Deliverable 6.1

Augmented reality visualisation technical specifications

Project acronym:	XCROSS - The next generation of railway crossing asset management technology
Starting date:	01/10/2024
Duration (in months):	30
Call (part) identifier:	HORIZON-ER-JU-2023-EXPLR-04
Grant agreement no:	101178744
Due date of deliverable:	Month 6
Actual submission date:	20-05-2025
Responsible/Author:	TXT e-tech
Dissemination level:	PU
Status:	Issued

Reviewed: yes

This project has received funding from the European Union's Horizon Europe research and innovation programme under Grant Agreement No101178744.

Document history		
Revision	*Date*	*Description*
1		First issue

Report contributors		
Name	*Beneficiary Short Name*	*Details of contribution*
Gianluca Quaglia	TXT	Main author
Ivan Shevtsov	ProRail	Reviewer

Table of Contents

Executive Summary

The deliverable aims to describe the hardware and software requirements of the Augmented Reality (AR) viewer in the context of the XCROSS project. Specifically, the application will allow the operator to visualise the heatmap of the crossing model being repaired to better support their work and view the details of the sections to be worked on. The data will be provided by the result of 3D scanning of the crossing and digital twin processing (activities carried out in WP3 and WP4). The application will also allow the user to enter information about the maintenance session and possibly view other previous sessions that have occurred.

The identification of requirements and use cases was approached as follows: first, the state of the art of AR technologies, hardware and software was analyzed, with a special focus on the use of these technologies in low-light conditions. At this point, possible hardware solutions available on the market were reviewed specifically, to assess the use of a tablet rather than an augmented reality headset. The tablet and headset models identified were compared with each other not only for specifications, but for usability in the field. The result of this phase showed that the use of a headset, such as Apple vision Pro or Meta Quest 3, would be difficult along with the safety devices used by operators. This led the choice toward a hand-held device, specifically a tablet that has a screen large enough to provide a good experience, and with tech specs suitable for the purpose, like the iPad Pro.

Once the hardware specifications were identified, it was possible to look for software requirements: application use cases. In this phase, the actions that the operator can perform while using the AR application have been described, from entering the session data and visualizing the heatmap in the real world to closing and saving the maintenance session data.

The use cases led to the identification of additional, more specific technical requirements for data exchange and communication with external components that will provide essential input to the AR application. As a result, a logical architecture was designed, encompassing all internal and external components of the application, including backend components for data storage and exchange.

Finally, some key implementation details are discussed, particularly the selection of the graphical engine for developing the AR application: Unity. This choice is justified by the cross-platform nature of the software, which allows the application to be implemented using high-level interfaces and libraries that are not tied to specific hardware. This will enable the future extension of the tool's use to devices that were initially discarded for logistical reasons, should these obstacles be overcome. Such an extension can be made with minimal effort.

Thus, following the top-down approach, the goal of identifying a hardware solution and describing the software requirements and use cases for the application has been achieved, providing a comprehensive overview for the design and development of the software specifically.

Abbreviations and Acronyms

Abbreviation / Acronym	Description
AR	Augmented Reality
MR	Mixed Reality
VR	Virtual Reality
LiDAR	Light Detection and Ranging
HMD	Head Mounted Display
HUD	Head Up Display
DOF	Degrees Of Freedom
SLAM	Simultaneous Localization And Mapping
VST	Video See-Through
OST	Optical See-Through
OHMD	Optical Head-Mounted Display

Table of Figures

1 Background

The present document constitutes the Deliverable D6.1 “Augmented reality visualisation tech specs” in the framework of the Flagship Project XCROSS – The next generation of railway crossing asset management technology as described in the EU-RAIL MAWP and contributes as well to the FP3 – IAM4RAIL Flagship Project.

2 Objective/Aim

The aim of the deliverable D6.1 is to define the hardware and software requirements for an AR-based application by identifying the most appropriate technological solution for the specific use case. This process will be driven by an in-depth analysis of the state of the art in AR technologies, with a particular focus on their applicability in low light conditions. By analysing the available AR hardware and assessing its compatibility with the operational requirements of maintenance operation, the deliverable D6.1 aims to determine the optimal solution. Based on this selection, the necessary software requirements will be outlined to ensure seamless integration and effective deployment of the application.

3 Application requirements

3.1 State of the art AR hardware and software

This section aims to describe the state of the art of AR technologies with a particular focus on technologies used in the maintenance field and specific cases for low light conditions. After an introduction and definition of the technology, we will look at some of the areas and products where Augmented Reality has brought benefits to its users.

3.1.1 Introduction to AR

Augmented Reality (AR) is one of the relatively old, yet trending areas in the intersection of computer vision and computer graphics with numerous applications in several areas, from gaming and entertainment to education, training, maintenance and healthcare (Shervin Minaee, 2022)

The term AR is part of the Extended Reality (XR) family of definitions. Extended Reality, in the field of immersive technologies, represents the concept of "non-reality," of new models of human-computer interaction. To simplify, we can associate "x" with the algebraic concept of a variable that represents all the opportunities inherent to this field, such as Virtual Reality (VR), Augmented Reality (AR) and Mixed Reality (MR) (Srl, H. (n.d.)., n.d.). The three technologies differ in the level of immersion and blending of the virtual world with the real world (Figure 1).

The Virtual Reality approach is based on the multimedia creation of objects, spaces and events in order to create an artificial replica of reality. The user is fully immersed in the virtual world using a headset (Kurzweil, 2003).

Figure 1: Graphical definition diagram of the terms VR, MR, AR - Source: (Srl, H. (n.d.)., n.d.)

The techniques of Mixed Reality and Augmented Reality, on the other hand, aim to achieve a blending between the real and virtual worlds. Specifically, these types of technologies are able to enhance the real world through the addition of virtual images and data. According to the definition proposed by researcher Ronald Azuma, AR represents a system that combines the real world and virtual reality in real time allowing a person to move freely in the three dimensions.

Unlike VR, AR technology does not create a new entirely virtual environment, but enriches the real world with new images, navigation information, and photorealistic objects that blend seamlessly with their surroundings. AR is able to recognize objects in the real world and superimpose virtual elements on them, trying to make them as indistinguishable from reality as possible and to achieve an aesthetically consistent and harmonious result (Bourdot, 2018).

