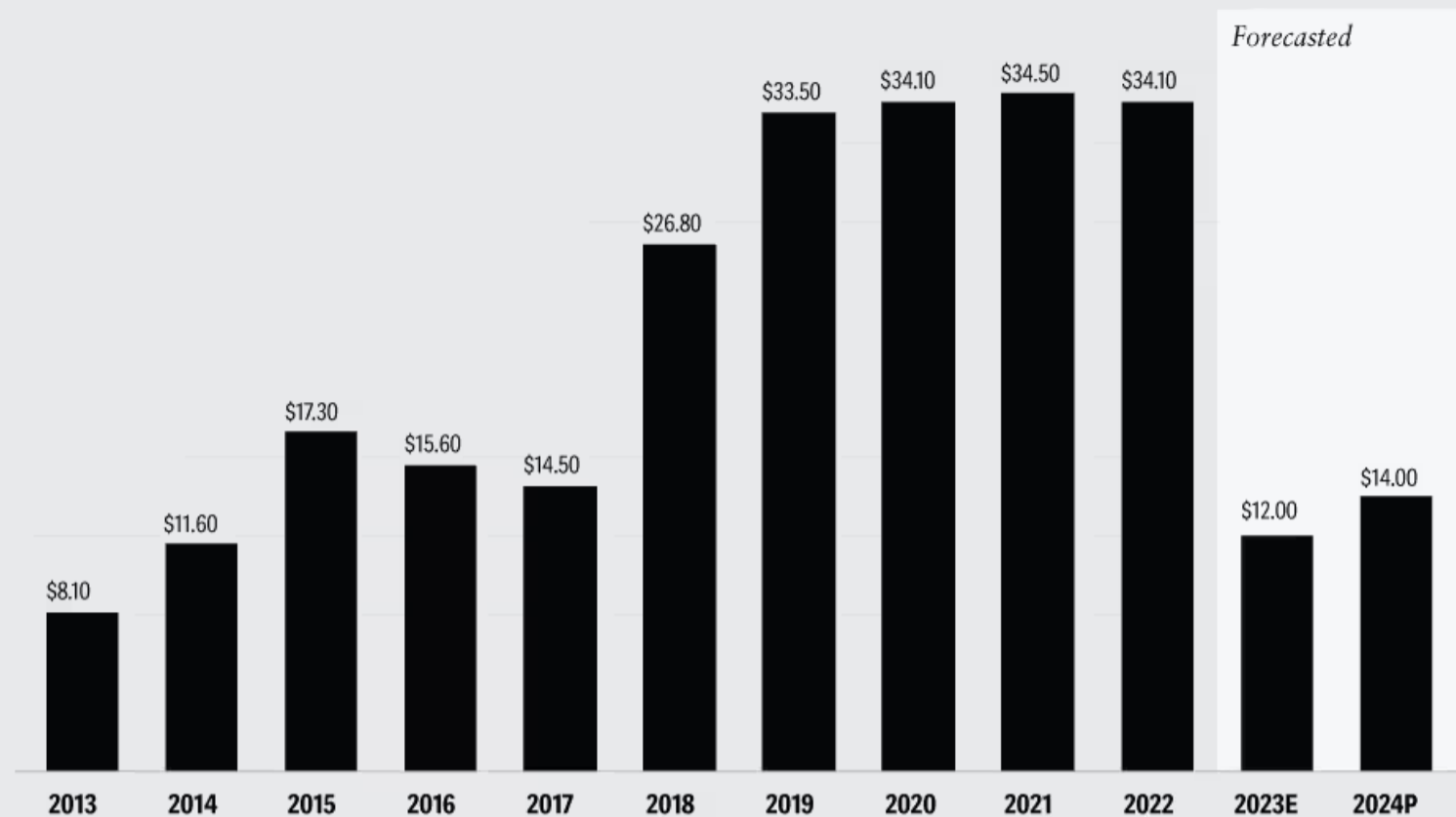
The failure of Silicon Valley Bank (SVB) in the first quarter of the year drew headlines, but that was only part of the story. Despite stock market strength in the largest (mega-cap) big tech companies, smaller tech was adversely affected in part by rising interest rates, investors focusing on later-stage companies, fewer Initial Public Offering exits (IPOs), a slowdown in mergers and acquisitions, and weakness in venture capital investing (also down about 60% from 2021 levels).¹ Despite these losses, the venture funding market (especially debt) should rebound over time, in part due to the rise of alternate lenders such as large private equity firms and alternative non-bank lenders gaining a foothold in venture lending.

There are over 50,000 VC backed companies in the US alone, and many are unprofitable and burning cash.² They tend to have around two-to-three years of cash but require steady capital infusions to meet their ongoing expenses. That money can come from investors who purchase equity stakes (ownership), known as venture capital (VC). Another source of capital for these companies is debt: They borrow the money from lenders and pay interest, but don't need to dilute their equity stake to the same extent as a VC deal. Venture debt has become an increasingly popular way for companies to raise money: Between 2017 and 2022, the US venture debt market went from 10% of the size of the US VC market in 2017 to 14% in 2022.³

Figure 1

A partial bounce back in venture debt activity appears to be underway

US venture debt activity (US\$ billion)



Note: E indicates estimated value, and P denotes predicted value.

