XR ON-DEVICE PERCEPTION AND AI

Eric Abbruzzese, Research Director

ABIresearch THE TECH INTELLIGENCE EXPERTS™ adeia

CONTENTS

STATE OF THE MARKET: AI AND XR

INTRODUCTION

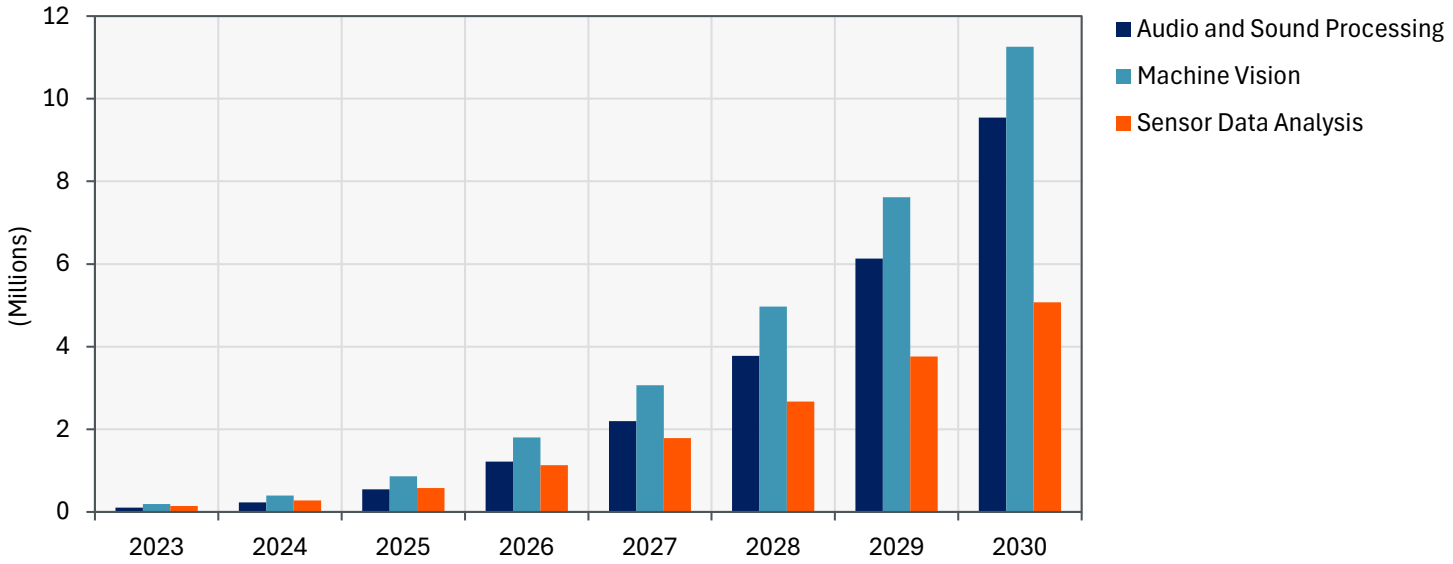
Over the past few years, the Extended Reality (XR) market has seen some significant shifts. After a wave of interest in smart glasses, on the back of Google Glass, and Virtual Reality (VR) headsets, on the back of the Oculus Rift, the market seemed prime for significant growth. And while XR has indeed matured and grown since that time, the focus areas and expected key use cases have not stayed constant.

Enterprise usage has focused on adding value in key areas. Remote assistance and training have been standout applications across enterprise verticals, with strong Return on Investment (ROI) and quick time to value. The consumer XR market has been less straightforward—VR has blossomed in media & entertainment, especially gaming, while Augmented Reality (AR) smart glasses have struggled to find footing. This is partly due to a lack of choice among consumer-targeted AR devices, combined with an unclear value proposition for potential buyers.