There are different types of technologies in which AR can be applied, each pursuing different objectives or applications, such as those mentioned below:

- **Marker-based AR** (Image Recognition): uses a camera and a marker (such as QR codes) as a reference. The system recognises the marker and overlays virtual content.
- **Markerless AR** ("By Location" or "Geolocation"): uses GPS, digital compass, accelerometer and speedometer to determine position. No need for physical markers; relies on the device's sensors. Often used for map navigation and location-based applications.
- **Projection-based AR**: uses artificial light projections on real-world surfaces. Enables user interaction by detecting changes in the light projection. Requires high computing power for real-time processing.
- **Overlay AR**: replaces part or all of the real-world view with augmented elements. Object recognition is critical to display the correct augmented content. Commonly used in medical imaging and industrial applications where enhanced object visualisation is required.

(Arena, 2022)

The numerous applications of AR require specific software and hardware. In general, however, the software used determines the coordinates of the real-world using cameras and tracking devices. The aim is to transfer this positional information to the AR application to enable the fusion of the real and virtual worlds by specifying the connection between them. This allows virtual objects to be integrated into the real world. These objects are controlled based on the movements performed by the user.

Hardware components can be divided into:

- **Input devices**: these are the devices that allow the user to interact with the AR application. The input can be of different nature, such as gestures on a touch device, a controller, or even voice commands.
- **Sensors**: these are the tracking devices in AR that are used to determine the position of the user or an object, enabling the fusion of physical and digital environments. This is achieved by visually capturing the physical world and its digital information, often using data from cameras or 3D models. Various tracking techniques are used, such as GPS and ultrasound, each with different settings and ranges. This improves the accuracy of the AR system.

- **Display**: used to visualise the results of AR applications, allowing users to see augmented content. Examples include head-mounted displays (HMDs), monitors and wearable devices such as glasses, gloves and clothing. HMDs typically have cameras that use holographic and optical techniques and are worn on the forehead for applications ranging from gaming to medical and engineering. Another similar device is the Head-Up Display (HUD), a compact and lightweight option designed to provide additional information while you're working. This category also includes devices such as Google Glass, Hololens and smartphones. The Figure 2 below shows some references

(Arena, 2022)

Figure 2: AR devices: Microsoft hololens, IPhone, Apple Vision Pro – Source: Google Images

As for the software part, the two most important players in the market, Apple and Google, have interpreted AR through their development platforms, ARKit and ARCore.

ARKit is an iOS app development kit. ARKit is an application programming interface (API) for iOS, iPadOS and VisionOS which lets third-party developers build augmented reality apps, taking advantage of a device's camera, CPU, GPU, and motion sensors. (Potuck, 2017)

ARCore is, on the other hand, a development kit for Android apps. The two libraries allow you to exploit the hardware components listed above, such as sensors, displays, cameras and others, available on board the various smartphones, tablets and head mounted devices. (Arena, 2022)

3.1.2 AR Field of use

With the rising popularity of augmented reality in recent years, it has been used in more and more new applications every day, which makes it hard to list all possible AR applications. However, we can group together some of the driving categories of this technology. (Shervin Minaee, 2022)

Education

According to the multimedia cognitive theory, people learn better, if multiple senses are targeted simultaneously. In this context, augmented reality has a great potential for better education because it can target multiple senses of people; an example of use in the education field is the Anatomy 4D application. A powerful AR tool that helps students study human anatomy by viewing and interacting with 3D models of the body, offering a detailed and immersive experience. (SÜNGER, 2019)

Health

A study conducted in April 2019 analyzed approximately 1300 publications from PubMed and Scopus. The results showed an increasing trend in the number of publications on AR in medicine, focusing on training and support. (Eckert M, 2019) . During the outbreak of the COVID-19 pandemic mentioned at the beginning of the section, Imperial College Healthcare NHS Trust in collaboration with Microsoft developed a mixed reality headset to protect frontline workers in contact with the Covid-19 virus: this allowed to reduce by 80% the time that healthcare workers spent in high-risk areas. It was also estimated that approximately 700 personal protective equipment were saved per ward per week.

Repairing and Maintenance

Another mostly used areas of augmented reality are repairing and maintenance services. Augmented reality is being used in repairing, maintenance, assembly, design, etc. from automobiles to advanced space vehicles (SÜNGER, 2019). In this field AR can guide the technician through what needs doing, in the preferred sequence. For example, it might start with removing covers and guards and then proceed to filter replacement and checking drain ports. AR is a valuable tool for supporting the adoption of predictive maintenance strategies.

With augmented reality, industrial maintenance can be performed faster and with fewer errors. That reduces mean time to repair, improves equipment availability and reduces unplanned downtime.

Specific benefits include:

- Faster problem diagnosis, with less time spent identifying relevant components and referring to diagrams
- Faster servicing and repair work
- Consistent servicing performance, regardless of who does it
- Less reliance on the experience of individual technicians
- Fewer errors in installation or servicing
- Enhanced training for new technicians
- Improved safety
- Real-time remote assistance

Ecosystems such as Vuforia Studio are widely used within the industrial world for the creation of monitoring procedures in the fields of automotive, manufacturing, life sciences and federal aerospace and defence (Figure 3).

Figure 3: Vuforia in automotive, manufacturing, life sciences and federal aerospace and defence fields of use - Source: Google Images

These tools allow you to pre-configure a virtual work environment that will be superimposed on a specific machine during maintenance. The system will be connected to real data coming from various sensors of the machinery itself to be able to control any maintenance sessions. (Advanced Technology Services, n.d.)

3.1.3 AR Applications in low-light conditions

As it was possible to understand from the description of the various components of an AR system (input, display and sensors), we could deduce that the performance of an AR application depends largely on how well the sensors are able to inform the software about where the device is located in the real space. For example, if we take into consideration the use of a smartphone, the main sensor that provides this information will certainly be the camera, together with any other depth sensors, if available. This means that to analyze the efficiency that an AR system can achieve today we can start by analyzing how well today's mobile devices are able to reproduce the image of the environment with good quality even in low light conditions.

Technologies to support photography under these conditions are called Depth-Sensing Technologies; these allow the camera to add depth perception, to improve focus, or communicate directly with the AR system.

Dual camera technology for example uses two cameras to simulate a binocular view and estimate the distance into the scene they are framing. This type of technology is one of the least expensive and found on almost all smartphones nowadays, but it remains nonetheless quite inaccurate in low-light situations.

Another noteworthy technique is named **Time-of-Flight (ToF)**; this uses an infrared sensor to estimate the time it takes for the signal emitted by the sensor to hit a framed object and return. This type of technology is more accurate than Dual Camera and is able to operate even in low light conditions.

The more precise of the last two named is **LiDAR (Light Detection and Ranging)**, a sophisticated technology that is entering more and more smartphones. Basically, LiDAR acts like radar, but it uses light instead of radio waves. It emits pulsed laser beams and measures the time it takes them to bounce back. This creates a detailed 3D map of the surrounding environment (Figure 4).

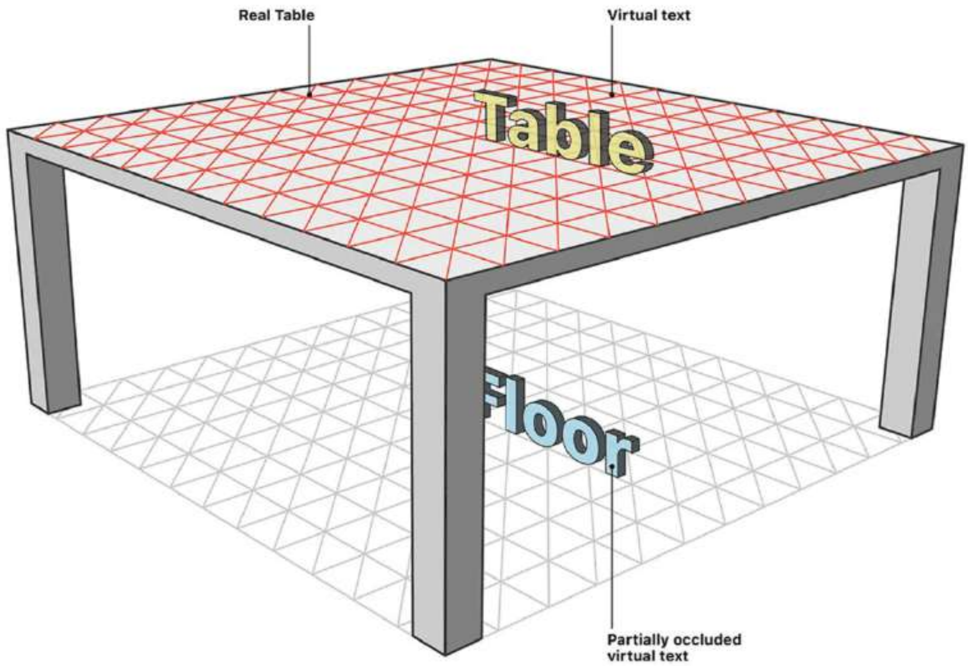

Figure 4: Applying virtual content in the physical world using depth information offered by LiDAR – Source: (Kao, 2024)

In particular, LiDAR offers two important advantages when using the camera:

- **Better Focus in Low Light:** Ever struggled to get your camera to focus in a dim room? LiDAR helps your camera find its focus quickly and accurately, even in challenging low-light conditions.
- **Sharper Night Mode Shots**: LiDAR helps your camera "see" better in the dark. It provides depth information that allows for better exposure and less noise in night mode photos.

These improvements in imaging along with the direct use of depth information makes augmented reality experiences more realistic and interactive. AR apps can use the depth information to seamlessly blend virtual objects with the real world.

This type of sensor is widely used in Apple phones and tablets, while it is not available in most Android devices.

(Kao, 2024)

3.2 Hardware requirements

The section aims to illustrate the hardware requirements considered in the development of the visualization tool. Different hardware devices for augmented and mixed reality will be shown, their pros and cons will be evaluated in a general context. Finally, placed within the context of real-world use within the maintenance pipeline.

3.2.1 Analysis of headset and tablet use

The state-of-the-art chapter defined the hardware features that an augmented reality device must include. The most commonly used devices that include these types of features are the smartphones we use every day, or larger tablets, which with their large display allow for an even more immersive experience.

Technological development in the field has also led to the construction of wearable visual aids, which allow the user to see virtual objects through glasses or headsets.

Below we look at some of the above hardware starting with the first category, smartphones and tablets. Let's consider Apple's top-of-the-line devices, which mount LiDAR devices mentioned in Chapter 3.13. These can calculate distances even in low-light conditions, which together with the use of multiple cameras allows the user to enjoy a more faithful experience.

- **iPhone 16 Pro Max**: one of apple's latest smartphones comes with very good specifications in terms of camera, display and tracking sensors. Here are some specifications below:
    - Display: Super Retina XDR display, 6.9-inch (diagonal) all-screen OLED display, 2868-by-1320-pixel resolution at 460 ppi
    - Camera: 48MP Fusion: 24 mm, ƒ/1.78 aperture, second-generation sensor-shift optical image stabilization, 100% Focus Pixels, support for super-high-resolution photos (24MP and 48MP).
    - Scanner LiDAR: can measure the distance to objects up to 5 meters, works in both indoor and outdoor environments, operating at nanosecond-sized speeds
    - Processors: 6-core CPU with 2 performance and 4 efficiency cores, 6-core GPU

(Apple Support, n.d.)

- **iPad Pro:** sold in multiple versions based on display size and resolution, 11 and 13 inches, the apple tablet features high-efficiency components, losing a few points on the camera, but without sacrificing excellent focal aperture and scan technologies.
    - Display: Ultra Retina XDR display, 13-inch, Tandem OLED, 2752-by-2064-pixel resolution at 264 ppi
    - Camera: 12MP Wide camera, ƒ/1.8 aperture, Auto image stabilization, TrueDepth Camera ƒ/2 aperture
    - Scanner LiDAR: can measure the distance to objects up to 5 meters, works in both indoor and outdoor environments, operating at nanosecond-sized speeds
    - Processors: Apple M4 chip, 10-core CPU with 4 performance cores and 6 efficiency cores, 10-core GPU

(Apple Support, n.d.)

Both devices present themselves as excellent solutions for hosting the visualization tool. The tablet has a camera with a lower resolution than the smartphone but has a larger display and a more powerful GPU. Higher resolution models can be used for better display of details.

Let us now analyze head-mounted devices. Virtual reality (VR) and augmented reality (AR) headset devices are wearable tools designed to provide users with immersive experiences by overlaying digital elements on top of the real world or by creating completely virtual environments.

Virtual reality (VR) headsets: these viewers immerse users in a three-dimensional digital environment that isolates them from the real world. Through stereoscopic displays and motion sensors, users can interact with simulated environments, creating immersive experiences in areas such as gaming, education, and professional simulation.

Augmented Reality (AR) headsets: these devices overlay digital information onto the real world, enhancing the user's interaction with their environment. Using cameras, sensors, and transparent displays, AR headsets integrate virtual elements into the user's visual experience, with applications in industrial maintenance, education, and design.

Mixed Reality (MR) headsets: These devices combine aspects of VR and AR, allowing users to interact with virtual and real elements simultaneously. Mixed reality enables a more integrated fusion of the physical and digital worlds, providing enhanced interactive experiences.

(Trofa, 2023)

Most headsets today offer all three possibilities, allowing the user to switch between a fully immersive mode and a see- through mode. VR viewers in fact, use real-time six degrees of freedom (DOF) and visual simultaneous localization and mapping (SLAM) to receive real-world information, such as 3D coordinates and object recognition. Some VR headsets incorporate images from monochrome cameras and extend the capabilities of **video see-through (VST)** to mixed reality (MR) (Anon, VR Headset See-through Tech to Help in MR Development, 2022)

The see-through technology used in current VR headsets is called VST, while the see-through technology used in augmented reality (AR) devices is called the **optical see-through (OST).**

An **optical head-mounted display** (**OHMD**) is a wearable device that has the capability of reflecting projected images as well as allowing the user to see through it. (Anon, Optical head-mounted display, 2022). The OST function of AR devices allows virtual images to be superimposed directly over real-world vision (Figure 5). The resolution of real-world images in AR devices is significantly higher than in VR headsets

Figure 5: OHMD Hololens glasses – Source: (Ashley, 2017)

(Anon, 2022)

After an introduction to the available technologies, we look at some specific products available.

- **Microsoft Hololens 2**: is a mixed reality OST device with no cables or external components in the way and is self-contained with WiFi connectivity. Its technical features and various functions have been improved from its predecessor, from the width of the field of view to the resolution and a new tracking technology (Figure 6). The device offers a 2K 3:2 aspect ratio display with a resolution of 1440x936, providing clear and immersive visuals (Anon, HoloLens 2, 2021)

Figure 6: Microsoft Hololens 2 POC – Source: Google Images

- **The Apple Vision Pro** is a cutting-edge spatial computer with two micro-OLED displays offering a combined 23 million pixels (Figure 7). It runs on Apple's M2 chip, paired with a dedicated R1 chip for real-time sensor data processing, ensuring seamless interactions with minimal latency. The device supports multiple refresh rates (90Hz, 96Hz, 100Hz) and features advanced sensors, including stereoscopic cameras, LiDAR, and eye-tracking technology. (Apple Support, 2023). This device offers ultra-high-resolution images allowing less and less distinction between the real and virtual worlds, making the experience even more immersive

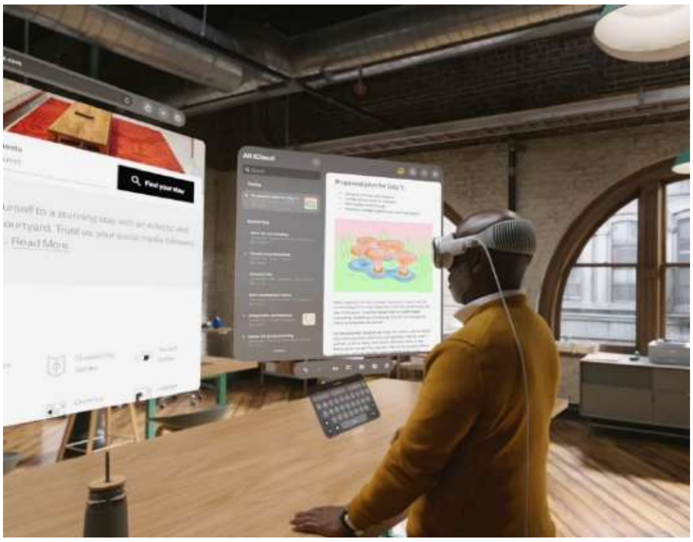

Figure 7: Apple Vision Pro POC – Source: (Sahah, 2023)

- **The Meta Quest 3** is a standalone virtual reality headset released on October 10, 2023 (Figure 8). The headset offers a resolution of 2064x2208 pixels per eye, utilizing RGB-stripe LCD+ panels with refresh rates between 90Hz and 120Hz. The Quest 3 is equipped with four tracking cameras, two 4-megapixel color cameras, and a depth sensor, enabling a colored pass-through mode for mixed-reality experiences. (Meta, 2015)

Figure 8: Meta Quest 3 – Source: (Meta, 2015)

3.2.2 Real use case

Having analyzed the devices of interest for our use case, let's take a closer look at the actual use case to see which is the best. As will be specified later, the AR Visualization Tool will be embedded within the maintenance process before and during grinding and welding operations.

To properly evaluate the right hardware device to use, it is necessary to consider its usability considering the circumstances in the field. We first take into consideration the operators' use of safety devices. These certainly include a welding mask with automatic darkening and equipped with a respirator to filter ambient air while working. In Figure 9 we can see some references.

Figure 9: Welding and grinding security devices – Source: (met 3MTM, 2016)

At this point we can make some observations about using a wearable device such as the headsets we listed above. First, if we were using an OST or VST viewer, it would have to be worn under the mask. This would obstruct the cameras that are usually placed at the ends of the viewer to provide a peripheral view, which would degrade the quality of tracking of the device.

Figure 10: M3 Speedglas Welding Helmet – Source: (met 3MTM, 2016)

In addition, the size of the products available on the market nowadays must be taken into consideration. These should be able to be worn inside the mask, which despite having modest dimensions, could not allow such a device to enter, see Figure 10.

We would also add that performance products such as Apple Vision Pro and Meta Quest 3 have cooling systems for their CPUs and GPUs that dissipate quite a bit of heat outside the visor. This would dramatically increase the temperature inside the mask, going to interfere with the airflow created by the respirator. (met 3MTM, 2025) (met 3MTM, 2016)

Having made these considerations, we can say that the usability of a wearable device, unless specifically integrated with a protective device, is complicated.

In the specific use case, we can consider the best device that does not obstruct the use of security masks, a tablet or smartphone. According to those mentioned, we may prefer to use iPad Pro, which provides a good compromise between performance and screen width for a good experience. This type of device can be easily handled during maintenance phases to interact with the digital twin and display the heatmap on the actual model. The device may possibly be protected by an industrial-use cover that goes to protect the device in all its parts.

3.3 Software requirements

The current chapter focuses on describing the software requirements of the application, in particular:

- **Use cases:** description of the utilization process within the maintenance pipeline.
- **High-level system architecture:** High-level architecture of the main components of the application
- Some **identified software solutions and tools,** considering the target hardware while keeping fallback options available in case the chosen ones encounter critical issues during development and/or deployment.

3.3.1 Description of the field process: use cases

We first show where the heatmap visualization application is placed within the maintenance pipeline, specifically the XCROSS project's workflow and data exchange. The application will be used by the operator throughout the maintenance session in order to manage the entire maintenance session. The diagram below provides a high-level overview of how the visualization and session management application interacts with external actions and data, see Figure 11. It also illustrates its connections with other steps in the process.

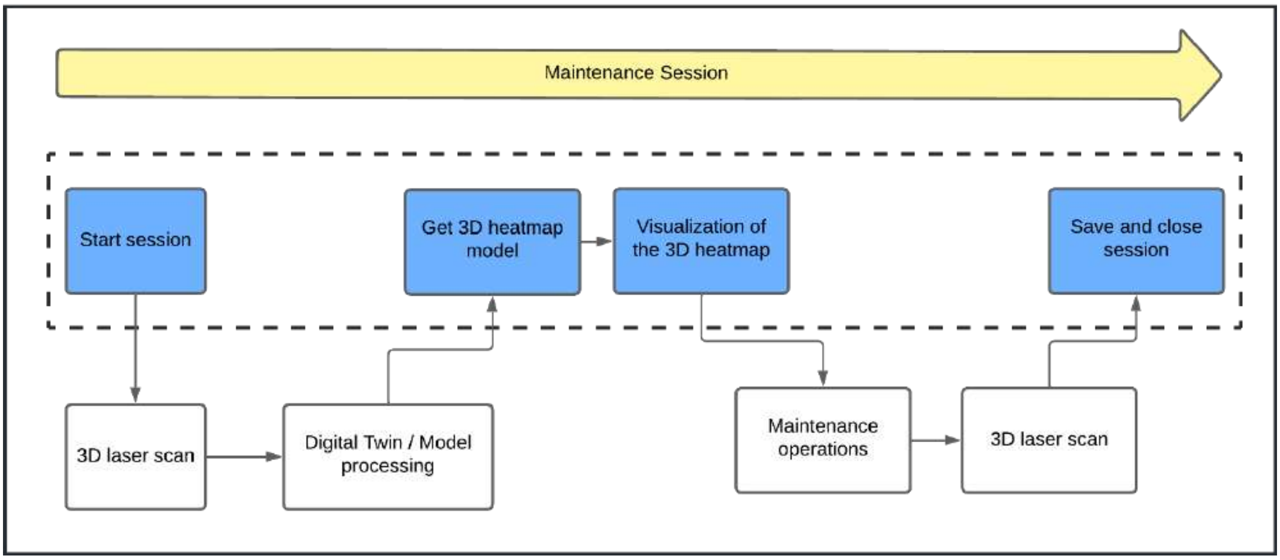

Figure 11: Maintenance Session graphic workflow

The blue blocks inside the hatched box represent steps in the application's workflow for session management and model visualization. They also include external components (backend) required for proper operation, data exchange, and storage.

As you can see, the application will be used before, during and after the maintenance session. More specifically:

- **Start session**: operations required to open a maintenance session, providing the necessary data; here a new session will be created in the backend, with the data provided by the user.
- **3D laser scan**: external operation of crossing 3D scanning and sending data to the digital twin.
- **Digital Twin / Model processing**: external operation for the manipulation of data received from 3D scanning and generation of output, the heatmap model.
- **Get 3D heatmap model**: once the output has been generated in the previous steps, the backend must provide the application with the 3D model to be displayed in AR.
- **Visualization of the 3D heatmap**: the 3D heatmap is displayed in the real-world using AR. The user will be able to place the model superimposed on the actual model.
- **Maintenance operations**: necessary actions performed by the operator on the crossing.
- **(2nd) 3D laser scan**: external operation of crossing 3D scanning after maintenance operations. A new 3D scan of the newly repaired model is performed.
- **Save and close session**: the last 3D scan (result of maintenance operations) is saved together with other information on the backend side session and the session is closed and made available for eventual review.

3.3.1.1 Flow of a maintenance session using the visualization tool

Starting from the simplified pipeline described above, we can list in more detail the steps that the user takes while using the application to better identify its features. In turn, these will help us identify more technical requirements on data exchange and components such as the backend.

Below we describe a path of use of the tool by an operator, referring to him as the user of the application:

1. The user starts the application to begin a new maintenance session
2. The user enters session profiling data (location, date - time, useful notes).
3. The user starts the maintenance session (Note: session data may be viewed and edited throughout its duration)
4. User performs external operation: 3D scanning and processing with digital twin
5. The application downloads the 3D model of the crossing, which has just been scanned and processed by the digital twin, to the local device.
6. At this point the user can:
    a. Begin the process of aligning and visualizing the model in augmented reality
        i. The application will start showing the user images generated by the device's camera

ii. The user is asked to perform certain steps to take landmarks within the real world (based on the alignment method chosen during development)
iii. If the process is successful, the user will be able to view the heatmap within the real world, superimposed so that it sits on top of the real model
iv. If the process is unsuccessful, and the model turns out not perfectly aligned, the user can perform small manual shifts, or decide to reposition the model by re-executing the same steps shown above

b. View the processed model in the 3D application, viewing it from different angles, zooming and/or rotating
c. Edit session notes, adding photos of the defect using the camera of the device

7. The user performs maintenance operations, having the possibility to view the heatmap during these operations
8. The user performs a second 3D scan of the model after maintenance
9. The application downloads the newly scanned model
10. The user can view the model in the application, as he did for the first model (6.b)
11. The user can add photos of post-maintenance
12. When the session is over, the user saves and closes the maintenance session

The saved sessions can be viewed later by the user. More specifically, for each closed session, the user will be able to:

- View session information and metadata
- View photos taken on the spot for the session
- View 3D heatmap models (model before maintenance) and the later scanned model

3.3.1.2 Data exchange

Once we have defined what the user requirements are, we can define how data is exchanged inside and outside the system. As you can see from the user requirements, the application will communicate with its **backend** to:

- Create a new session
- Download 3D models (heatmap and scan)
- Save current session information

Based on what has already been determined, a **session** can be identified by:

- **Unique Id**: self-generated unique identifier to identify the session
- **Place of session**: may be textual or georeferenced
- **Session start date time**: autogenerated by the system at the time of start
- **Photo**: of the defect and/or post maintenance work
- **A 3D heatmap model**
- **A 3D model of the post-maintenance scan**
- **Notes**: text comments on the session
- **End session date time**: autogenerated by the system at the closing of the session

The data for each session will be saved on the device used on the site (ideally a laptop or other type of PC), and handled by the backend which will take care of receiving the 3D data as output from external actions within the pipeline and share such data with the application running on the mobile device. Similarly, the backend will allow to read local data on request from the application and save them at the time of closing a session.

The data will be made available on a second service network whenever internet coverage is available, so that it can be reached on all other devices.

Session files will be provided in an exportable format so that data (such as photos) can be integrated with third-party systems.

3.3.2 System architecture

We can now identify a high-level architecture of the system. Based on what has been said so far, we can imagine a draft of the physical architecture simplified in this way:

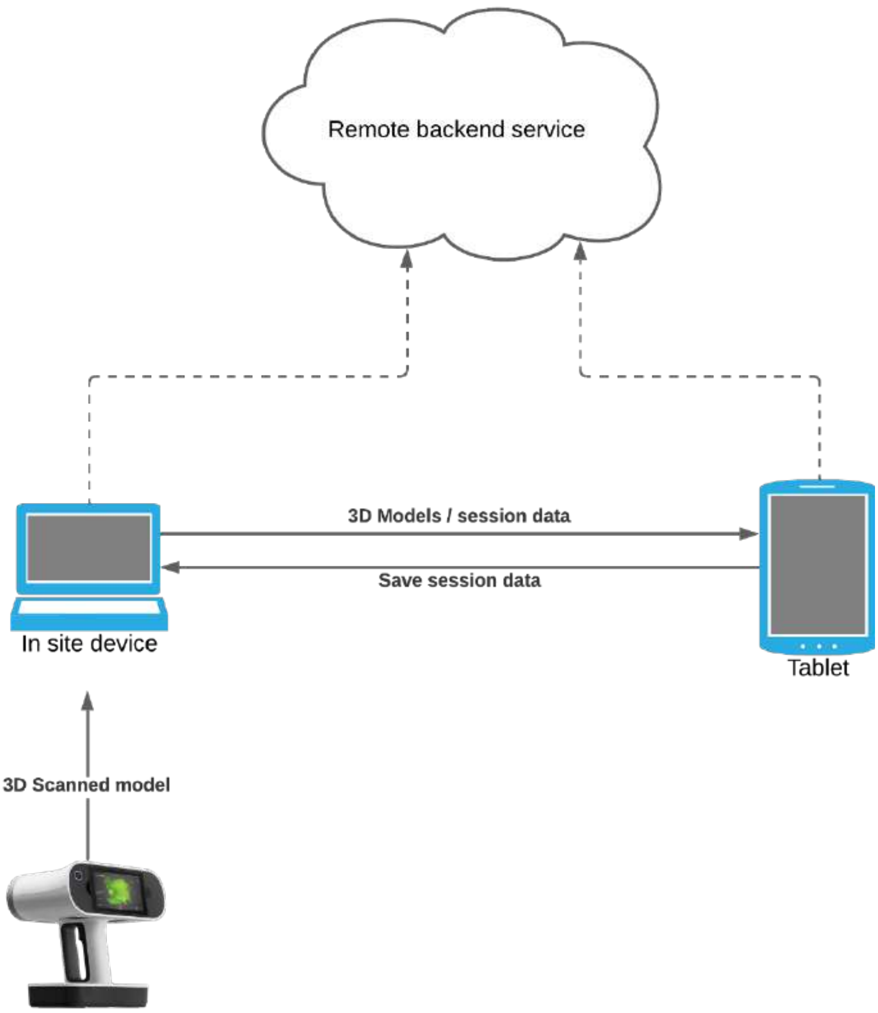

Figure 12: Physical architecture

The device used in site will be directly connected to the tablet for data exchange if it is not possible to have a sufficiently powerful internet coverage for data exchange (especially for 3D models). Otherwise, both devices will exchange their data using a remote backend service (Figure 12).

The logical components of the architecture are shown in the Figure 13 below:

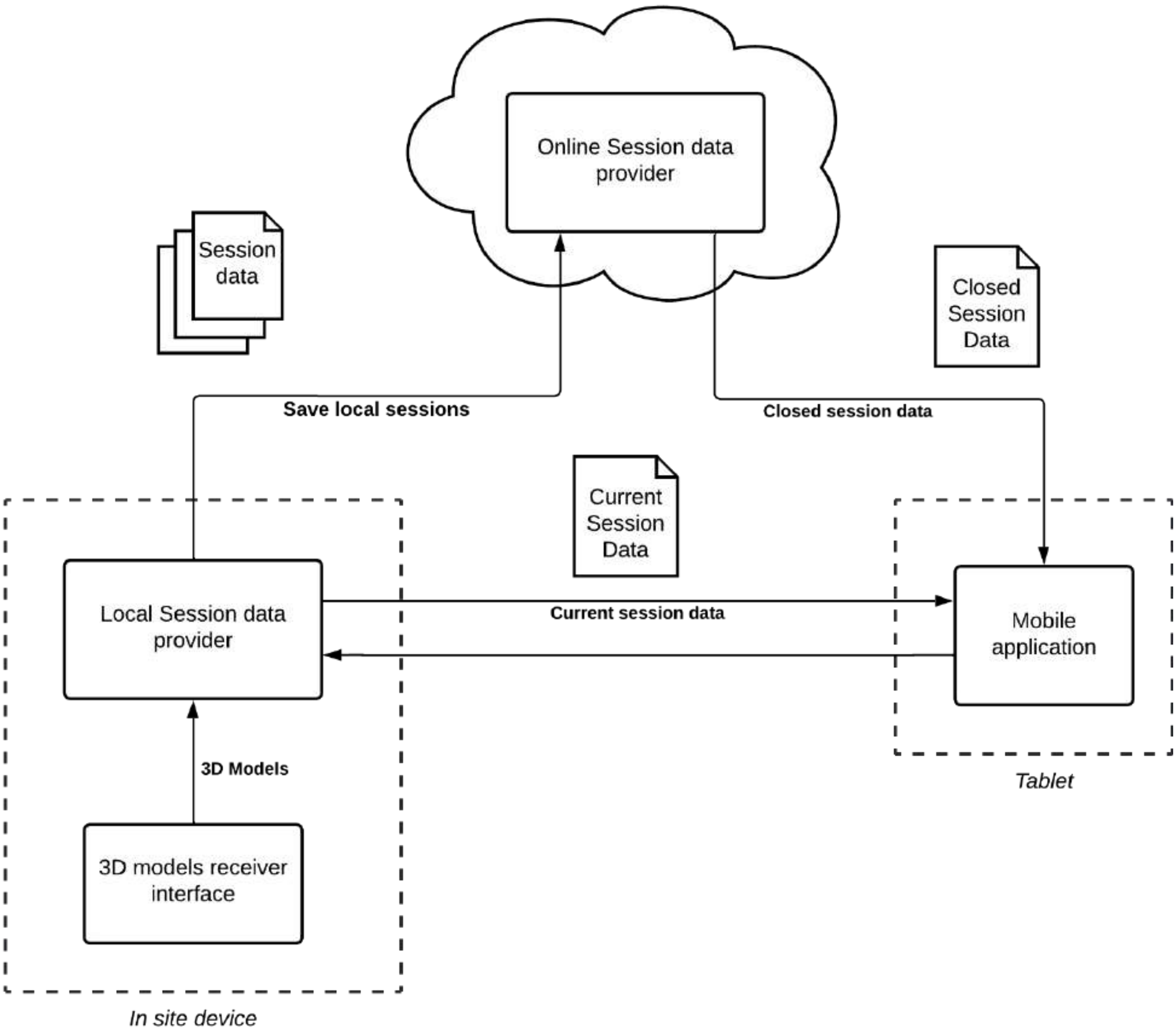

Figure 13: Logical Architecture

More specifically, the architecture is composed of the following logical components:

- **3D models receiver interface**: component that interfaces with the scanning services and process to receive the 3D models to be displayed within the application and to be saved in session data. The interface will be implemented according to the needs identified by the technologies that provide the output in accordance with the application requirements
- **Local session data provider**: takes care of receiving the 3D models from the interface described above, and providing this data to the mobile application. The component also interfaces with the online service in case of loading locally saved procedures during offline sessions
- **Online session data provider**: is the online version of the component described above. It interfaces with the mobile application during an online session to provide and save necessary data
- **Mobile application**: performs the features identified in user requirements and use cases

3.3.3 Development tools

As mentioned at the beginning of this section, we can show some implementation choices that are consistent with the design choices. However, there is flexibility to consider alternative solutions. These alternatives do not compromise the overall implementation during development. One of the most important is that of the graphics engine for building the mobile application, in this case **Unity**.

Unity is a cross-platform game engine developed by Unity Technologies, first announced and released in June 2005 at the Apple Worldwide Developers Conference as a Mac OS X game engine. Over the years, it has evolved to support a wide array of platforms, including mobile, desktop, consoles, and virtual reality (Wikipedia, 2025). This extensive platform support allows developers to create a game once and deploy it seamlessly across multiple platforms, ensuring a consistent player experience without the need to rewrite code for each platform. (Unity, n.d.)

The engine architecture allows the application to be developed almost independently of the target platform, using their native features for each of them. A graph of the architecture for XR application development on Unity can be seen in the Figure 14: Unity XR Development Tool Architecture – Source: . As we can see, the engine provides a set of tools that enables the development of powerful AR/VR/MR applications by providing high-level interfaces that communicate with provider implementations based on the platform on which the application runs. These include the two libraries mentioned in the state of the art, particularly ARKit, which will be used by the iPad Pro device to take advantage of its AR capabilities.

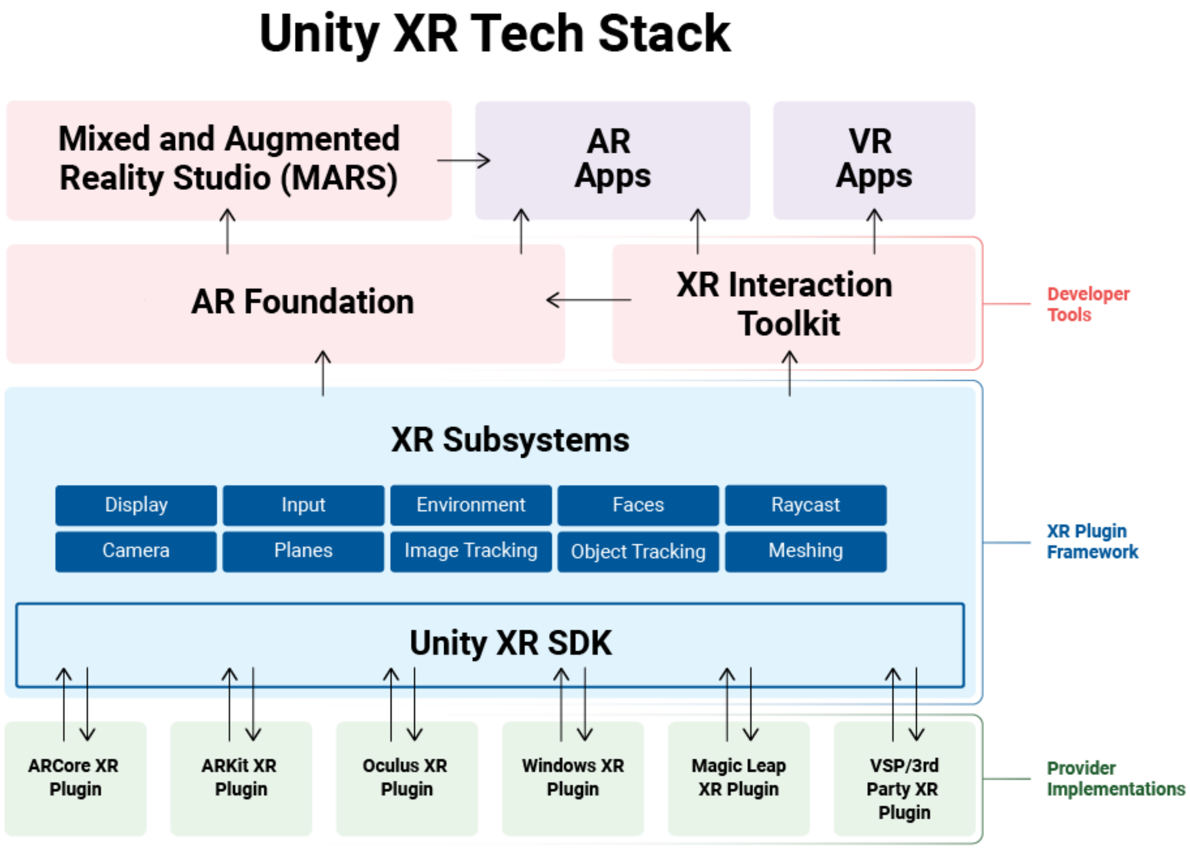

Figure 14: Unity XR Development Tool Architecture – Source: (Unity, 2025)

Providers also include libraries for developing immersive applications on head mounted devices. This will eventually allow the tool to evolve if it is possible to use those devices based on the considerations made in chapter 3.2.2. (Unity, 2025)

4 Conclusions

This deliverable focuses on defining the hardware and software requirements for the AR viewer within the XCROSS project. The main objective was to provide the operator with a tool to view heatmaps of the crossing model being repaired, which supports their work and helps them focus on the sections of the crossing requiring maintenance. The application is based on 3D scanning and digital twin processing, allowing the user to input maintenance session data and view past sessions.

The methodology followed in this work included analyzing the state of the art of AR technologies, with a special focus on their use in low-light conditions, and then reviewing available hardware solutions. We compared the use of tablets versus augmented reality headsets and ultimately selected a tablet (iPad Pro) due to its suitability for the task and compatibility with the operator's safety equipment. Once the hardware was selected, the software requirements and use cases were outlined, including the flow of tasks an operator would perform using the AR application. This led to the design of a logical architecture, encompassing both internal and external components of the application, including backend components for data storage and exchange. Furthermore, we chose Unity as the development engine due to its cross-platform nature and flexibility.

The findings show that a tablet provides the best usability and compatibility for the project, avoiding the limitations imposed by AR headsets. The identified use cases and software requirements have helped in shaping the AR application's architecture, ensuring its effectiveness in real-world maintenance tasks. However, one limitation lies in the challenges related to implementing AR in complex environments, particularly when operating in extreme lighting conditions or using less powerful devices.

The best solution identified, a hand-held tablet, provides a practical and effective approach for the project, aligning with both the operational needs and technical requirements.

References

Advanced Technology Services. (n.d.). *Augmented Reality in Maintenance: How is it Helping?* Retrieved from https://www.advancedtech.com/blog/augmented-reality-in-maintenance/

Anon. (2021, May 22). *HoloLens 2*. Retrieved from https://en.wikipedia.org/wiki/HoloLens_2

Anon. (2022, February 13). *Optical head-mounted display*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Optical_head-mounted_display

Anon. (2022). *VR Headset See-through Tech to Help in MR Development*. Retrieved from Counterpointresearch.com: https://www.counterpointresearch.com/insights/vr-headset-see-tech-help-mr-development/

Apple Support. (2023). *Apple Vision Pro*. Retrieved from https://www.apple.com/apple-vision-pro/specs/

Apple Support. (n.d.). *iPad Pro - Technical Specifications*. Retrieved from https://www.apple.com/ipad-pro/specs/

Apple Support. (n.d.). *iPhone 15 Pro Max - Technical Specifications*. Retrieved from https://support.apple.com/en-us/111828

Arena, F. &. (2022). *An Overview of Augmented Reality.* Enna: Università degli Studi di Enna Kore.

Ashley, J. (. (2017, March 13). *How HoloLens Displays Work*. Retrieved from The Imaginative Universal: https://www.imaginativeuniversal.com/blog/2015/10/18/how-hololens-displays-work/

Bourdot, P. (2018). Augmented Reality, Virtual Reality, and Computer Graphics: 5th International Conference. *AVR 2018.* Otranto.

Eckert M, V. J. (2019). *Augmented Reality in Medicine: Systematic and Bibliographic Review.* JMIR Mhealth Uhealth.

Kao, J. (2024, October 28). *LiDAR in smartphone photography*. Retrieved from 9meters: https://9meters.com/technology/lidar-in-smartphone-photography

Kurzweil, R. (2003). *Foreword to virtual humans.*

met 3MTM. (2016). *Heavy-Duty Welding Helmet G5-01 | 3M United Kingdom*. Retrieved from 3m.co.uk: https://www.3m.co.uk/3M/en_GB/worker-health-safety-uk/personal-protective-equipment/new-featured-products/3m-speedglas-heavy-duty-welding-helmet-g5-01/

met 3MTM. (2025). *SpeedglasTM G5-01-serie lashelm | 3M Nederland. (2025)*. Retrieved from 3mnederland.nl: https://www.3mnederland.nl/3M/nl_NL/p/d/b5005081009/

Meta. (2015). *Meta Quest 3 512GB*. Retrieved from Meta.com: https://www.meta.com/it/en/quest/quest-3/?srsltid=AfmBOopxFFb0Bk-usnkF3-

NPy9QZ2UNYBBWmDmfB2oICxHLIhh2FoPXk

Potuck, M. (2017, September 1). *Apple announces ARKit for iOS 11*. Retrieved from 9to5Mac: https://9to5mac.com/2017/06/05/apple-announces-arkit-for-ios-11/

Shervin Minaee, X. L. (2022). *Modern Augmented Reality: Applications, Trends, and Future Directions.* Guangzhou: Sun Yat-sen University.

Srl, H. (n.d.). (n.d.). *Differenze tra XR, VR, MR e AR*. Retrieved from Blog Applied: https://www.applied.it/it/le-differenze-tra-xr-vr-mr-e-ar-nws-36

SÜNGER, İ. &. (2019). *Augmented Reality: Historical Development and Area of Usage.* Journal of Educational Technology and Online Learning.

Trofa, F. L. (2023, January 5). *Realtà virtuale: cos'è, a cosa serve, i visori e le applicazioni di business*. Retrieved from Tech4Future: https://tech4future.info/realta-virtuale-visori-applicazioni

Unity. (2025). *Unity - Manual: XR architecture*. Retrieved from Unity Documentation: https://docs.unity3d.com/2022.1/Documentation/Manual/XRPluginArchitecture.html

Unity. (n.d.). *Maximize multiplatform game development | Unity*. Retrieved from Unity.com: https://unity.com/solutions/multiplatform

Wikipedia. (2025, March 20). *Unity (game engine)*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Unity_%28game_engine%29

Appendices