Source: Data for 2013 to Q1 2023 based on PitchBook data as of March 31, 2023; 2023E and 2024P values are Deloitte's estimate based on our analysis of partial data for 2023 and our prediction for 2024.

Deloitte Insights | deloitte.com/insights

Technology venture debt grew by 2% in 2022. But based on a reported US\$2.8 billion in Q3 for tech venture debt—which would aggregate to about US\$12 billion annually⁴—Deloitte expects 2023 will experience a 60% year-over-year decline. That funding level is well below the 2021–2022 peak, but still in line with the 2015–2017 levels. However, although data suggests a medium-term decline in available venture capital, we anticipate venture debt to continue to be a growing asset class over time, with venture debt recovering and growing into 2025 and beyond. We expect to see growth of up to 20% of overall venture funding market size, relative to the growth of venture capital, by 2027.⁵

What are the benefits of venture debt, and how is it different from venture capital?

Venture debt financing is an option for startups that have usually already received at least one round of venture capital funding. Securing venture debt provides founders of fast-growing startups a way to raise money while retaining more equity in their companies. Unlike venture capital, which trades equity in the company for capital, venture debt allows founders to retain their equity, but the loan must be paid back over time.

Venture debt loans often require startups to pay interest, fees, and offer warrants to the lender, which can be taken as equity down the line. The returns on venture debt lending are lower than traditional VC funding, but lending is also less risky.⁶

Although there is good data available for the global VC market, and there is robust data for US venture debt, there is a scarcity of good global venture debt data, both current and historical. Deloitte believes that the trends seen over time in the United States are also true globally: Venture debt was about 10% of the VC market size pre-pandemic, rose to about 15% by 2022, and in 2023 the venture capital and venture debt market in 2023 will both decline by more than 50%, meaning that venture debt will still be about 15% the size of VC.⁷ We estimate the US market for both venture capital and debt to be about half the global total, suggesting that global venture debt in 2023 will be around \$20–25 billion, and will be over \$30 billion in 2024. We anticipate, however, that the global market might feel the impact of a venture debt downturn less prominently in the shorter term, because much of the venture debt market in Europe and Asia is served by state-backed venture debt.

What's driving the current decline of venture debt?

The slump in lending was mostly brought on by rising interest rates and risk aversion.⁸ Still, with both venture capital and venture debt funding shrinking and strategic investors becoming more risk-averse, it's becoming increasingly difficult for unprofitable companies to raise funds; we predict this trend will continue until well into 2024.

In the short term, venture debt deals could become smaller, rates higher, and generally harder to obtain. Similar things are happening in VC: Fewer deals, smaller deals, lower valuations ("down rounds," or deals done at valuations lower than previous rounds), and more restrictive terms.⁹ Companies that might have preferred to borrow might need to raise capital and dilute their equity, divest less profitable business segments, or be acquired by competitive buyers during a vulnerable time. We expect to see strategic VC deals (either equity stakes or outright purchases) rise in 2024, as cash-rich, mega-cap tech companies with high valuations invest or buy out smaller companies that can't raise capital or debt at acceptable prices.

With technology venture funding more difficult to secure, we may also see strategic buyers seize the opportunity to acquire or obtain stakes in early-stage companies that are dependent on more competitive venture funding. While VC funds run by large corporations (venture arms) might have preferred to purchase percentages of early-stage tech companies, we may begin to see them make moves to acquire them outright to gain a competitive edge.¹⁰ We may also see lenders begin to ask for more warrants in their deals, allowing them to buy stock at a predetermined, lower strike price.

Venture debt market prediction for early-stage tech companies

In the wake of these developments, early-stage tech and telecom companies should weigh carefully how to proceed. With venture debt more difficult to obtain, early-stage tech organizations should prepare to show robust revenue and strong margins to be competitive in securing difficult-to-obtain funding. Where the focus was solely on revenue growth and market share in 2021, companies may want to reduce costs and show a faster path to profitability. Venture debt will likely become expensive, so even companies that are able to secure funding should still plan for steeper costs.

Less venture money and high venture debt cost are also likely to affect the innovation ecosystem. Given that options to buy and acquire may be fewer (because promising startup ventures with positive cashflows could be fewer), tech companies can focus within their own enterprises on talent strategy to improve their rates of innovation in cost-effective ways. Those efforts might include offering specialty training programs for highly valued employees and considering relationships for research and development within the tech ecosystem to help ensure they don't lose their competitive edge.