Thanks to AI, this is beginning to change. Another wave of interest is expected over the next few years, focusing on enabling Artificial Intelligence (AI)-centric use cases across domains. Virtual assistants, such as Siri and Google Assistant, are already integrating AI capabilities to enhance interaction, accuracy, and overall capability. This is especially pertinent to smart glasses—devices that maintain natural user vision and overlay data, rather than fully replace user vision as in VR. Smart glasses target more varied usage patterns, such as in public and on the move. This is especially true when targeting the consumer market, compared to the enterprise space where usage can often be limited to a specific environment (e.g., a factory floor, office, etc.). Smart glasses have the potential to be an all-day wearable device, in a similar vein to smartwatches with greater capability.

Chart 1: Shipments for TinyML Devices in XR Devices with Edge AI Chipsets by Use Case World Markets: 2023 to 2030

(Source: ABI Research)



Having constant access to an AI-enabled device has significant potential value. The combination of real-time machine vision, hands-free input and interaction, and visualization capability is only possible with XR. Head-worn devices are excellent for this, capturing and leveraging many types of data in parallel. "Regular" data act similarly to any modern connected device—applications used, interaction parameters, location, etc. XR can add some unique data types to this list—forward-facing cameras for machine vision and (when applicable) eye/gaze tracking are most distinctive.

Bringing in these multiple data types and sources is a processing challenge, and one that AI can both streamline and add to. Cameras, spatial and environment data, voice, gesture, gaze, location, and more can be captured simultaneously and leveraged for various applications. AI thrives with as much data as possible, and XR presents a unique host device for enabling that data access and delivering a compelling use case for users—AI-centered and otherwise.

DEVICE CAPABILITIES AND CONSIDERATIONS

To power these applications, devices need a certain level of compute capability to handle on-device processing. Many chipsets available today offer dedicated AI sections on a chipset to handle this processing with the greatest efficiency. When dealing with local processing, a balance must be reached between power efficiency and outright computing power potential.

The camera is perhaps where most impact is seen when combining AI and XR devices. While not an outright requirement for an XR device, front-facing cameras are basically standard. Tapping into that camera system ranges from a requirement for function—in the case of spatial tracking—to a value add in something like object recognition.

Any XR experience requires some form of input to allow interaction with the device, and AI plays a role here as well. Voice recognition and control is commonplace for smart glasses, and going forward, will be directly tied to virtual assistants. AI helps with accuracy and reliability, and can speed up response rates with on-device Natural Language Processing (NLP). For more complex interaction, XR devices can use eye tracking thanks to the positioning of the device. Gesture control is also an option, using a combination of cameras and sensors to track hand and finger movement. Gesture control is not technically hands-free, so it has not been used as commonly as voice control, but the presence of forward-facing sensors and cameras, combined with AI sensor processing, does make it an option.

Latency must also be considered. Much of the most powerful AI services are cloud-based today. This is acceptable for the target applications of those services—either not user-facing, or non-time-sensitive when users are involved. In XR, however, everything is user-facing and some data types and experiences will require low-latency processing and visual feedback. Voice-driven virtual assistants may not need to be low latency, but real-time object recognition and visual overlay need to keep up with a user's movement. If a device is trying to overlay visual information in real time, and must send data round trip to and from a network, it is very difficult to align data updates with that user movement.

KEY USE CASES

There are plentiful use cases that will be valuable and impactful through the combination of AI and XR, but the following offer a selection of the most interesting—thanks to potential scale, broad value, and uniqueness. These use one or more of the above enabling elements, and create a compelling, valuable experience for the end user. Each has varying processing requirements, which are important to understand when examining the hardware capabilities of future XR devices.

TRANSLATION

Translation has been a quintessential technology use case for decades. Real-time visual translation is already being done, and XR devices present an exciting next step to streamline the experience for users. It is easy to market—not something that can be said of all XR experiences—with easily understood application and value. It also integrates very well with the existing virtual assistant platforms and is a focus of broader AI efforts as well. Advancements in quality and speed are coming, which XR devices will showcase to users well.

Translation also has relatively low compute requirements, making it a win-win. In terms of hardware, the only requirements are a camera and some compute for machine vision. Value can be added through dedicated AI processing for high-efficiency and low-latency translation, but even simple character recognition pushed to the cloud and back can be valuable. Real-time visual overlay of translated content has more significant visual processing requirements and needs faster round-trip processing speeds (time from camera recognition to visual update to the user).

SPATIAL CONTENT

Spatial compute was a hot topic for some time and is expected to be so again soon. Spatial content leverages environmental information to enable Three-Dimensional (3D) experiences in space. Sometimes referred to as 6 Degrees of Freedom (6DoF), spatial content, put simply, allows for movement in an environment. Through an XR device, this can be either a full VR experience, with a fully digital world, or a mixed reality experience using optical or camera pass-through.

To enable this requires significant sensor hardware and processing. First, the device must be able to track itself in space and the relative movement within that space. This is usually done through Simultaneous Localization and Mapping (SLAM), a machine vision technology more commonly known through autonomous vehicle usage, but equally applicable for XR spatial tracking. This inherently must be done with an on-device processor, as visual updates need be as close to real time as possible.

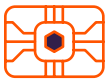
Second, the device must be able to display 3D content and map it to the environment. AI can enable complex environmental understanding, with machine vision enabling object recognition not only to enable tracking in space, but to capture semantic data about the surrounding environment—people, specific objects, potential hazards, and more can be identified and passed to the user. Once again, this needs to be as close to real time as

possible, so on-device processing often wins out. There are some situations when a hybrid compute approach is possible—such as movement tracking done on-device and graphics processing done in the network. This can alleviate some local processing load at the expense of more complicated content delivery and opening additional failure points (the network).

REMOTE COMMUNICATION AND COLLABORATION

Having gained traction as a COVID-19 reaction and sticking around since then, remote communication and collaboration is a high-value XR use case across both smart glasses and VR headsets. This has taken a few primary forms. Enterprises have been gradually growing capabilities in XR training, thanks to proven improvements in retention and recall for VR training content compared to paper and video training. Smart glasses are excellent for hands-free remote assistance applications, work instruction, and verification. Value is yet to be proven in the consumer market, although there is interest and activity already, mostly through a social media focus.

Depending on the level of capability delivered for the experience, compute requirements can range from mild to significant. In the simplest implementation of a remote assistance use case, for instance, a two-way audio and video feed is the only real requirement—enabling that hands-free is the primary value. This would only need enough compute power to enable camera capture and streaming to the network. More complex features can be added to that, and commonly are—3D tracked annotations add a level of clarity, for instance. Full spatial compute experiences are also possible, with multiple remote users tracked in space in a shared virtual environment. This has been done, but is a challenging implementation—XR devices involved must be able to handle SLAM tracking for the device and push that to the network in real time, while also doing graphics processing for the visual element, whether it be just a piece of content like a 3D model, or an entire virtual environment. Efficient machine vision is a requirement, and dedicated AI processing to handle this at low latency is a boon.



SHOPPABLE VIDEO

Shoppable video is a relatively new format in the retail advertising space, embedding product information and an interactive link into a video containing the item. This is most popular on social media platforms, but is being used increasingly in first-party sources like brand/store applications and web pages. It is driven by creating as direct a link as possible between content and product. While XR is too nascent to see any significant activity in shoppable video yet, the idea is being explored and developed rapidly.

XR can enable a unique experience through immersive visuals and interaction not possible through existing technology platforms. With 3D capability, products can, of course, be displayed in three dimensions. With machine vision, overlay and spatial compute offers additional capability, such as through virtual try-on experiences or high-fidelity, scale-accurate visualization of larger products like furniture or vehicles.

Simple 3D model visualization is relatively lightweight, although can scale dramatically depending on the level of fidelity desired.

The idea is not entirely new in the XR space; a similar idea was being explored at the peak of metaverse hype, and while that terminology has fallen off in recent years, the value behind the idea remains, and companies are exploring the potential of virtual product and embedding virtual products in other forms of media. Video games have successfully leveraged a similar idea, with branded content offered to players as a microtransaction. XR offers an additional addressable user base, whether through games or social/communication experiences.

Depending on the level of visual fidelity and type of content the product is embedded into, compute requirements can vary. Simple 3D model visualization is relatively lightweight, although can scale dramatically depending on the level of fidelity desired—ultra-high resolution models can sometimes stress mobile device (and similar) chipsets. AI can play a somewhat unique role here; it can enable shoppable video, even without direct product linking through machine vision recognition for products. This idea is already being leveraged without AI through "similar item" recommendations, and AI can extend that to be more automated and accurate. XR can bring a spatial element to this as well.

SOLUTION EXAMPLES

Chipsets are the foundation for processing, AI and otherwise. In the XR market, chipsets look very similar to the smartphone space, sometimes with exact chipsets shared between the device types. Qualcomm is the inarguable leader in the XR chipset space today—the company powers nearly every XR device that is not an Apple product currently. After seeing its mobile-targeted Snapdragon chipsets being used for smart glasses and VR headsets, the company spun out a purpose-built XR chipset line in the Snapdragon XR and Snapdragon AR lineups. AI is at front and center for these chipsets, with dedicated AI acceleration blocks, low-latency camera pass-through, and significant sensor support.

Apple is the "other half" of the XR market today. The only dedicated XR device currently is the Vision Pro VR headset, although XR capability is embedded into the company's iPhone and iPad lineup as well through the ARKit Software Development Kit (SDK). Vision Pro has

the most robust sensor suite of XR devices today, with a focus on high-quality pass-through for mixed reality experiences and accurate gesture control. Apple is the biggest proponent of "spatial compute" as a term and is positioning spatial compute and VisionOS as a significant element of its go-to-market strategy for the next few years. Vision Pro uses the M2 chipset, along with a purpose-built R1 chip for real-time sensor processing. AI has not been a direct element of the Vision Pro go-to-market strategy, but is expected to be going forward as the company bakes its Apple Intelligence portfolio into more devices.

The big question mark in the XR space today is Google. All expectations point to a wide-reaching XR effort this year, likely through the Android platform, but nothing has been confirmed.

Meta needs to be mentioned for both AI and XR efforts. The Meta Quest VR headset lineup is the most successful XR product launched so far, with tens of millions of shipments. The company has also, more recently, ramped up AI efforts, and even opened much of its technology around machine vision and spatial tracking for wider use. The Meta Ray-Ban Glasses include the Meta AI assistant, the company's cross-platform AI assistant platform that can use the camera data for machine vision. On the VR side, Horizon OS is now open for third parties to use, offering Operating system (OS)-level access to those AI and machine vision capabilities. Currently, Meta is partnered with Qualcomm and uses its chipsets in both Quest VR Head-Mounted Displays (HMDs) and Ray-Ban smart glasses; however, MediaTek will be powering the company's next smart glasses product, with details currently scarce.

The big question mark in the XR space today is Google. All expectations point to a wide-reaching XR effort this year, likely through the Android platform, but nothing has been confirmed. The company does have a confirmed partnership with Samsung and Qualcomm, with Google's role likely focused on the software and content enablement side. Expect Google-flavored versions of its AI across key applications, especially translation and its broader virtual assistant.

Outside of the big tech incumbents, the market is much more fragmented. There has been such significant consolidation in the AI space that the AI landscape touching XR is small—all the hyperscalers are building out AI capability across segments. Some GPU vendors are well-positioned to power cloud-based AI—NVIDIA, especially, thanks to both its AI and XR strengths, is positioned favorably. There is room for more on-device competition as well, whether through single chipsets with AI accelerators onboard or with purpose-built AI chipsets.

SUMMARY AND RECOMMENDATIONS

XR offers a plethora of opportunities to leverage AI, from chipsets to end-user applications. AI has breathed new life into the head-mounted form factor, with promises of high-value AI-powered applications both new and old. XR as a device type is a perfect fit for AI, offering valuable data to feed AI platforms and streamlining the interaction with those AI solutions through hands-free input. The ubiquity of cameras on XR devices means machine vision is always on the table, and it is up to developers to integrate that data into unique solutions for a unique product category. Some applications present challenging compute scenarios, especially when combining high-fidelity graphics processing, multi-sensor spatial computing arrays, and low-latency demands for real-time applications. With the AI market consolidating, there is still significant innovation potential when combined with XR, whether through large companies offering end-to-end solutions with AI throughout, of through more focused vendors looking to provide a tailored experience.

INVESTMENT PRIORITIES (FOCUS ON HARDWARE/CHIPSET REQUIREMENTS)

- **Identify the "sweet spot" for hardware and processors.** There will be a fight for the best balance between outright capability and battery life. This will vary by target use case somewhat, but for broad mass-market efforts, customer feedback will dictate what that sweet spot is. This also considers form factor, on-devices versus cloud-processed content, and price point (for both hardware and content/services). Just as was seen with graphics processing in the smartphone market, where the most powerful was rarely the best choice, so too will it rarely be that the most powerful or capable AI-enabled chipset is the best. Mass-market XR devices will compete more with smartwatches than with Personal Computers (PCs), smartphones, or even standalone VR devices of today, so the balance between compute power and efficiency is most important. "All day" battery life will not be feasible in the short term for XR devices, so battery life will be a key differentiator over outright AI capability.

- **Open source is a boon for (almost) all. Investing and partnering with those who favor open source helps to build a sustainable ecosystem.** Only a very select list of companies can operate entirely siloed, and this holds especially true for XR. Being early to support a broad list of vendors—be it chipsets, OEMs, developer platforms, or others—positions you favorably in a market that has been fragmented and tumultuous thus far. Adding AI to the mix only enhances the potential complexity, and once again solidifies the value of shared market building and cross-platform support.

GO-TO-MARKET OPTIONS

- **Identify a few high-value use cases and plan a market entrance accordingly.** Broad, mass-market applications in a novel technology segment are incredibly difficult to build and launch. There are likely incumbents already active or soon planning to be, with the advantage of an existing addressable market—whether that be hardware or software developers, end users, or all of the above. Fighting these incumbents with a like-for-like product is a challenge to say the least, but a slightly differentiated offering can either compete or make a compelling partner. As for AI on XR devices, the market is quite nascent with room still to compete, but that is changing quickly.

- **Be cognizant of AI consolidation.** It has already been significant in the pure AI space, meaning the small innovators will mostly be acquired and integrated into other systems sooner than later. For the XR space, this is both a blessing and a curse. On one hand, the XR space has suffered from fragmentation significantly over the years, so avoiding further fragmentation on the AI side will be beneficial. At the same time, it consolidates viable competition to a handful of companies that will fight to maintain and monetize their own ecosystem. There are some genuine open-source efforts happening in the AI and XR space, most notably from Meta today, that hope to alleviate some of this ecosystem lockdown.

IMPORTANT TIMELINES

- **The next 2 years will bring a significant increase in competitive activity across all relevant fronts.** There will be new hardware offerings from tech giants that hope to become the next smartwatch, with AI as a cornerstone offering. These will tout mature virtual assistants and integration with existing ecosystems, mostly coming from the smartphone market. In all segments, expect a balance between both more competition and more interest, meaning differentiation and visibility will be paramount.

- **In the next 3 to 5 years, activity will shift from innovation to maturation.** There will be less focus on outright compute specifications and more on how developers can leverage what is available to create something unique and valuable. The pure hype from AI has mostly shifted to realistic expectations and a desire for a marketable product. Support for the widest selection of hardware possible helps to attract developers (more users, more incentive to create), which, in turn, helps to sustain content/platforms/ecosystems.

Published September 2024
157 Columbus Avenue
New York, NY 10023
Tel: +1 516-624-2500
www.abiresearch.com

We Empower Technology Innovation and Strategic Implementation.

ABI Research is uniquely positioned at the intersection of end-market companies and technology solution providers, serving as the bridge that seamlessly connects these two segments by driving successful technology implementations and delivering strategies that are proven to attract and retain customers.