Financial institutions should seize the opportunity to gain a competitive edge

Financial institutions should consider the considerable gap in the venture debt market, which means there is a significant opportunity for other venture debt lenders to fill the market demand. With banks behaving in ways that are more risk-averse, we're already beginning to see large private equity firms and alternative non-bank lenders start to fill the venture debt space.¹¹ We anticipate that these lenders will drive the recovery of the technology venture debt market.

Bottom line

During this rebound phase, early-stage start-ups should plan to keep more cash on hand to weather economic uncertainty if funding becomes more difficult to obtain down the road. They should also plan for the risk that debt may be challenging or impossible to obtain, and they may need to raise further equity rounds at lower prices. These developments in the market are likely to encourage early-stage tech companies to build more stable growth over time.

Regulators had seen early signs that SVB was in trouble as early as 2021, however, bank managers failed to address the problems.¹² In the wake of the bank failure, questions have been raised about the role of regulatory bodies in preventing bank failures. In 2018, regulations in the Dodd-Frank Act which went into effect during the 2008 recession were amended to exempt banks with assets between US\$100–250 billion from keeping sufficient liquid cash for thirty days of withdrawals on hand at all times.¹³ Some have argued that this change, combined with SVB management's failure to address concerns, appears to be a significant contributor to the bank's failure.¹⁴ As a result, many observers anticipate more regulation,¹⁵ which could contribute to the cooling of venture debt markets in the short term.

For financial institutions, the collapse and rebound of the venture debt market presents a compelling opportunity to gain market share. Institutions moving into the venture lending space should proceed cautiously to manage heightened risk. For example, lenders may need to ask for more warrants, in order to buy stock at a predetermined price, possibly at lower strike prices, in order to help reduce their risk factors.

Nonetheless, one possibility that venture debt could unlock for the tech startup ecosystem is serving as an additional pathway to raise funding. And it could work well both for the lender (less risky) and the borrower (access to funds of smaller ticket size). Beyond traditional VCs, venture debt could serve as an alternate asset class for tech startups to keep the innovation engine on. For instance, venture debt lenders can help startups by offering lighter funding in the range of US\$5–8 million to support pilots and prototypes (for example, generative AI solutions) via strategic joint ventures, academic collaborations, joining hands with adjacent tech players as part of an industry consortia, or participating in ideation labs and other accelerator programs.

Moreover, startups are sprouting in new areas such as sustainable tech (for example, ESG software and analytics, AgTech) and generative AI and private large language models. These new areas are expected to require a continuous flow of funds to help those startups launch innovative solutions to help generate value for their customers. Venture debt can serve as another plausible alternative investment avenue for the hundreds of new/emerging startups, beyond the established investment avenues such as VCs and PEs.

All said, tech start-ups may still have to think carefully about their funding strategies in the coming year as venture debt funding recovers. Still, as the market recovers, tech start-ups will likely be well-positioned to build toward sustainable growth in the future.

By **Bree Matheson**
United States

Duncan Stewart
Canada

Sriram Prakash
United Kingdom

Karthik Ramachandran
India

Endnotes

1. Mike Dickson, “[Are the "Magnificent 7" tech stocks ready to share the road?](#),” Nasdaq, August 04, 2023.
2. PitchBook Venture Monitor, Q1, 2023.
3. PitchBook Venture Monitor, Q4, 2022.
4. Ibid.
5. PitchBook Monitor Venture, Q1, 2023.
6. CB Insights, “[Venture debt vs. venture capital in one graphic](#),” November 18, 2020.
7. PitchBook Venture Monitor, Q1, 2023.
8. John Turner, “[Why did Silicon Valley Bank fail?](#),” Economics Observatory, March 17, 2023.
9. Dean Takahashi, “[Global VC deals declined in Q3 for the second quarter in a row, hitting 3-year lows](#),” Venture Beat, October 3, 2023.
10. Karthik Ramachandran, Duncan Stewart, Sriram Prakash, and Gillian Crossan, [TMT divestitures make a comeback 2023 deal values in tech, media, and telecom may bounce back strongly](#), Deloitte Insights, November 30, 2022.
11. Nate Bek, “[Startup leaders sound off on the future of venture debt in fallout from Silicon Valley Bank collapse](#),” GeekWire, March 24, 2023.
12. Christopher Rugaber, “[Fed official: Bank rules under review in wake of SVB failure](#),” AP News, March 29, 2023.
13. Ibid.
14. Rachel Siegel, “[Fed says it must strengthen banking rules after SVB’s collapse](#),” Washington Post, April 28, 2023.
15. Phil Rosen, “[The Fed is about to tighten the screws on America’s biggest banks](#),” Markets Insider, July 11, 2023.

Acknowledgements

The authors would like to thank **John Forster** for his contributions to the article.

Cover image by: **Manya Kuzemchenko**

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee (“DTTL”), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as “Deloitte Global”) does not provide services to clients. In the United States, Deloitte refers to one or more of the US member firms of DTTL, their related entities that operate using the “Deloitte” name in the United States and their respective affiliates. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms..