POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

# Supporting group-work in a mixed reality educational experience

Tesi di Laurea Magistrale in
Computer Science Engineering - Ingegneria Informatica

Author: **Leonardo Scapolan**

Student ID: 964719
Advisor: Prof. Nicoletta DI BLAS
Co-advisors: Prof. Marco GRIBAUDO, Matteo VALORIANI
Academic Year: 2023-24

# Abstract

For decades, Mixed Reality (MR) has been gaining attention across various sectors, particularly in the realm of education. Its primary objective is to replicate objects and experiences that may be difficult, hazardous, or simply unattainable within the confines of a traditional classroom, like machinery, archeological remains, anatomy... Interestingly, while the development of 3D content has flourished, there has been a noticeable lack of parallel emphasis on the educational activities that, when combined with this content, create a comprehensive educational experience. This thesis work is part of a larger project called Mixed Reality in Education (MIRE), by Vodafone Italy, a company (FifthIngenium), HOC-LAB at Polimi, and a group of Alta Scuola Politecnica (ASP). MIRE aims to bridge this gap, by exploring the development of educational activities in MR, with a specific emphasis on supporting groups engaged in collaborative projects. The specific focus of this thesis work has been the design, implementation, and testing of one of the activities within his scenario: collaborative brainstorming. Functional requirements stipulate that the application should support discussions among multiple individuals (students) with a moderator (the instructor). As such, an application based on the concept of a "word cloud" was developed, where discussed elements are connected to spheres, with their relative size indicating the degree of agreement with that concept among participants. The application was implemented in Unity and deployed on several MR devices available in the market. The Word Cloud activity underwent usability and perceived effectiveness testing in two real-world contexts. Notably, the latest round of testing involved students and teachers in a real school context. The test results, in addition to confirming the usability of the application, indicated a positive reception by the intended users and a favorable perception of its utility. Since it's an application designed to support learning, future developments include a broader and longer-term assessment of its educational impact. The results of the project were published as full paper at a major international conference.

**Keywords:** Mixed Reality, Augmented Reality, Collaboration, Devices, Education, Evaluation, Extended Reality, Word Cloud

# Abstract in lingua italiana

Per decenni, la Realtà Mista (MR) si è diffusa in diversi settori, tra cui la formazione. È stata principalmente usata per replicare una serie di oggetti che potrebbero risultare pericolosi o impossibili da avere all'interno di ambienti scolastici, come macchinari, reperti archeologici, elementi anatomici... Tuttavia, mentre lo sviluppo di contenuti 3D è fiorito notevolmente, non c'è stata molta enfasi sullo sviluppo di attività didattiche che, combinate con i contenuti, creano un esperienza didattica completa. Il lavoro svolto per questa tesi è parte di un progetto più grande, intitolato Mixed Reality in Education (MIRE), che mira a colmare questo divario. L'obbiettivo principale di questo progetto è esplorare lo sviluppo di attività didattiche in MR, con un taglio specifico sul supportare gruppi impegnati in progetti collaborativi. Il focus specifico di questo lavoro di tesi è stato il design, l'implementazione e il test di una delle attività in questo scenario: il brainstorming collaborativo. Il requisito funzionale era supportare la discussione tra diversi individui (studenti) con un moderatore (l'insegnante). È stata sviluppata un'applicazione basata sul concetto di "word cloud", nella quale i concetti di cui si sta discutendo sono associati a delle sfere, le cui dimensioni relative rappresentano il grado di accordo sui concetti da parte dei partecipanti. L'applicazione è stata implementata in Unity e rilasciata per diversi dispositivi di MR disponibili sul mercato. L'attività Word Cloud è stata poi testata dal punto di vista dell'usabilità e dell'efficacia percepita in due contesti reali. In particolare, l'ultimo test è stato effettuato con studenti e insegnanti in un vero contesto scolastico. I risultati ottenuti, oltre ad aver confermato l'usabilità dell'applicazione, hanno indicato una accoglienza positiva da parte degli utenti finali e una percezione favorevole della sua utilità. Trattandosi di una applicazione creata per supportare l'apprendimento, gli sviluppi futuri includono una valutazione più ampia e a lungo termine del suo impatto educativo. I risultati sono stati presentati come "full paper" a una conferenza internazionale.

**Parole chiave:** Realtà Mista, Realtà Aumentata, Collaborazione, Dispositivi, Istruzione, Valutazione, Realtà Estesa, Word Cloud

# Contents

# 1 | Introduction

MR is a field in computer science that, especially in the last decade, has started to become more and more relevant. It also became widely known among people without any computer science background thanks to several products released in the market. This was made possible by the advancement in technology, allowing MR devices to become affordable even by the masses.

The increasingly widespread introduction of Mixed Reality (MR) in the educational field, in particular, poses specific challenges. The sector has worked extensively on the production of educational content, but it has mainly focused on the virtual replication of "objects," such as large, expensive, or dangerous machinery to teach their use, anatomical parts (e.g., a human heart), or objects and buildings from remote historical epochs for archaeological studies. However, a learning experience is not only about content but is also composed of educational activities that, when combined with this content, create a comprehensive educational experience. To draw an analogy, it's akin to selecting the most suitable textbooks for a school lesson but neglecting to consider the activities in which students are expected to engage. Undoubtedly, any meaningful educational experience hinges on the seamless integration of content and activities. In this aspect, the literature reveals the existence of a gap. While many efforts have been made in creating content, very few have been made in developing teaching activities that, when combined with the former (or even with "traditional" content), can lead to comprehensive learning experiences in MR.

This thesis aims to bridge this gap by focusing on a particularly interesting teaching activity: collaborative brainstorming. This activity allows for the development of higher-order learning skills, such as critical thinking, communication, and argumentation abilities. It can be introduced in an educational context either in combination with MR content or with "traditional" content (learned from books, lectures, and handouts), making it highly flexible.

Therefore, the objective of this thesis was the development of an MR application to support collaborative brainstorming, covering all stages of development from design to implementation and testing in two real-world contexts.

The activity was developed as part of the MIRE project. Its outcomes were presented as a full paper titled "Supporting Project Work with Mixed Reality" [4] at the international EdMedia conference held in Vienna in July 2023. Furthermore, these findings contributed to the development of an educational product by the stakeholder companies, catering to schools, universities, and the corporate sector.

# 2 | State of the Art

Extended Reality (XR) is a technology that, especially in the last 10 years, has found its way into being utilized in several fields. The one that I'm most interested in for this thesis is Education, especially for what concerns applications that focus on collaboration. In the following sections, I will study the current state of XR, primarily in those fields, focusing both on its strengths, while also addressing its shortcomings and introducing how I plan to overcome them.

## 2.1. Extended Reality

XR, also often referred to as MR, is a concept that has been discussed for several decades now, and, especially in the last few years, it has become something that people involved in the world of computer science need to know. It also became a topic of discussion in recent years among the general public, most of whom have already probably come into contact with or even used a system that employs XR. Some notable examples are previews of products in Augmented Reality (AR) that are starting to pop up in online stores, virtual reality Virtual Reality (VR) gaming thanks to popular and affordable VR headsets, museums, and artistic sites with exhibitions empowered by the use of MR. Another field where XR is taking hold is the medical one, where it is used for example to train surgeons.

XR is an "umbrella term" that refers to a class of technologies that aims at extending the physical world with digital enhancements. A widely accepted taxonomy of XR was proposed in 1994 by Paul Milgram and Fumio Kishino in their publication "A Taxonomy of Mixed Reality Visual Displays" [15], with the concept of a Virtuality Continuum (VC).

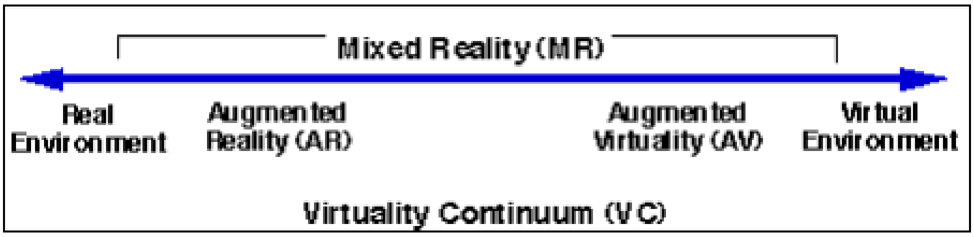


Figure 2.1: Virtuality Continuum (VC) representation from [15]

The VC is being represented by Milgram and Kishino as a line: on one side they place the real environment consisting of only real objects, including both what would be seen without the use of any device, but also reproductions of real scenes on digital displays recorded from a camera for example. On the other side, they place the virtual environment completely made up of virtual objects, examples of which are computer simulation or traditional video games. Everything in between is considered MR which in turn can be divided into several subgroups based on the proportion in which they mix real and virtual environments.

Going in order from the first to the second, they begin with Augmented Reality (AR), which refers to representations of real environments where few virtual objects are inserted. This technology has become quite commonly used both in commercial environments, with product previews as 3D models that can then be viewed in the real environment, and in the entertainment industry with video games like Pokemon GO, where the digital creatures could be seen immersed in the environment captured through the smartphone camera.


Figure 2.2: Examples of AR

Augmented Virtuality (AV) instead is kind of the opposite: an almost completely computer-created environment is enriched by the presence of real objects. This is also a widely used technology that is especially exploited by the cinema and television industry as an example. The technique used is the employment of a green screen in front of which people and objects are positions, that then get filled either in post-production or directly in the live feed with the desired virtual environment. It is often used during weather forecasts and in the shooting of fantasy and science fiction films.


Figure 2.3: AV used during a weather forecast

Another insightful definition for Extended Reality (XR) can be seen from Microsoft in their online article "What is mixed reality?" [14]. Here they define MR as the combination of Human, Computer, and the Environment, in contrast with conventional reality, which is just the interaction between the first and the third. The possibility to make Computers and the Environment interact is made possible by a perception algorithm that through the data collected by several sensors can recreate a digital model of the physical environment. Lastly, the possibility to use more advanced human-computer interaction paradigms like head and body tracking, hand recognition, and speech recognition is what truly makes the combination of the three elements into a fully-fledged MR experience.

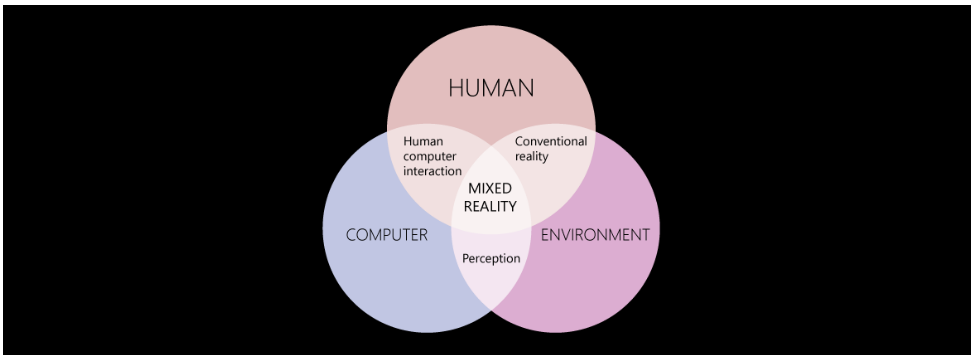


Figure 2.4: The interactions between computers, humans, and environments from [14]

Of all the possible applications of XR, the ones I'm most interested in, and that I will discuss in greater detail in the following sections, are XR in education and XR as a tool to support project groups.

## 2.2. Extended Reality in Education

Several applications of Extended Reality for education have been studied, developed, and reviewed over the years. In 2021 Juan Garzón published an article "An Overview of Twenty-Five Years of Augmented Reality in Education" [10] in a special issue of the journal "Multimodal Technologies and Interaction" in which, after a review of several other publications on the topic, gives a summary of the major trends that characterized AR in education from 1995, earlier year in which he was able to find an Augmented Reality application exclusively targeted to the world of education, to 2020. He identified three generations of AR in education: the first was characterized by highly specialized application, almost completely targeted to Science, Technology, Engineering and Math (STEM) subjects. One of their main limitations was usability since most of them were complex. This caused overexertion of the users and the high cognitive requirement hindered the learning process. The other big limitation was cost, coming from the high price of the hardware technologies (be it head-mounted, heads-up, or handheld displays), the need for specialized maintenance, and the skills required to develop content. Altogether this meant that this kind of application was not diffused, since most institutions deemed the cost not worth the benefits of the technology. Nevertheless, in the same years, Garzón was able to find some studies reaffirming the benefit of employing AR in education mostly coming from the ability to support interaction between the real and virtual world. The second generation, starting from 2010, was characterized by the resolution of the two main critical issues in the first generation (cost and usability) that I previously discussed. This was mainly achieved thanks to the ability to deploy AR applications to mobile devices, an already diffused and relatively cheap hardware support, and the release of software development kits and libraries to simplify the development of AR, allowing to save time and money that can then be redirected to the improvement of usability and user experience. The number of systematic reviews on the topic also started to increase, mostly focusing on the effect that AR has on the world of education, the main benefits, the bigger challenges and limitations, and finally the direction that they believed research and development should focus on in the future. Finally, the last generation, which is the one that started in 2020, and that we are experiencing now, was enabled according to Garzón by three main innovations. First, there was the entrance into the market of new and powerful smart glasses like Hololens and Oculus Rift. Second, there was the introduction in 2017 of WebAR, allowing users to experience AR applications directly from the browser, without having to download and install dedicated apps on their devices. Finally the integration with Artificial Intelligence (AI) allowed to create more immersive experiences utilizing algorithms capable of recognizing physical space, user hands, and voices.

Another important topic that Garzón's article touches on is the main issues that one may find in AR for education. These are:

- accessibility, as in ease of use of applications from everyone, especially with people presenting some kind of disability;
- usability, mostly caused by the involvement of multiple senses, affecting user attention and by consequence usability;
- lack of dissemination caused by the fact that applications must be often downloaded and kept on devices, and by the fact that they often support only a device, or just a few;
- pedagogical approaches in the sense that most of the AR applications focus on content, lacking any pedagogical structure.

Lastly, Garzón showed the trend of the number of studies published on the Web of Science (WoS) on the topic of AR in education over the 25 years.

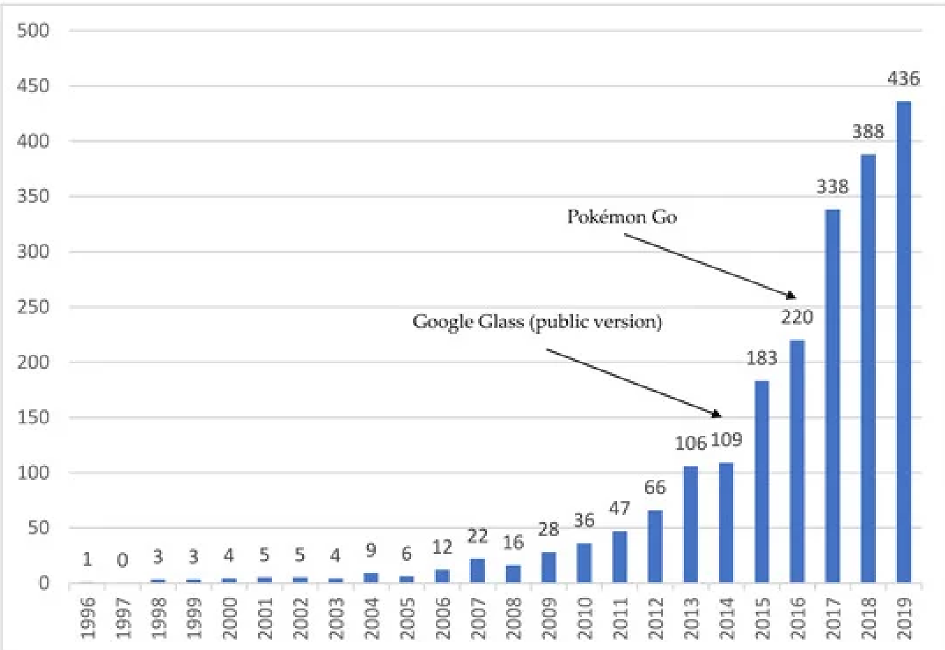


Figure 2.5: Number of studies of AR in education per year in the WoS from [10]

As can be seen, after a stale first period coinciding with the first generation that was just discussed, from 2010 there's an exponential growth of publications on the topic. Two notable jumps can be seen in the graph, the first is attributed to the public release of Google Glasses in 2014, while the second is the release of Pokemon GO in 2016. "These two events brought AR to prominence technologies, attracting many developers worldwide to create AR applications for education." [10]

Focusing on a narrower, but still recent time window, Chen et al.[7] in 2017 compiled an extensive review of 55 studies between 2011 and 2016 on the topic of AR in education. Several are the findings on the topic brought up in the publication. As a starter, they showed, just like Garzón, how the number of documents and articles on the topic started growing each year, indicating an increase in interest in the field. They also found that most of the studies were targeted towards all levels of education, from primary schools to Bachelor, with more focus on higher education, and that AR was mostly used "for motivating students, explaining topics, adding information [...]. It seems possible that AR has been applied in settings with this target group to improve the educational experience of the students and motivate engagement..." [7]. Quite interesting is the distribution in subjects where AR has been applied: the majority (55%) of the studies were targeted towards STEM, where it was deemed effective because it allowed to visualize complex concepts and situations that are not easily observable in the real world either because of the context in which they present themselves, or because they require highly specialized equipment. Also notable is the presence of studies (16.36%) focused on social science, particularly in language learning and visual art appreciation. Other fields of application were "health" and "Service" with 4 studies each. The devices mostly used during these studies were smartphones and tablets, with one that employed also Oculus Rift Glasses. The main benefits of AR highlighted by the 55 studies analyzed by Chen et al. were: the improvement of learning performances, motivation, engagement, and enjoyment.

The benefits that MR provide to STEM education were further proven by other publications on the topic, where it was found that "mixed reality applications are effective in students' motivation for collaborative work in science teaching. In this context, it can be stated that mixed reality applications can make a significant contribution to the students' potential of collaboration with their peers. In the research, it was concluded that students were willing to use the augmented reality applications in science teaching" [5] and that "The overall results do suggest that VR and MR are both suitable and safe technologies for learning, potentially enabling new approaches to teaching [...] the benefits, such as increased engagement and positive emotions, suggest that VR and MR would be good as supplements to traditional learning methods [...] XRs have a myriad of possible uses, as many environments and interactions can be accurately simulated within virtual environments, enabling access to learning materials that may not otherwise be available to learners [...]. For example, dangerous environments or chemicals" [1].

## 2.3. Extended Reality for Collaboration

The benefit of utilizing XR to aid in collaborative works is a topic that has already been studied and whose benefits have been largely highlighted. Mark Billinghurst and Hirokazu Kato published in 2002 a study[6] where those advantages were already observed, especially when AR gets compared with traditional digital supports. These perks were also shown to be relevant both in remote and in presence collaboration. The main ones that are still relevant to this day and that I would like to report are the following:

- the use of traditional digital support, like a big screen, often results in the separation of the communication space and the digital task space. This division can be removed through the use of AR technologies
- communication does not happen only through words; some notable examples are the way we interact with physical objects, where our gaze wanders, and how we move our bodies. All of these are often lost while collaborating using traditional digital supports, especially in remote settings. AR, instead, allows for more cues to come through during conversation, and by proxy collaboration
- there can be interaction between the digital world and the physical one
- each user can control their viewpoints and have personalized ways to display objects based on their preferences and needs.

In 2014 Iulian Radu performed a meta-review of AR in education[18]. One of the main benefits of applying the technology in the context that he found was the improvement of collaboration between students both in tasks performed outside and inside the classroom setting.

Then, in 2015, Fonseca et al. [9] performed a study whose main goal was to see whether the utilization of AR in a university setting would benefit the academic experience of the students. In particular, they tried using the technology to allow architecture and building visualization students to visualize and interact directly with 3D models of buildings. They found out that AR improved the motivation, engagement, and performance of students. These benefits were even more evident in remote project work and in situations where collaboration was key.

The benefits of collaborating in XR were also highlighted by another review of the literature in 2020 from Pidel and Ackermann [16]. Their study focused on mostly studies performed at most five years from their publication. The main benefits they noticed were:

- an improvement in effective communication given by the ability to visualize and get immersed in the topic of the conversation, making participants less reliant on their imagination;
- an improvement in engagement, especially in remote collaboration settings;
- the ability to modify the users' point of view to simulate another perspective, like the possibility to apply a filter to simulate colorblind vision, or the ability to scale up virtual rooms to simulate a child's perspective. "This could be used advantageously for empathy training, as well as a new method of accessibility testing in architecture and design" [16].

However, they were also able to analyze some of the drawbacks:

- the technology is expensive and evolving rapidly, limiting who can participate. Furthermore, the wide array of platforms makes cross-platform development harder and, if ignored, can further limit the potential audience;
- motion sickness and eye strain can still be felt by the users, especially after longer exposure;
- single tasks may require different mediums. Having to take on and off often bulky headsets interrupts the flow of work;
- the additional complexity of VR and AR translates to more possibility of system of user error

## 2.4. Limitation, Challenges and how to overcome them

All the positive effects of XR in both Education and Collaboration I just discussed, gives me the confidence that this is a field of study that should be further explored. However, there are still several limitations that have to be overcome and challenges that have to be kept in mind while going forward.

As was presented in the previous sections, the main field of application of XR in education was STEM, due to the ability of the medium to reproduce the objects of study of these subjects, especially ones that are difficult or dangerous to see with traditional instrumentation. This makes XR mostly focused on the visualization of and interaction with objects. I would like to go beyond this approach and try to use the medium to also effectively represent abstract concepts and their correlation between themselves and concrete objects. I believe that this would make the medium more suitable for subjects whose main focus is not STEM related, like History, Social Studies, Philosophy, and various Languages to cite a few.

This last point also nicely introduces the next one I would like to make: at the moment, what we have is a huge library of two-dimensional and three-dimensional content ready to be utilized in XR, that right now are mostly employed in a vacuum, or in very simple predetermined experiences. What I would like to do, especially now that XR is starting to become a widely known and moderately used technology in several different fields, is to give all the tools necessary to effectively utilize both this content and the new ways I will introduce to represent abstract concepts, to produce more complex, while still highly customizable experiences without requiring extensive knowledge on XR and any knowledge on how to program for it. This is especially relevant in the world of education, where people who have the necessary knowledge to introduce an effective pedagogical approach to XR in most subjects, especially in the case of not STEM ones, lack the technical knowledge required to implement XR experiences. This also solves one of the main issues that Garzón highlighted[10] and that I reported in the previous sections.

Still, on the topic of the issues reported by Garzón, I will also design my proposed solution to account for a high grade of accessibility and usability. I will also try to solve the problem of dissemination by making my solution available on as many devices as possible. Lastly, I will also focus on the possibility of accessing the experience remotely seamlessly, allowing users even in different parts of the world to join a session and contribute meaningfully.

I will further expand on these points in the following chapters while introducing my solution, how I implemented it, and how I tested it.

# 3 | Technological background

Another important aspect that needs to be analyzed before introducing my solution, is the current technological landscape surrounding XR concerning some of the various devices that one could use and the functionalities already designed or implemented on XR for collaboration, education and the representation of concepts.

## 3.1. Devices

In the following section, I will talk about all the devices that I chose to develop my application for, talking about their main features, strong points, and weaknesses. Then I will go over my reason as to why I chose these supports. Lastly, I will conclude with a short recap and one last comparison between the presented devices.

However, before I start, I have to answer one pretty obvious question: Why go through all the trouble to support multiple devices in the first place? The answer lies in the fact that every device has its own set of characteristics, and with them, a set of strong and weak points, making them more or less suited for certain situations. What's more, giving a wider array of available devices to the end users, allows them to choose between the the one that suits their preferences the most or that they feel the most comfortable with. It also helps with the problem of dissemination that XR application faces and that I discussed in the previous chapter. By having the application available on multiple devices, privates and institutions have the freedom to choose what device to use it on, whether it is one that they may already own or one that they have to buy. This overall has the effect of making the application available in more situations. Having said that, let's start by going through the various devices I decided to support.

### 3.1.1. Hololens 2

"HoloLens 2 is an untethered holographic computer"[13] developed by Microsoft and released in the second half of 2019. It's the evolution of the first version of Hololens, which improves upon several aspects of the visor while introducing new features. As the definition just given suggests, it's a self-contained device that has built-in all the components needed to function. Notable ones are:

- a set of see-through holographic lenses that uses waveguide technology to draw the holograms. They also come paired with an eye-based rendering system that calibrates the device to the user's eye to improve the placement of the hologram, optimizing the viewing experience
- four visible light camera and a depth sensor that, thanks to machine learning driven algorithms, are responsible for hand tracking, spatial awareness, and recognition
- two infrared cameras pointed at the user for eye tracking and iris recognition
- an 8MP RGB camera capable of capturing mixed reality pictures and video up to 1080p 30FPS
- m Inertial measurement unit composed of an accelerometer, a gyroscope, and a magnetometer
- speakers that support spatial audio
- a 5-channel microphone array
- Wi-Fi and Bluetooth modules to empower connectivity
- Qualcomm Snapdragon 850 chip, accompanied by a second-generation custom-built holographic processing unit, 4GB DRAM and 64GB of storage
- lithium batteries that give the device 2-3 hours of battery life in active use.

Figure 3.1: Hololens 2 (credits [1])

Thanks to these, the device is capable of:

- project holograms over a 96.1 (diagonal) degrees field of view over the transparent lenses, allowing the user to see at the same time both the real world and the virtual objects, making them feel like part of the same environment
- track the user's head and body position over six degrees of freedom
- anchor holograms to the real world, letting their position persist over a long period and different sessions
- tracks the user's hands, allowing them to interact with holograms by touching and grabbing them
- track the user gaze to understand where they are looking and adapt the holograms to their eyes
- listen to and interpret voice commands
- allowing to record the experiences by producing mixed reality pictures and videos

[1] https://www.medtrixhealthcare.com/holo-lens

- provide a comfortable wearing experience thanks to the balanced design, which keeps the center of gravity located around the middle of the user's head.

allowing the users to enjoy one of the most advanced and complete mixed reality experiences available. All these capabilities however come at a cost, namely a base price tag of 3500$, one of the highest in the market for this kind of device. However, it is justified given that Hololens still boasts the most advanced and well-responding set of sensors among all the devices I will present. The same, however, cannot be said for its computational power, which in the years has fallen behind the one of newer devices, some of which I will also talk about. Another strong point of the Hololens 2 is the option to use it while wearing glasses, which may require some extra steps or be outright impossible with some of the devices I will speak of in the following sections. Overall Hololens 2 is a highly versatile device that can be used in several contexts like a production site, a medical studio, or a school. It is also easy to pass around between people thanks to the fact that it is a standalone device. The main issue with this is that, thanks to mediocre battery life, the only way to use it for a session longer than a couple of hours is to have it attached to some form of power supply, be it a plug or a battery pack.

### 3.1.2. Meta Quest Pro

Meta Quest Pro [12] is the independent extended reality device developed, as the name suggests, by Meta. It's the successor of the Meta Quest and Meta Quest 2. Its main components are:

- display realized from the combination of patented pancake lens and optics technology and two (one per eye) 1800x1920 pixels LCD panels
- specialized local dimming technology capable of controlling over 500 LED blocks independently
- Mechanical continuous lens spacing adjustment and a dial that adjusts the lens distance from the user's eyes
- five infrared eye and face tracking sensor pointed at the user's face, each with a 120-degree field of view
- two controllers with built-in rechargeable batteries with around ten hours of battery life, haptic feedback, and precision pinch motion
- dedicated Qualcomm Snapdragon 662 mobile processor and three camera sensors per controller

- Ten advanced VR/MR sensors
- built-in audio system, with a dynamic bass extension, compatible with 3.5 mm earphones
- 3-microphone array
- Wi-Fi 6E enabled and Bluetooth 5.2
- Snapdragon XR2+ platform, which is optimized for VR, 12GB RAM, and 256GB of storage

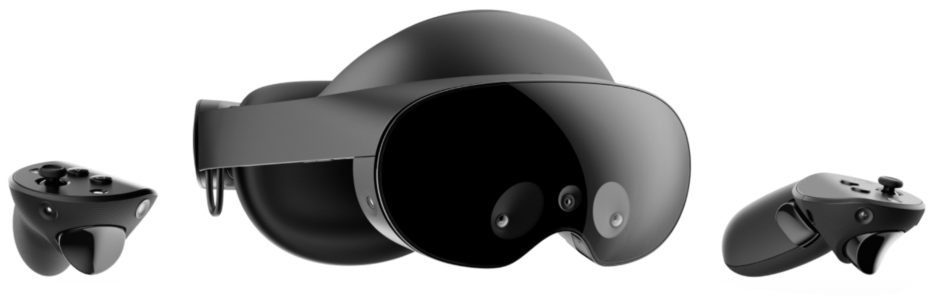

Figure 3.2: Meta Quest Pro (credits [2])

These allow the device to:

- draw objects over a 106-degree horizontal times 96-degree vertical field of view with an overall system resolution of 22 pixels per degree of field of view, sharpness of 4.9 distinguishable line pairs per millimeter at the center, and 4.25 at the edge, high-level of color contrast, over a DCI-P3 color gamut
- have a stereoscopic mixed reality pass-trough, which creates a natural 3D view of the world by combining the feedback of multiple sensors. Compared to a monoscopic solution, it creates fewer distortions and gives off a better depth perception, resulting in a more natural and higher-quality experience.
- performs 6 degrees of freedom inside-out SLAM tracking thanks to the VR/MR sensor and cameras, useful both to perform spatial mapping and controller tracking,

[2] https://www.meta.com/it/quest/quest-pro/

giving a 360 degree of motion in the virtual space

- give the option to the user to use both controllers or their hands to interact with the mixed-reality space
- adjust both lens spacing and distance to improve the viewing experience
- real-time eye, face, and expression tracking
- provide experiences supporting spatial audio
- provide a comfortable wearing experience

Worth mentioning is also the possibility to adjust the level of VR immersion between fully open peripheral vision, to partially blocked thanks to an accessory that comes with the device or to fully immersed with an accessory that is instead sold separately. The base price for this device is around 1200$, with the expected slight variations based on the region of purchase. This lower price tag with regards to the previously introduced Hololens 2 is possible because of the overall cheaper-to-realize viewing technology (displays and optical lenses stacks versus the Hololens holographic lenses empowered by waveguide technology) and slightly worse hand recognition, which however is mitigated thanks to the native support for controllers, which the other is lacking. Spatial recognition instead is on par with one of the previous devices. Another big difference between the two is in how they create the mixed reality space: while the Hololens projects the holograms over a transparent set of lenses, allowing the users to see both the real world and the virtual objects, the Meta Quest Pro instead uses its AR/VR sensors and cameras to provide a pass-trough experience, which means that it reproduces through the displays what the user would see without the visor on. Of course, this comes with a couple of drawbacks: first, the representation of the real world may present some slight distortion, especially for objects distant less than 20cm from the cameras. Second, there may be some lag in the reproduction of the camera feed on the displays. Third, in case of sudden shutdown for any cause, be it battery depletion or software error as some examples, or malfunctioning, the user finds himself completely visually cut off from the external environment until restart or removal of the headset. This last point makes the device not suited to be used in situations where contact with the environment must be never interrupted, like for example when a technician is operating or doing maintenance on a machine, where not being able to see might produce a high safety risk. Lastly, because everything that the users see is created by the display, instead of just a few selected items, the Meta Quest Pro may produce a stronger sense of motion sickness if used for long sessions and may strain the eyes more easily. One last disadvantage of the Meta Quest Pro is that, while it can be used with glasses too, it is slightly uncomfortable, or outright impossible, to do so with some types of frames, especially bigger ones. Overall I believe that for the use that concerns my application, the Meta Quest Pro is a suitable device since most of the drawbacks aren't much of a concern for my use cases.

### 3.1.3. Magic Leap 2

Magic Leap 2 [11] is another independent mixed reality headset developed by the homonym Magic Leap. It's the evolution of the first Magic Leap, upon which it vastly improves. Unlike the previously introduced Meta Quest Pro, it allows the user to see the environment through its transparent lenses over which it draws the virtual objects, much like the Hololens 2, rather than being a virtual reality headset that reproduces the environment thanks to cameras and sensors feed. It's made up of three components:

- An highly ergonomic headset that weighs just over 260 grams and sports:
  - two lenses that work as the AR display, with a 70-degree diagonal field of view, 1440x1760 pixel resolution, 120Hz refresh rate, 37cm to infinity range of operation, and adjustable 20 to 2000 nits brightness
  - a 12.6M pixel auto-focus RGB camera capable of recording 4K at 30fps or 1920x1080 at 60fps video
  - Several sensors including 3 wider FoV world cameras, a depth camera, an RGB camera, an ambient light sensor, 4 eye tracking cameras, 4 inertial measurement units, a 3-axis accelerometer and gyroscope, a 3-axis magnetometer, 2 altimeter
  - 2 Built-in Stereo Speakers and 4 Microphones
- A powerful compute pack attached to the headset through a cable that, thanks to its split design, can be fastened to one clothing that is composed of:
  - AMD 7nm Quad-core Zen 2 x86-64 core with a thread count of 8, a maximum frequency of 2.4GHz, 512kB L2 cache per core, 4MB total L3 cache and a custom 14-core computer vision processing engine (CVIP)
  - AMD RDNA 2 with max Gfx frequency capped at 1.1 GHz as GPU with one shader array, four workgroup processors, two RB+ render backends, and 1MB L2 Cache
  - 16GB of RAM and 256GB of storage
  - Wifi 6 and Bluetooth 5.0

  - around 3 and a half hours of usage time

- A 135g controller with a touch-sensitive trackpad and capable haptic feedback, which the visor keeps track of thanks to higher accuracy 6DoF optical tracking

Figure 3.3: Magic Leap 2 (credits [3])

This gives the device the ability to:

- Show virtual objects while allowing the user to maintain contact with the environment
- Track the user position
- Allow the user to interact with the virtual objects both with their hands and the controller, based on personal preference or the amount of precision required to perform certain actions
- Track the environment to allow interaction between it and the virtual objects
- Listen for users' voice input and act on them
- Track the user gaze to be used as an additional input
- Reproduce spatial audio
- Produce mixed reality video and pictures

[3] https://www.magicleap.com/news/magic-leap-announces-pricing-and-us-general-availability-date-for-magic-leap-2

The headset unfortunately cannot be used with glasses, however, prescription lenses can be mounted directly into the headset thanks to quick swap inserts. Magic Leap 2 comes at a starting price of around 3200 dollars. Its main selling points are the possibility to both use hands or a controller to interact with the virtual objects, although the hand recognition is slightly worse than the one in the Hololens 2, the lightweight of the headset, and the high computational power, especially if compared to the Hololens 2 again. The main drawback is the fact that the user has to keep close to them the compute pack, which is attached to the headset through a not detachable cable. In case of stationary use, this is not a problem, since the user could simply leave it on a nearby surface. To use it on the go, instead, the user has to either hold the compute pack in one hand, making it impossible to interact with both hands with the virtual objects or hang it in a pocket. This last option is made possible thanks to the split design of the pack, which allows half of it to be put inside the pocket, while leaving the other half outside, allowing easy access to the power button, and the LEDs and keeping the cooling vents in contact with an open environment. Both this last point and the fact that it has to be used either without glasses or with prescription lenses mounted on them make the device less prone to be passed around. In my case, for example, it would be much better used by teachers rather than students.

### 3.1.4. XREAL light

All the devices that have been discussed until now were standalone visors that had hardware both to show the virtual objects and to perform all the necessary computations. However, in the market, there exists also devices that only archive the first goal, while leaving the second to external hardware to which they are connected. An example of such a device is the XREAL light [20] which can be connected to a wide selection of Android phones (for a comprehensive list, please, refer to the official XREAL website[19]). As such, the specifications for the computational power of this setup vastly depend on the phone used. What remains constant instead, are the hardware characteristics of the glasses:

- 3840 by 1080 pixels OLED display with fingerprint-resistant coating
- a wide set of sensors and controls including screen brightness adjustment button, accelerometer, gyroscope, magnetometer, proximity sensor, ambient light sensor, 2 Computer vision cameras, and a 5MP camera
- 2 open-ear speakers
- dual microphone array

- prescription-lens frame

Figure 3.4: XREAL light and smartphone (credits [4])

These, together with the computational characteristics of the connected smartphone, allow the device to:

- Display virtual objects over the environment over a 52-degree field of view (amounting to around 42 pixels per degree of resolution) with up to 280 nits perceived brightness
- Be able to perform human and environment understanding, including 6 degrees of freedom head tracking, hand tracking, image tracking, and plane detection
- use the connected phone screen as a touch-pad to interact with the virtual objects

The XREAL light comes at a relatively low price of around 600$, far lower than the one of the other mixed reality devices I've presented up until now, even if one has also to buy a compatible smartphone, for which he would need around 400 extra dollars. The lower price tag comes however with a few disadvantages. The device has a much lower field of view compared to the other devices seen until now. It also has the worst hand recognition capabilities, giving less precise control over virtual objects. Additionally, it misses some of the other mixed reality interaction paradigms like voice recognition and eye tracking. Finally, just like the Magic Leap 2, it cannot be used with glasses, but prescription lenses can be easily mounted on the device. Overall the device can still be used to take part

[4] https://www.xreal.com/light/

in mixed reality experiences of all kinds, although I advise using it in situations where active participation is limited, due to the lower accuracy of the interaction mechanism. It is also a great device to consider to be bought in high quantities thanks to the lower price tag.

### 3.1.5. Smartphones

The last family of devices I focused on for the fruition of mixed reality experiences is smartphones. Out of all the presented options, this is the most accessible of them all, given that most people nowadays possess one, and even if it is not the case, one that is capable of running mixed reality applications can be bought for less than 400$. All this accessibility however comes at the expense of the quality of the experience. The field of view is widely reduced since the user can only see the virtual objects through the screen. For the same reason, immersion is also lower. Lastly, interacting with virtual objects is also worse, because it can only be done by touching the screen or using voice commands. Hand recognition and eye tracking are not supported. However, mixed reality applications for smartphones can still have 6 degrees of freedom. This is mainly achieved through the use of either a mix of GPS and the various sensors (like accelerometers, gyroscopes, etc.) mounted on the device, or algorithms based on image recognition or point clouds. These systems, however, are not as reliable as the ones from the other devices I talked about until now, resulting in a lower stability of virtual objects' position and not allowing any persistence between different sessions. Just like for the XREAL light, I believe that the best use cases for smartphones are when a high amount of them is required or when users need to take a less active role.

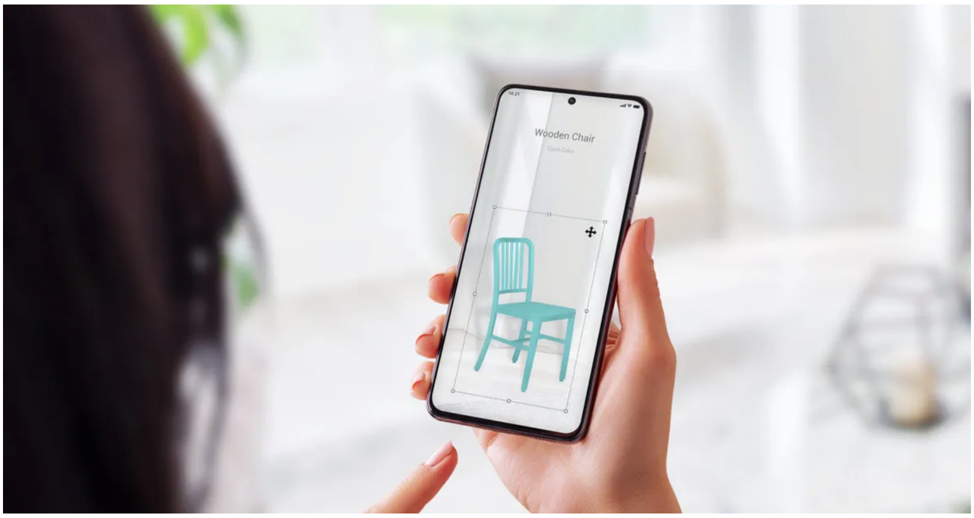


Figure 3.5: AR with smartphones (credits [5])

### 3.1.6. Final Comparison

Before giving a recap of what was said about the devices and comparing them, I would like to give my reasons as to why I chose these devices in the first place: it's because together they give a good representation of all the MR devices in the market. In fact, for almost every characteristic that one of those may have, there's at least one in my list that has it too. This allowed me to get a deeper understanding of how each of these characteristics impacts the overall experience, while also giving me the ability to make some educated guesses on how other or future devices might perform in different situations and use cases based on what characteristics they present. Here's a short recap of the major features of each device:

| | **Hololens 2** | **Meta Quest Pro** | **Magic Leap 2** | **XREAL light** | **Smart-phone** |
|---|---|---|---|---|---|
| Display technology | Augmented Reality | Virtual Reality + Pass-trough | Augmented Reality | Augmented Reality | Augmented Reality |
| F.O.V. (diagonal) | 96.1° | 143° | 70° | 52° | Highly dependant on model |
| Main interaction mechanism | Hands | Controller or Hands | Controller and Hands | Track-pad and Hands | Touch Display |
| Other interaction mechanisms | Voice Commands, Eye Tracking | Voice Commands, Eye Tracking | Voice Commands, Eye Tracking | Voice Commands | Voice Commands |
| Supports spatial audio | yes | yes | yes | no | no |
| Can be used with glasses | yes | depends on the frame | no, but prescription lenses can be mounted | no, but prescription lenses can be mounted | yes |
| Approximate price | 3500$ | 1200$ | 3200$ | 600$ (+400$ for smart-phone) | 400$ |

Table 3.1: Table of devices and their main characteristics

[5] https://airlapp.com/blog/realta-aumentata-ar/

## 3.2. Tinalp

This project was developed as an Activity to add to Tinalp, a platform created by FifthIngenium, a company whose software development offices are based in Milan. Its objective is to create advanced experiences in mixed reality for training and education purposes. The main goal of Tinalp is to create what FifthIngenium calls an "Augmented classroom", or in other words, a mixed reality space where both teachers and students can create lessons empowered by the use of both traditional two-dimensional media like PDFs, presentations, and videos, and innovative 3d content. However, it is important to not mistakenly think that Tinalp is just a collection of predefined and static mixed reality educational experiences. Both the flow and the contents of augmented lessons can be created, customized, and deployed by teachers and students alike, through the use of an intuitive and code-free online editor.

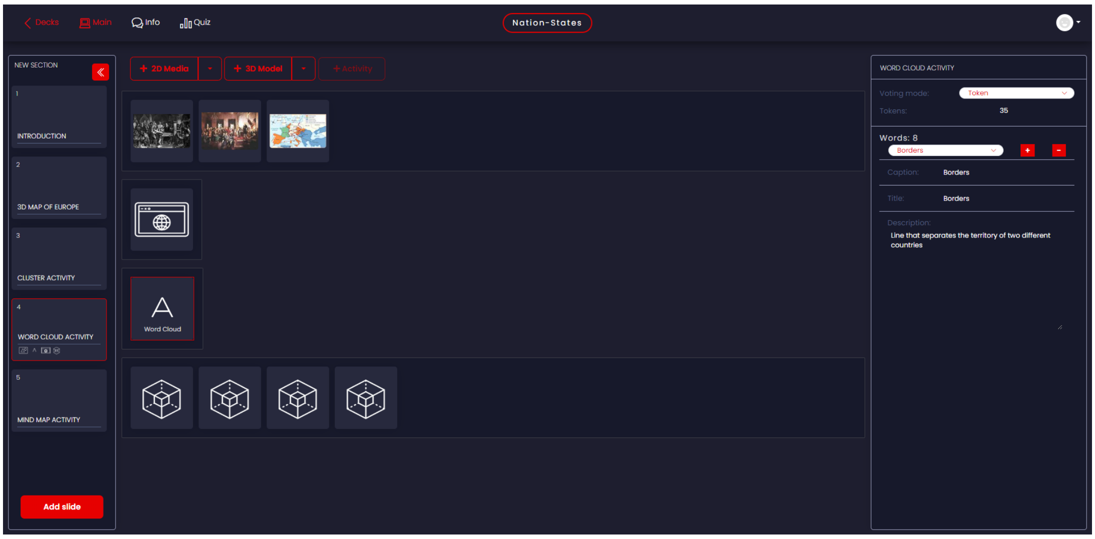


Figure 3.6: Interface of the web editor

Both the web application and the structure of these lessons are heavily inspired by PowerPoint. The editor can be used to create what is called a "Deck", a concept akin to presentations, which in turn is divided into "Steps", the equivalent of slides. Inside these steps is where the content of the lessons can then be added. Once the deck is created, it can be accessed anytime from the mixed reality space. The previous picture shows the page inside the web editor where one can add and modify the content of the deck. The section on the left shows the various steps that the deck is composed of and it can be used to navigate between them. The central section instead shows all the content currently available in the selected step. Trainers can use the buttons at the top of the page to add more content and upload images and 3D models. By selecting one of the contents present, a third section on the right will appear where all the details and the available options for the selected piece of content are shown.

The augmented classrooms can then be carried through with the content created in the web editor. To participate, teachers and students can use one of the many devices supported by Tinalp. The list now contains all the devices I previously introduced, and it is in constant expansion. In situations in which there are not enough devices for everyone, the lesson can be projected onto a screen, using a mobile phone as a mixed reality capture system. These sessions can be joined both locally, so the users share both the physical and the virtual space, or remotely. In the second case, the avatars of the other users are shown and the interaction between people can be carried through also thanks to the voice communication system that is in place. All this is made possible by the fact that the session is synchronized between all the devices that are taking part, meaning that everything that one sees and does in the mixed reality space is also replicated in real-time to all other devices that are participating.

Users joining an augmented classroom are also divided into two categories, trainer and trainee, which mirrors the structure teacher/student that most educational environments follow. Moreover, one of the trainers is also considered the owner of the session. By default, it's the trainer that started the augmented classroom, however, if by any chance they disconnect, the ownership switches to another trainer. The owner has control over the flow of the experience, deciding which deck and step the session is currently in and what content is shown. This is done through the use of an apposite menu that only the owner can see and interact with. Trainees instead have their own little personal menu that allows them to do simple actions like leave the current session, set again their mixed reality space, or raise their hand. This last action notifies the owner that they may want to speak or interact with what is currently being shown to the class (more on this later). Most of the actions that involve control of the content of the deck can be initially done only by trainers. They can interact with 3D objects by moving, rotating, and scaling them directly with their hands. Clipping planes can be used to show only sections of the models. Additionally, if supported by the model, 3D objects can also be split into parts that can then be individually made invisible or interacted with independently. Trainers can also change the page of a PDF or a presentation that is currently shown or use a digital pointer to draw attention to part of them. They can also pause a video that is being played, resume it, or seek a particular moment using a slider. Trainees instead initially act only as passive viewers, however all of these capabilities can also be given to by the room owner, either directly from another tab in their menu or by approving a request of control sent by one of the trainees.

Contrary to what one could think, the augmented classrooms made possible by Tinalp can be used not only for subjects like chemistry and mechanics, where 3D models are already widely used to represent things like molecules or mechanical devices but also for subjects where abstract concepts are studied. This is made possible by the ability to represent both texts and images as 3D elements inside the augmented classroom and the possibility to add various activities to each step. These exploit the aforementioned ways to represent text and images and 3D elements to create more structured, yet highly customizable, experiences. Currently, the following activities are available:

- Cube quiz: users are supposed to answer questions using a cube and the concept represented on its faces through the use of images or text. to do so they need to place the face that they believe referees to the correct answer upward. Both the question and the concept represented on the face of the cube are chosen during the creation of the activity by the trainer.
- Cube poll: similar to the previous activity, users are required to use a cube and the concept represented on its faces to express their opinion on a topic. The results are displayed in real-time on a panel visible to everyone. Again the question and the faces of the cube are chosen by the trainer during the creation of the activity.
- Card Memory: users will play an extended reality version of the famous memory game, in which they'll try to uncover the conceptual link between what's represented on the faces of the various cards through the use of text and images. Both the content of each card and its paring are decided by the trainee during the creation of the activity.
- Cluster activity: users are required to divide concepts represented on cards and cubes, placing them in several areas scattered around the mixed reality space. Again, the content of cards and cubes, together with the topic that each area represents are decided by the trainer during the activity creation.
- Mind Map: the last of the already existing activities that are currently available in Tinalp is one in which, using a set of pre-prepared concepts added to the activity by the trainer during its creation, users can create a three-dimensional mind map. The concepts are represented either by cards or cubes containing text or images, or by simple 3D text, and can then be connected using lines or arrows.

A more detailed overview of various activities currently available can be read in a previous thesis written by Matteo Di Marco[8].

Inside Tinalp is also possible to create surveys that can then be administered to the trainees who are participating in a session. These come with a lot of the features usually available in any survey application like the possibility to mark some answers right and to randomize the options, just to cite a few. What's more, they also offer some features that are possible only in an extended reality space, like the ability to connect a question with a part of a 3D model to give more context. The results of these surveys can then be accessed through the web application that I talked about earlier. Other than the results of the various surveys, the application also records other kinds of data about the sessions that may be useful to a teacher in evaluating the effectiveness and the engagement of the augmented classroom, like how long each user joined a session for, just to give a simple example. However, these data are currently inaccessible through the web portal, which is a feature currently planned.

One of the key strengths of Tinalp is the promotion of active learning as in the act of involving students in lectures and activities in which they are meaningfully engaged in the learning process. Several studies have been carried out through the years to promote the usefulness of active learning, as in the improvement of concept retention and engagement by students after applying this method to their learning process. Michael Prince does an excellent job in his article "Does Active Learning Work? A Review of the Research"[17] in collecting the results of several studies on the topic and reporting them. In particular, they show that the introduction of student activities interleaved in the lectures improves short-term concept retention by 35% and long-term concept retention by 11% and that the substantial use of interactive and engaging methods can more than double the students' understanding of the topics explained.

## 3.3. The Mixed Reality in Education project

This thesis work, which I will shortly begin to present in greater detail, is not just part of a product called TINALP, but it's also part of a multi-annual project within a "framework agreement" between Politecnico di Milano and Vodafone Italia S.p.a. The project is named MIRE and it aims to

> put forth a new approach, proposing to use MR to represent concepts instead of objects, designing educational experiences where MR is used to prompt students to recall, select, and combine, concepts either in an individual or in a cooperative way, via playful and competitive activities. MR thus becomes the trigger for learning content [...] and competencies. [3]

MIRE also involves HOC-LAB, a laboratory from Politecnico di Milano and the, at the time startup, and now company FifthIngenium s.r.l.s. The second got involved in the project after their victory at the contest "Action for 5G" hosted by Vodafone Italia in 2020, after which TINALP was chosen as the technological platform to implement their novel approach.

### 3.3.1. The project in greater details

Thanks to the advances in both the technologies that enable XR content consumption, with the devices that I already largely talked about and in the technologies that empower connectivity, with the advent of 5G, the focus in creating MR application can be shifted. It is now possible to switch from how to overcome technological challenges in implementation, as it has been until recently and which has often limited the focus on content, to how to design and use them as support for learning activities. This is what MIRE proposes to do. Thanks to the features of TINALP that I previously talked about, educators can utilize the "No-CODE" visual web editor to focus on the design of the experiences, orchestrating and providing the content for it, without needing any technical knowledge. This way their focus can switch back to the pedagogical aspect of the experience, rather than on its technological foundation. Other than experience creation and customization, another big focus of the MIRE approach is data display again thanks to the tools provided by TINALP.

The educational experiences have been designed to promote the representation of concepts rather than objects. To do so, the MR space is populated by 3D models that resemble simple geometrical shapes. Each one of those has the role of conveying content, be it a picture, a quiz, a video, and so on. Activities have then been designed to make participants interact with those shapes, and by proxy with the content they represent, in different ways to achieve specific goals. They also were made to follow some specific requirements: promote active learning, favor social interaction, high engagement through gamification, and most importantly improve the motivation for the study of the background content by the students. A comprehensive list of activities together with several use cases for them can be consulted in the article [3]

This approach was later tested thanks to a first implementation of some of the activities in a high school as reported in a second article on MIRE[2]. The results of this testing were pretty promising, showing how it can strengthen the acquisition of even non-trivial content and boost collaboration between peers. However, it also elevated some critical points of the system, namely students would like to receive immediate feedback on the result of the activity through scores or visual cues and they would like more enticing and look-alive visual elements and environments.

### 3.3.2. Supporting Project Work with Mixed Reality

Following this first phase of the project, a group of students from ASP designed an evolution of the system based on the ability to support project work. They took inspiration from the agile methodology, which consists of dividing the job into smaller tasks, assigning them to members of the project to complete, and then optionally giving those a priority, decided either by the creator of the task or by some or all the project contributors. Usually, tasks are created to cover about two weeks, called a sprint in the terminology of the agile methodology. Then periodically, both during and between the sprints, there are checkups to see the progress on tasks and eventually adjust them.

At the core of this concept evolution was the division of the project that is going to be supported by the MR tool into "thematic islands" containing a set of sub-activities. These are carefully crafted to aid in the island's main purpose and were made following the design principle of activities that I presented in the previous section. The islands that were devised are:

- "concept island" with activities supporting the discussion over concepts and keywords, like brainstorming and word clouds;
- "island of planning" containing activities for planning and problem-solving, like flow chart creation;
- "decision island" with activities supporting debate and surveys to facilitate discussion and decision between the members of the project.

The tasks for this phase have been divided as follows: the three tutors supervised the entire work, with a specific focus on the pedagogical aspect (Prof. Di Blas, HOC-LAB), the technical one (Matteo Valoriani, FifthIngenium) and the application to the business world (Cinzia Campanella, Vodafone). The brainstorming on the general concept of supporting a working group was carried out by the students of the ASP project; the identification of the Word Cloud and the definition of the requirements was carried out in collaboration between them and me, the technical implementation was entirely my responsibility while the testing was again carried out in collaboration.

# 4 | Design

In this chapter, I describe the design and implementation of one of the activities identified to support a work group: the Word Cloud. The design was guided by the following educational requirements:

- the educational experience must support the discussion of different concepts
- the experience must allow users to propose concepts
- the experience must allow users to endorse concepts
- the experience must be collaborative
- participants belong to two different ranks: those who have the authority to "manage" the experience and the simple participants.
- the system should show users the relative degree of approval between keywords
- the system should clearly show to all users when any action is performed on the Word Cloud

In the design, We also took into account certain constraints and the "lessons learned" derived from the literature. We directly addressed:

- the problems highlighted by Garzón[10] that I talked about in section 2.2 (accessibility, usability, lack of dissemination and pedagogical approaches)
- the issues that I discussed in section 2.4 (the focus on STEM subjects and the focus on content rather than experiences)
- the observations gathered in the second MIRE paper[2] and reported in section 3.3 (More feedback and look-alive elements)

## 4.1. Use Cases

I will begin the study of the design of the Word Cloud by exploring some use cases that guided us in the creation of the activity.

### 4.1.1. Brainstorming session

Anna, Bruno, Carla, and Dario are in a group that was tasked to design a mascot for a new line of ice cream products. They are at the beginning of their work and are required to do a session of brainstorming to get the ideas flowing. They however find themselves in two distinct locations. Everyone boots up their MR device of choice. Carla, the team leader, who will also act as moderator of the upcoming brainstorming activity, creates a session for the other members to join. Once everyone is in, Carla informs everyone that they will have 5 minutes to add keywords to the cloud and starts the Word Cloud activity. After the time limit is reached, a discussion starting from the Word Cloud that has been created begins. During the discussion the participants start reorganizing the keywords by directly interacting with them, to give order to the great number of ideas that have been brought forward during the 5 minutes. They notice that a lot of keywords refer to possible colors for the mascot. After a small discussion, they reach an agreement on the color palette to use and Carla proceeds to delete all the color-related keywords that are not part of the defined palette. In the end, they settle on a series of aspects that their mascot needs to represent and a set of characteristics that they want it to have, ready to move forward with their design.

### 4.1.2. Philosophy review

After a lecture on the philosophies of Immanuel Kant, a classroom of students is required to revise the main concept of the topic under the supervision of their teacher Sara. They decided to use the Word Cloud as a means to support the review process. The teacher starts the activity and the more eager students immediately add a few initial words like "imperative", "critical", noumenon", "categorical imperative" and "cogito ergo sum". Some students decided to use voice dictation to add their word, while others preferred the keyboard because either they felt shy about speaking to a device or because their word was in Latin. After a while, other students start to add new words and begin voting on the ones added by their peers they feel are most relevant. Words like "categorical imperative", "judgment" and "critical" start to get traction, while others like "predisposition" and "cogito ergo sum" remain unpopular. After 10 minutes, the teacher decides to stop the students to start discussing the result of their activity. First, she removes all the words that are unrelated to the topic at hand like "cogito ergo sum" explaining to the students why that is. Then she proceeds to put together all the words that represent the same concepts. After that she asks some students to help her reorganize the cloud to have all the words referring to the same part of the Kantanian philosophy near each other, all the while she explains the rationale behind their actions. Finally, she decides to prompt the discussion on why the students decided to vote so much for some words while leaving others without much support while also giving her professional opinion on them.

### 4.1.3. Task priority assignment

A group of eight people, including Mark, the Project Manager (PM), is working on a project utilizing the agile methodology. After having divided the work to do in 12 tasks, they decided to use the Word Cloud as a way to decide their priority. The PM already prepared a Word Cloud containing one word for each task using the web editor. Everyone gets his MR device of choice ready and enters the session created by Mark. Every participant gets a total of 40 votes to distribute between the various tasks following the criteria that the higher the number of votes a task has, the higher priority it has. After everyone uses up all their votes, the PM takes back control of the activity and starts discussing the results with his fellow project members. After the activity is completed, Mark uses again the web editor to get how many votes each task got and uses this information to assign the priority to each one.

## 4.2. Word Cloud Design

The Word Cloud, as the name suggests, is an agglomeration of distinct words or small groups of words (which from now on I will also be referring to as "word" for the sake of simplicity), each one representing a concept. As such we decided to visually depict the cloud as a group of spheres each one containing one word.

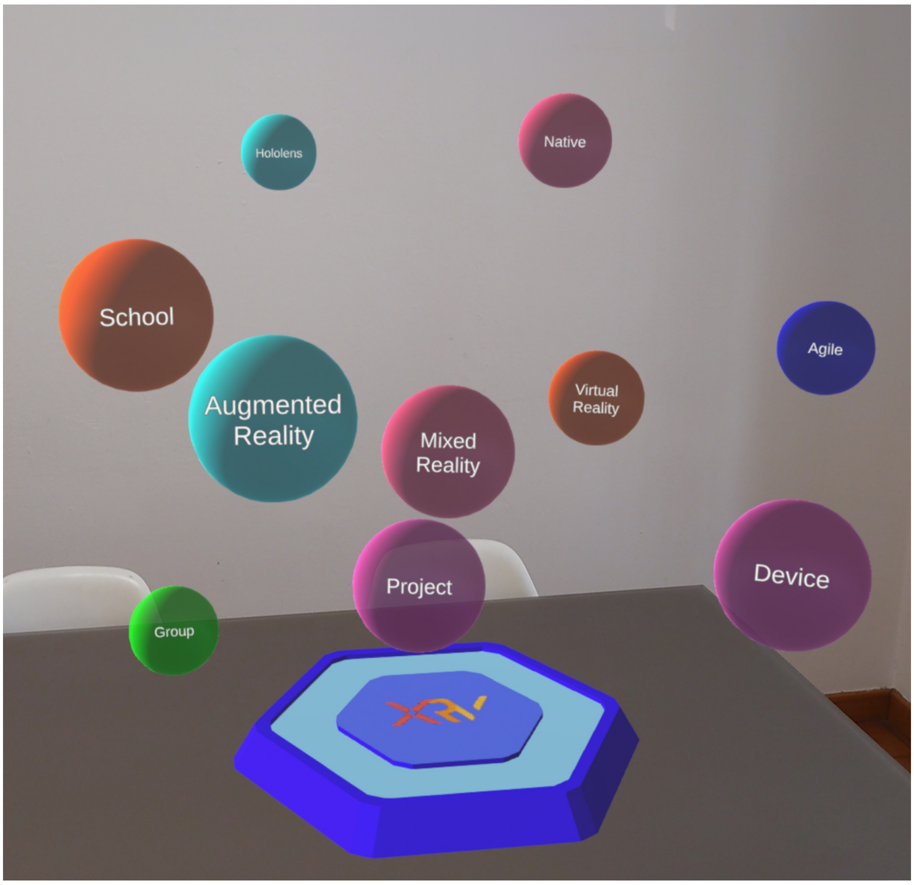


Figure 4.1: A Word Cloud

These spheres are adorned with a moving hold animation, which is a technique often used by animators to make characters and objects hold a pose without making them feel unnatural by adding to them some small movements they perform even when they are supposed to stay still. This significantly contributes to enhancing the illusion that the Word Cloud is not merely a virtual object but rather something actively created and interacted with, physically occupying a distinct space. This also helps to create more enticing and look-alive virtual elements, which was one of the main complaints brought forward by students during the first MIRE testing session.

Many of the scenarios where the Word Cloud is going to be used present a differentiation in the roles of the participants. In particular, there always seems to be at least one user whose role is moderating the activity and optionally taking part in it. As such we decided to create two different tiers of users for the activity. For the sake of consistency with how they are implemented, I will refer to them by using the Tinalp terminology of trainer and trainee, where the first is meant to represent the user with the moderation privileges, while the second is the standard user. Unless specified otherwise, all actions can be performed by both types of users

After studying the previously presented use cases we decided to divide the Word Cloud into two separate phases. During the first one participants of the activity are allowed to either add new words to the cloud or add or remove votes for existing ones. To add a new word users can either use the keyboard offered by the device or a dictation system. This was done to offer an alternative for systems or situations in which one of the two is not available or feasible, all to improve usability and accessibility. Once a word is added a sphere containing it appears in the cloud unless it is already present. If so a vote to the word is added instead. Voting can also be done by using a specifically designed menu containing a list of all the words in the cloud. Every time a word gets a vote the corresponding spheres become slightly larger and vice-versa giving immediate feedback on the change to all the users, again addressing another of the complaints raised by students who tested MIRE the first time. Three voting modes were designed:

- multiple: each user can vote for any word any amount of times;
- singe: each user can vote for every word at most once;
- token: Each user has a limited number of votes that can be distributed between the various words.

While the first is the most basic way of having a voting system that can be used in many different situations, the others all have their strong points making them better suited for different situations. They can also be used to give more structure to the activity. For example, the single mode is perfect for the Word Cloud used as a support for brainstorm sessions, where usually users add new keywords or agree with ones already present, so having the possibility to vote multiple times for the same word is not necessary, if not detrimental to the activity. The token mode instead is perfect to either directly limit the number of votes each user has at his disposal or to give higher weight to each vote since having a limited number of them means that giving one to a word also implies not giving that same vote to the others. Trainers can also remove any word using the same menu that they use to vote. This feature is intended to be used in situations where some participants decide to add words that may be inappropriate or offensive. At any time during the first phase, the trainer can decide to pass to the next one.

In the second phase, words cannot be added or voted anymore. Instead, users can interact with the words of the cloud directly. Initially, all methods of interaction are only available to trainers, who at any point can make these available to any trainee of their choice. This is done because these methods of interaction are implemented as a mean to interact with the ideas that each word represent. Often the people that have the role of trainer correspond to the ones that have the role to orchestrate the activity that the Word Cloud is meant to support. As such they should either be the ones to interact with the ideas first or to decide whether is the case for some or all of the other participants to do so. This limitation is also useful in situations where there is a high number of participants and where moderation is required to avoid chaos and disorder. Users who can interact with the Word Cloud can:

- move spheres around to reorganize them following some logic;
- resize spheres;
- delete spheres by squashing them or by using a menu. If done using the first method, while the sphere is being scaled down, it will also start to vanish, clearly indicating to the user that they are about to delete it;
- merge spheres by holding two of them at the same time and making them touch. Again this behaviour is transparently indicated to the users by the appearance of lightning between two approaching spheres and a merging animation.

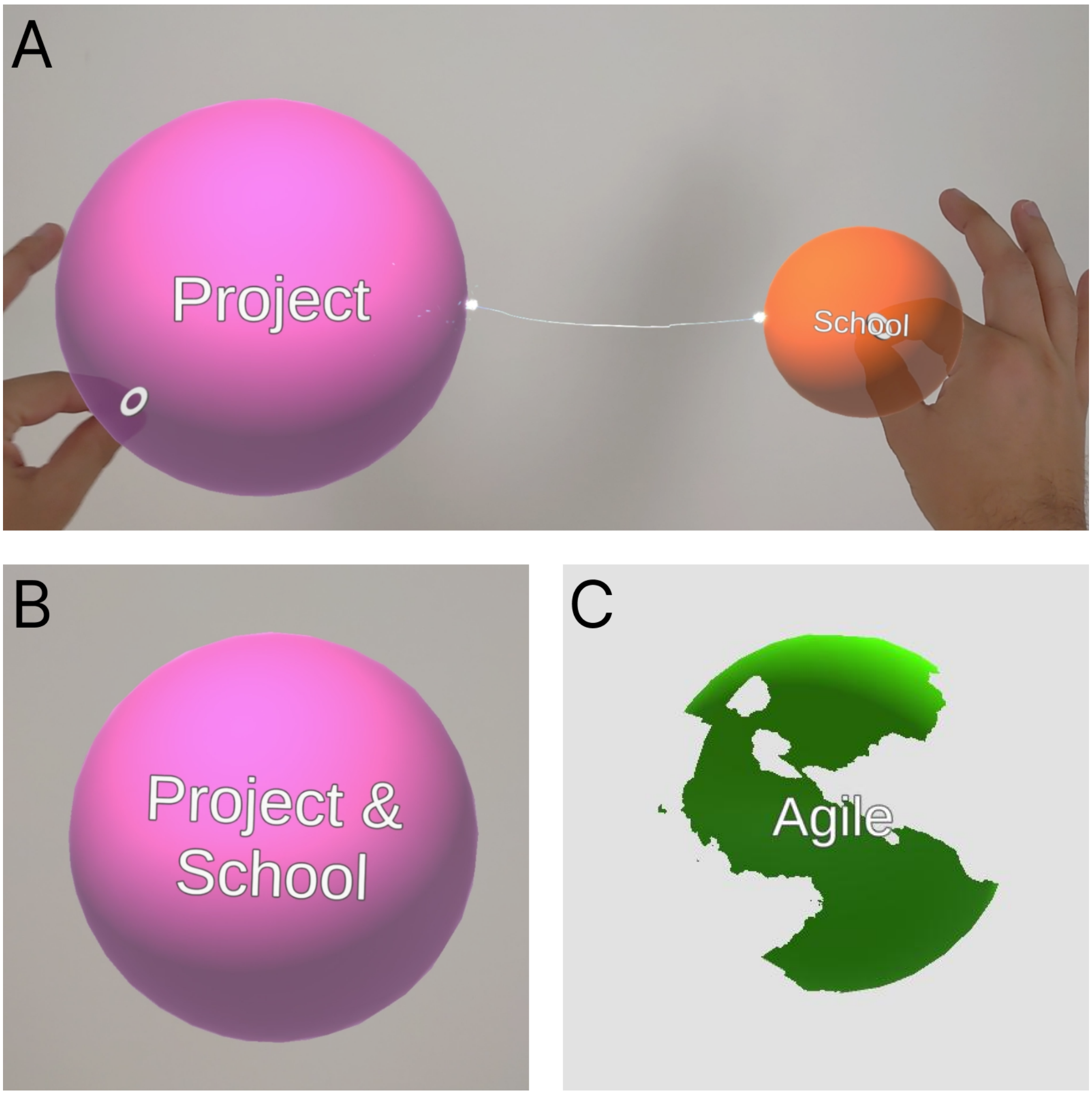


Figure 4.2: Two spheres about to be merged (A) and their merged form (B). A sphere about to be deleted(C)

Additionally, a not intractable view of the Word Cloud snapshotted at the moment the first phase ended and the second one began can always be toggled on by trainers or trainees with permission to interact.

At any point during the activity, the current state of the Word Cloud can be saved and persisted. This is also done automatically when trainers decide to switch from the first to the second phase. The saved data can then be accessed at any point through a web portal. The same web portal can also be used to configure the activity by choosing the voting mode, the number of tokens in the case of the token voting mode, and by optionally adding some words that should be already present in the cloud at the start of the activity.

To access some of the functionalities we described, a menu is necessary. We opted for a menu that by default presents a tag-along behavior, always staying between 30 and 60 centimeters from the user. This behavior can be toggled using the pin button or by grabbing the menu. The interface can also be manually moved, rotated, and scaled.

While designing the Word Cloud, several usability and accessibility aspects were taken into consideration. Some I already discussed in the previous paragraphs, however, I would like to present now a more complete list:

- have the text inside the spheres always displayed in the direction of the user, allowing them to move around without compromising readability;
- distribute the spheres in a way that allows a high percentage of them to be visible from any angle and ensure that every sphere is visible from a sufficient number of angles;
- ensure that words are spelled correctly by showing them to the user and asking for their confirmation before their insertion in the cloud;
- showing both the list of words with their number of votes also in a menu;
- choose a color palette for the spheres that makes them clearly distinguishable even for color-blind people;
- choose a color for the text inside the spheres that has high contrast with the aforementioned palette and an adequate font size;
- allow in the second phase to see a snapshot of the Word Cloud taken at the end of the first phase;
- give a moving hold animation to the spheres;
- give visual feedback for the actions of a user to every single participant in the activity;
- allow users to scale up or down the menu, according to their preferences or necessities.

Lastly, another important aspect of the Word Cloud is the fact that users should be able to participate in the activity and see both other users and their contributions to the cloud even if they are connected remotely. This is achieved by showing the avatars of users in other physical settings, by allowing communication directly inside the application, and by showing the immediate effect of people's actions in the Word Cloud.

# 5 | Implementation

Having shown the design principles and philosophies that guided us in the creation of the activity, I would now like to focus on the implementation of the activity. This process was entirely done by me, following the design choices described in section 4.2 concerning how the system should behave while expanding upon it with all the necessary information required for the implementation. I opted to use Unity as the framework to implement the application. This came with mainly two advantages: the possibility to simplify the deployment process of the applications to several devices and the possibility to use the Mixed Reality Toolkit (MRTK).

MRTK is a package for unity developed by Microsoft that aims at abstracting and simplifying the implementation of the most common mixed reality interaction paradigms and functions. It also offers several ready-to-use components, shaders, and materials optimized for MR devices.

The Word Cloud is also implemented as an activity inside of Tinalp, allowing me to exploit several of its features like the web editor, the local and remote session management system, inter-device communication, and the trainer and trainee separation system that I talked about in section 3.2.

## 5.1. Note

Before delving into how the Word Cloud was implemented itself, I shall first look into the elements that it is composed of. In particular, I decided to first develop a more general and transversal concept of note, which every word of the cloud is then part of. This choice came with the following benefits:

- it allowed me to have at my disposal a system for the representation of groups of information that could be reused in several different contexts
- it let me predispose the Word Cloud to extensions like, for example, the future possibility to associate to each word a description and some content like text or images.

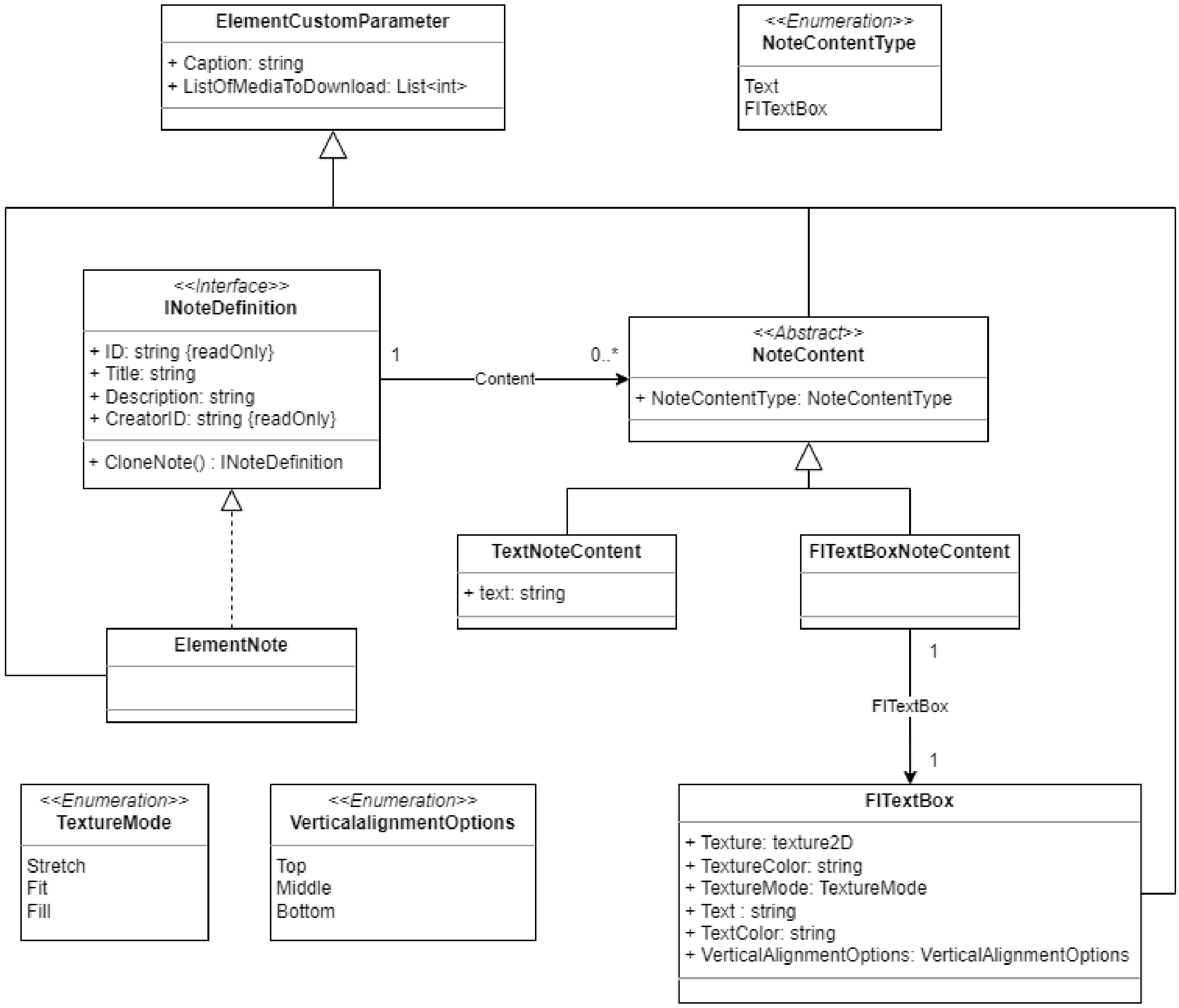


Figure 5.1: Note Data Model - UML

The main characteristics of a Note are described by the interface INoteDefinition, allowing me to possibly define multiple implementations based on the context of use. For my purposes, the interface was implemented in the class ElementVotableNote. Said characteristics are a unique identifier, the ID of the user who created it, a title, a description, and optionally some content, which could be in the form of plain text or an image, with the possibility of extension in the future. In particular, the support for containing pictures is offered by the class FITextBox which "combines one image, a brief sentence, and some customization parameters"[8]. As can be noticed from the data model, several of the classes involved in the definition of a note inherit from ElementCustomParameter. This class is necessary to keep track of and download all the media (like the aforementioned images) that the element needs.

The system also supports the possibility of creating a new note with just a title. In the future I would also like to implement the option to add a description and other content to the notes directly inside a session, however for now I decided to focus on other features. To do so I designed a simple menu that allows the user to insert the title of the new word either by clicking the text box and then by using the keyboard that the device provides, or by using a dictation system.

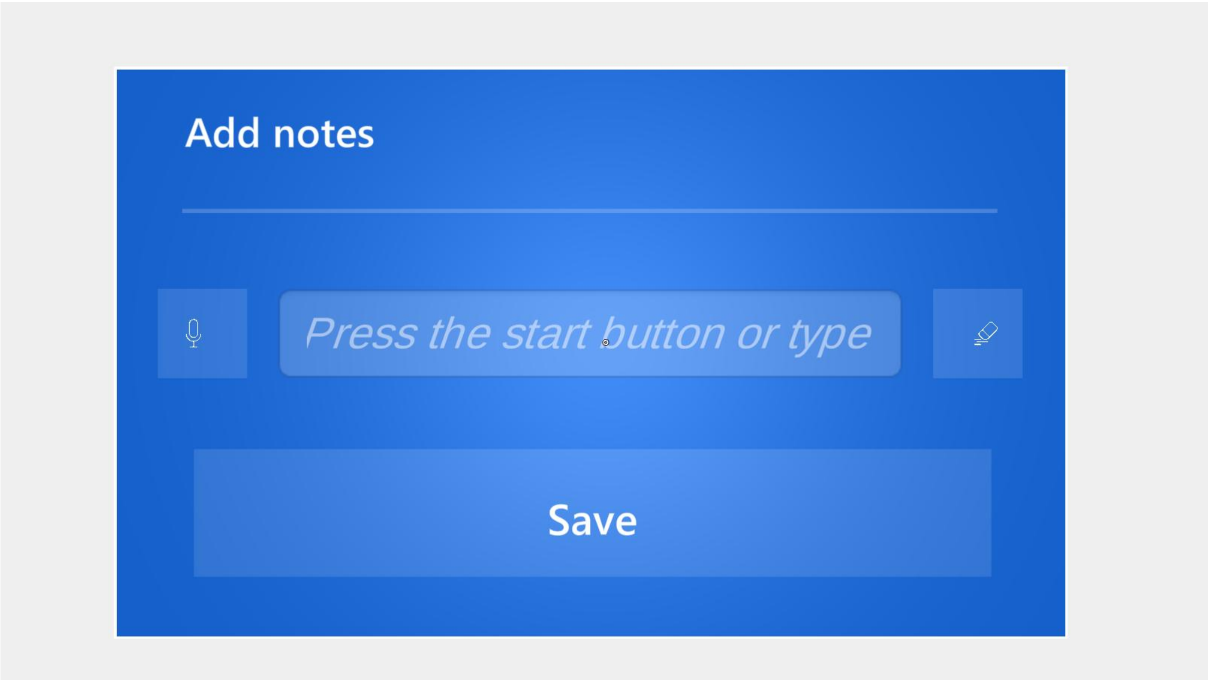


Figure 5.2: Menu used for adding new notes

### 5.1.1. Dictation system

As just stated, I allow the users to add words using a dictation system. This was done to make the function more usable on devices where the use of the system keyboard is unpleasant or slightly uncomfortable, enforcing the portability to multiple platforms that, as I already talked about plenty, is very important in XR.

I explored several service providers in search of the best compromise between delay from voice input to text output, complexity of the language model, memory usage, cost, ease of integration, and availability of the service on multiple devices. In the end, I was able to identify three possible candidates to choose from.

The first was the native dictation system of Windows-based devices, Windows speech APIs, which offered really fast and precise translations according to the testing I had done, without hindering the performance of the application. The main drawback of the system is its availability only on Windows-based platforms, making it unavailable on several of the devices I wanted to support. This is by far the easiest to access in an MRTK based application since it is the default implementation of the IMixedRealityDictationSystem, the interface used by all the components in the toolkit that require dictation.

The second system I analyzed was VOSK Offline Speech Recognition API, which offers several already trained free-to-use language models, that could then be further trained based on one's specific needs. While the other speech recognition systems all require an internet connection to run, this one offers instead some light models that can be then used directly on an offline device. The main drawback of this solution is the fact that said models, while relatively small, still require around 300MB of RAM to run while also compromising the precision of the translation. Alternatively one could also host one of the bigger and more complete models on a server and then access it through a custom-built API, effectively running a private version of a speech-to-text service, removing the restriction on RAM and precision of the translation.

The last one is the service that is also used by the Voice SDK of the Oculus Integration SDK created by Meta, called Wit.ai. This is again an online service that can then be integrated into a Unity app through the aforementioned Voice SDK. Although slightly slower than the Microsoft alternative, it offers comparable performances under all other metrics. Where it shines however is the possibility to be used on any device that Unity can deploy to.

Although very interesting, especially for its offline capabilities and its ability to be customized, I decided to not go with the VOSK API due to the slightly higher integration requirement, and the tool that the offline models would take on devices with low RAM like the Hololens2. It is however an option that I will keep on my radar, especially because of its low-cost nature, requiring only the amount of money necessary to run a server to host the language model and the API to access it.

Initially, I decided to go with a combination of the Windows speech APIs, which would be used on all Windows-based devices to achieve the best possible performances on those systems, together with the Wit.ai service, which would be instead used on all other devices, giving the system complete coverage on all supported devices. However during testing the Window service became unavailable for several days, prompting me to switch to the other service on all devices. Since the majority of the testing happened with this configuration, and the results were promising, I decided to keep it for the foreseeable future.

Everything I just talked about for the dictation system was done to ensure its availability on as many devices as possible, in order to combat the Dissemination problem that I exposed during the discussion on the State of the Art of XR in Education form section

2.2. It was also done in order to let it be used inside other parts of Tinalp, like, for example, in a possible future conversational agent.

## 5.2. Votable Note

The next step that I need to trace to get to the Word Cloud itself is an evolution of the concept of note that would allow users to express some sort of consent to its content. The direction I decided to take is the concept of a votable note. As the name suggests, each note is associated with a way to keep track of how many votes it has received by the users. The choice to use the concept of vote also opened up the possibility to associate meanings other than simple consent to it, like, for example, how much each user thinks that the content of a note has priority over the others. This may be useful in situations like the one in which the content of various notes represents some tasks to do for a project, which is something that comes up often when working using the agile methodology.

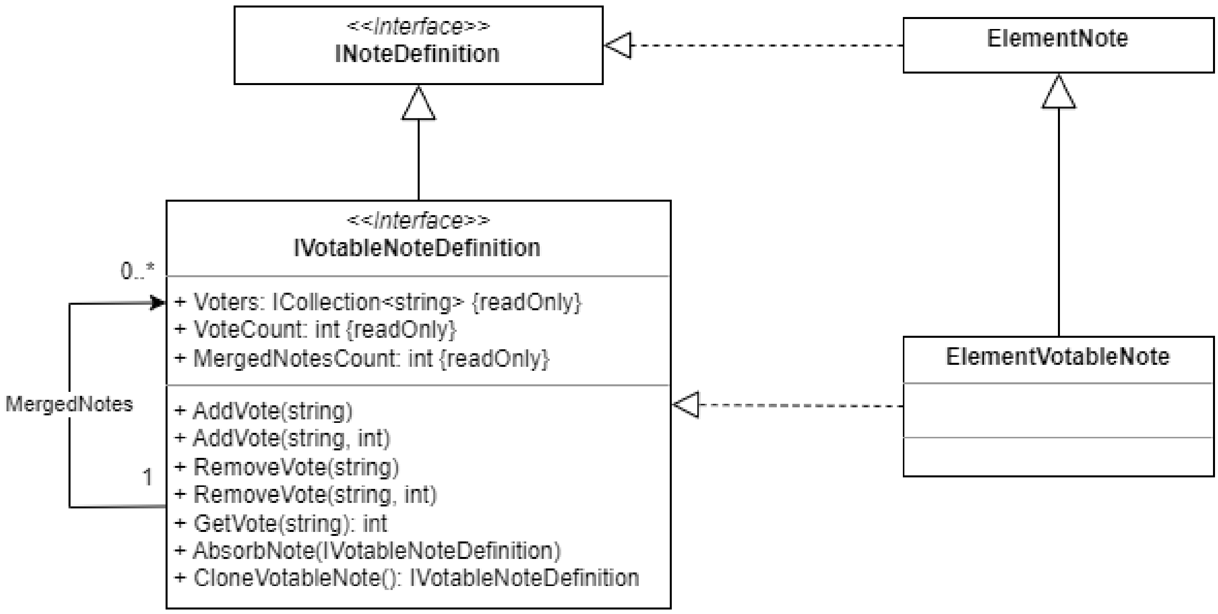


Figure 5.3: Votable Note Data Model - UML

Again the characteristics of a votable note are defined in the interface IVotableNoteDefinition for the same reasons stated in the previous section, which was then implemented with the class ElementVotableNote. This interface offers all the necessary attributes and methods both to keep track of and modify the number of votes received by each user and to merge two votable notes.

For what concerns the first set of functionalities, I decided to expose both the total number of votes and the votes cast by each user(through the method GetVote(string): int) together with a list of all users who voted the note. I also decided to make these values modifiable only with methods that can just add or remove votes, rather than setting them directly. This last choice was made keeping in mind that this data structure would have been used in a multi-user environment. Some of these modifying methods might in fact be called out of order on different clients, depending on how the backend that is responsible for communication implements message ordering. Using only methods that add or remove votes, I can implement a form of eventual consistency in which slight inconsistencies might show for a limited period on the vote counts (caused for example by the fact that some messages might be received in different orders on various devices), before eventually sorting themselves out once all the devices have received all the messages.

I opted to implement three voting modes to cover a wider range of use cases:

- Single: every user can vote every note at most once
- Multiple: every user can vote each note as much as he'd like
- Token: every user has a set total number of votes that he can distribute between the various notes as he pleases.

I also decided that every time a votable note gets created, it also automatically receives a vote from its creator. This is mostly relevant in the single voting mode since it is only natural to assume that a word that someone decided to add also has their vote. The second set of functionalities instead aims at giving the ability to absorb other votable notes, gaining their votes and contents. A copy of both the original note and the absorbed notes are also kept and made accessible.

Together with this data model, I also designed a menu that allows the users to access all the information and perform almost all the operations that I just described. The only one that this interface does not provide is the act of merging two notes, which has to be implemented in other ways, as it happened in the case of the Word Cloud. Instead, I decided to give the option to delete one of the notes from this menu.

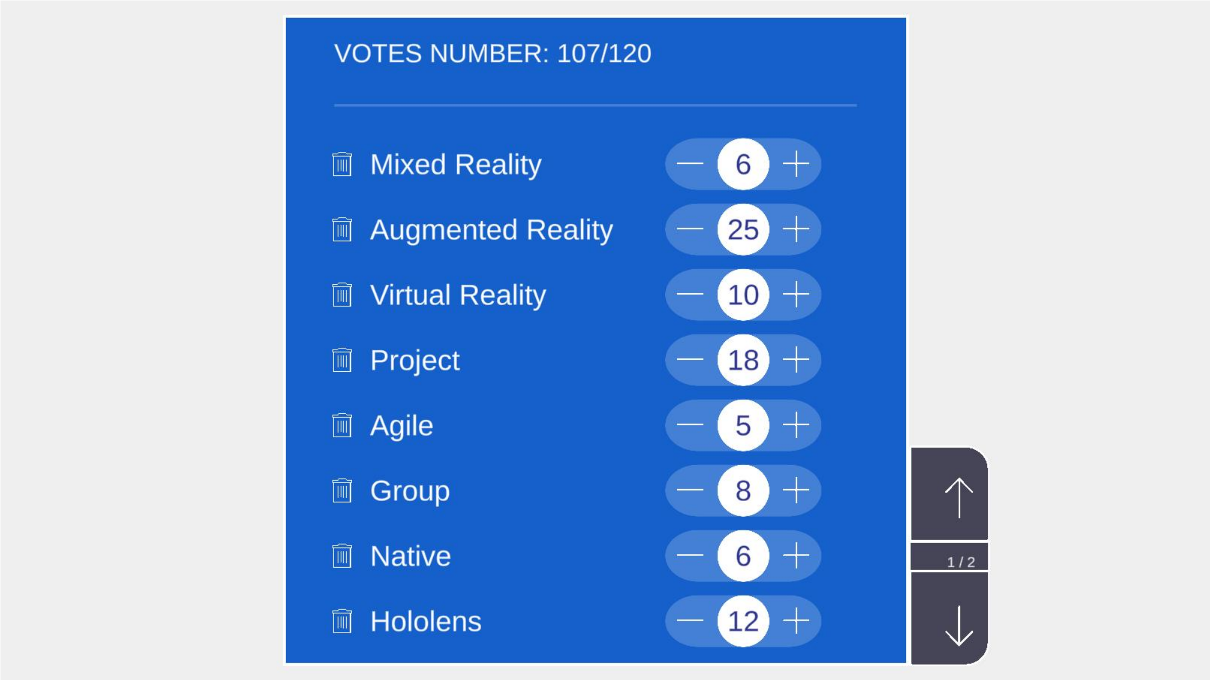


Figure 5.4: Menu used for voting

## 5.3. Word Cloud Implementation

Finally, I can show the implementation of the Word Cloud itself:

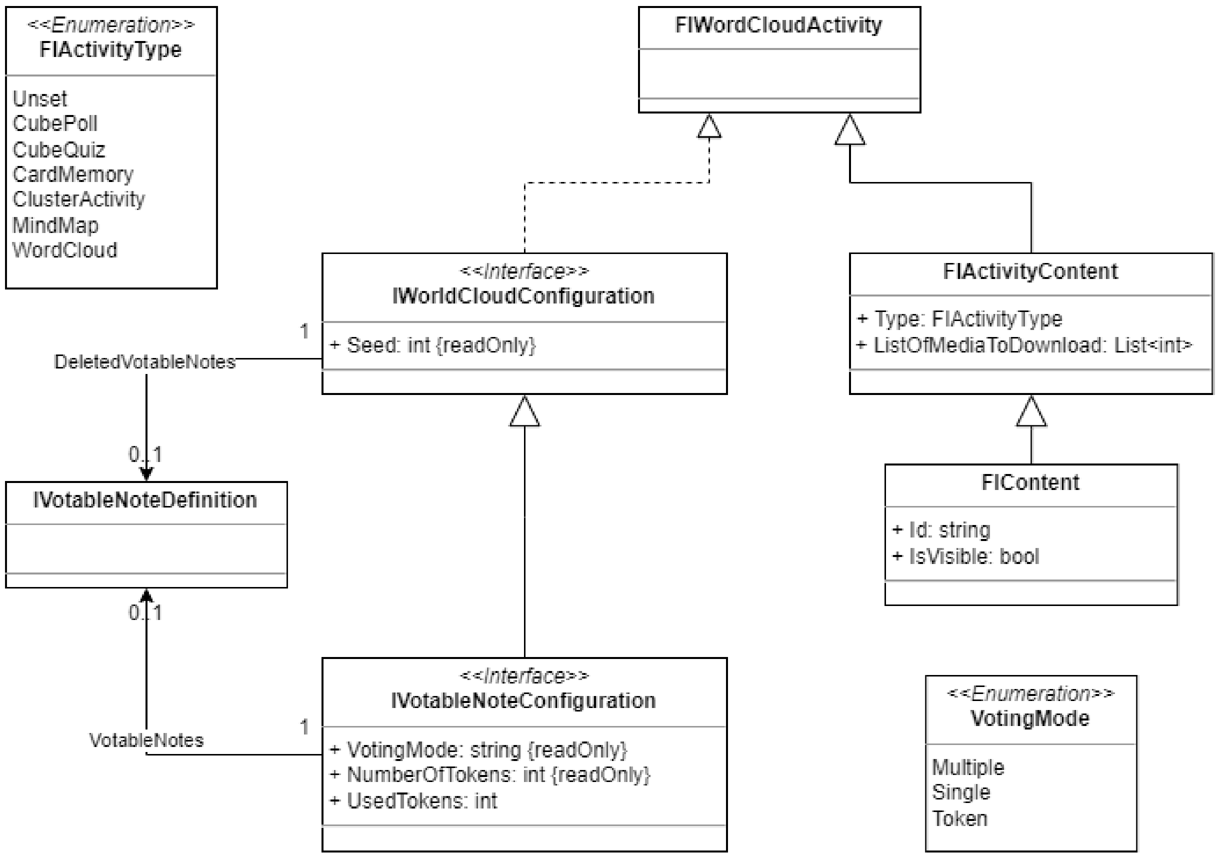


Figure 5.5: Word Cloud Data Model - UML

Focusing first on the data model, it can be seen that FIWordCloudActivity, which is the class that contains all the data necessary for describing the activity, inherits from FIActivityContent. This is the base class used by Tinalp to identify all the activities that are distinguished from each other through the Type attribute. It also contains the list of identifiers of all the media that should be downloaded for the activity ready to be used by the component managing the downloads from the back end. Finally, it inherits from FIContent, another Tinalp class that instead is responsible for identifying all types of content, a string that works as its unique identifier, and a boolean that indicates whether the content should be visible, which in the case of activities, for now, is always set as true.

To make the Word Cloud as context-agnostic as possible, I again used an interface to contain all the necessary data, to easily allow for future implementations of the Word Cloud in other situations. Then I also decided to split the content into two interfaces, where the second inherits from the first: one contains all the data necessary for a generic activity that utilizes votable notes, while the second adds all the rest of the data necessary for the Word Cloud.

Starting from the content of the first interface, there are:

- VotingMode: an instance of an enumeration indicating the voting mode to be used;
- NumberOfTokens: the number of votes each user has at its disposal when the selected mode requires to keep track of it;
- UsedTokens: The number of tokens that the user already used to vote when the selected mode requires to keep track of it;
- VotableNotes: A list containing all the words currently present in the Word Cloud.

Meanwhile the second adds:

- Seed: The seed used for the random number generator is used to compute certain parameters in the position of the words. More on this when I will talk about the positioning of the spheres in the Word Cloud;
- DeletedVotableNotes: a list containing all the deleted words from the cloud.

As I already mentioned in the design chapter (4.2), we decided to represent the words in the cloud as spheres containing text. Going into more detail on how this was technically achieved, the spheres themselves are created using a simple spherical model that is rendered using front-face culling algorithms, rather than the usual back-face ones. The final effect obtained is of objects in the shape of carved bowls that have the opening always facing the user. In the center of this "bowl" I then decided to place the text with an added MRTK component that always makes it face the user too. Then, as mentioned in the design section, I added to the sphere a moving hold animation. In this case, it consists of three sinusoidal independent movements along the three axes with different periods and amplitude smaller than 10% of the diameter of the sphere. This results in a continuous small oscillatory movement.

The spheres are then arranged into a specific topology inside the cloud to ensure that from any point of view, a high percentage of them is visible and every sphere is visible from enough angles, as per our design decision I explained and motivated in the previous chapter (4.2). However, before delving deeper into this topic I should introduce one more technical feature of Tinalp: during the application setup, users are required also to position a platform that functions as a fixed reference point for all other objects' positions. This can be done either by scanning a QR code, placing it manually, or downloading its position using a pin created during its upload. The platform is more often than not placed on top of some surface. Since the Word Cloud will be created on top of this platform, I would like to make sure that the spheres of the cloud won't end up under said surface. This will be reflected in the topology I implemented and that I will now describe. The spheres are first arranged along concentric orbits on the xy-plane centered on top of the Tinalp platform. Then an offset along the z-axis is randomly computed and applied for each sphere. The offset in absolute value cannot exceed the radius of the orbit the sphere occupies. Both the number of spheres per orbit and their diameter are dynamically computed based on the maximum size that the spheres can have and the amplitude of the moving hold oscillation to ensure that any two of them could never overlap. A further optimization was introduced to reduce the amount of space: bigger spheres are pushed further along the z-axis than smaller spheres, reducing the radius that each orbit must have to ensure no overlapping. Both the order in which the spheres are filled and the seed to use for the random number generator are defined in the data model, ensuring that the same Word Cloud has the same structure across both different sessions and different devices, and by extension different users' points of view. Although the topology is not three-dimensional itself, the illusion created through the random offset along the z-axis creates a pleasantly looking 3D Word Cloud as could be seen in Figure 4.2, while coming with the benefit of being pretty simple to understand and handle given its two-dimensional nature.

Moving the focus to the second phase of the activity, the main method of interaction with the spheres is grabbing them. This behavior is regulated by a component of MRTK called ObjectManipulator together with another one called NearInteractionGrabbable, whose role is to map the various ways to take hold of an object based on the interaction interfaces offered by each device to a grabbing behavior. When a user is holding two spheres at the same time it can merge them. While they are approaching a bolt of lightning forms between them to clearly indicate to all the users that something is about to happen. After they touch a merging animation begins where the bigger one absorbs the smaller one. All these behaviors are handled by a component of my creation called MergerController that keeps track of a pair of spheres grabbed by any user and handles both the creation, progression, and life-cycle of the lightning animation and the merging animation. While holding a sphere, a user can also decide to delete it. To do so they need to scale it down. After a certain size threshold parts of the spheres start to gradually disappear. The effect gets more prominent as the spheres get reduced in size. When it completely vanishes, the word is deleted from the cloud. If at any point the sphere gets released, it will be restored and scaled up to a minimum allowed size if smaller. The vanishing effect is achieved through the use of a shader that uses the alpha channel as a threshold: the part of the spheres with a value lower than the threshold are not rendered. This value is assigned to each point of the sphere using a texture as a map. The alpha channel is then updated based on the current scale of the sphere: a smaller one means a higher value.

Another important aspect of my implementation is the communication between users in the same session. The Word Cloud utilizes the same system used in the whole Tinalp application, which is a message broker that receives and relays messages either in broadcast or unicast. It also runs on the same assumption of reliable messaging as the whole of Tinalp. Although this assumption is not always true, in part due to the protocol used being UDP, in practical application this problem can be overlooked as shown by extensive testing. As a consistency model, I decided to use eventual consistency. This was achieved by implementing all the operations that require message exchange in a way that ensures the same result, independently from their order of application. For example, I can modify the votes only through add and remove operations rather than set operations. Although there can be moments where there are slight inconsistencies. This was in my opinion the best compromise between user experience and program efficiency. The only point in which the system ensures that all devices share the same state of the Word Cloud is in the passage from the first phase to the second one.

# 6 | Experimental Evaluation

In the following section, I will describe the evaluation sessions performed on the application described in the previous parts of this thesis. These evaluations were conducted within the actual operational environments where the software was intended for deployment, namely educational institutions and corporate settings. Additionally, evaluations were carried out in environments less conducive to the software's intended use cases. In the latter scenario, the emphasis primarily revolved around gauging the software's usability. Conversely, in the former context, the evaluations encompassed a dual objective: assessing usability while also soliciting feedback pertaining to the relevance and effectiveness of the developed activities within their intended contexts of use. These sessions will be presented in chronological order since the results of one evaluation round affected the redesign of the software in view of the next, which in turn influenced the tests. All of them were done in collaboration with some members of the ASP group of students who worked on MIRE and FifthIngenium.

## 6.1. Vodafone

The first situation in which we decided to let an external user test the application was during a meeting at the Vodafone Experience Center in Milan. At this gathering were five Vodafone employees from different sections like marketing, communication, and HR. Two of them are also extremely familiar with the agile methodology and they apply it often in their line of work. The main goal of this meeting was to present the whole MIRE project and in particular the Word Cloud as the first of the activities fully designed and implemented. It was done mainly to gather feedback on how effective it would be to use mixed reality as a tool to aid a group of people working following the agile methodology. To do so we decided to make the people present try the Word Cloud.

Going into more detail on how the activity was tested, the setup was the following: after a presentation about Tinalp by one of the senior employees of FifthIngenium and another about the MIRE project in general done by one of the ASP students, I presented the activity and its function to the listeners. Then the two people more familiar with the agile methodology were invited to take part in the Word Cloud activity using Hololens 2 in the same local session as me, who acted as the room owner and trainer. The experience was also streamed through the point of view of one of the participants on a screen, to make sure that the other presents at the presentation could see the activity unfolding. For about 20 minutes they were free to explore the first phase of the activity while everyone was encouraged to comment on it.

Unfortunately, due to the nature of the event, we were only able to gather informal comments on the activity, rather than administer surveys or ask pre-prepared questions. Nevertheless, the observations made by the people present at the meeting were quite insightful, especially regarding the application of this activity inside a company that adopts the agile methodology. The first things they commented on were the reservations they had before our presentation about the system in general, especially its application during workshops. In particular, they were worried about the inclusiveness in situations where several people needed to participate, utility, and effectiveness. However, they also told us that these concerns were cleared during the presentation. In particular, the possibility of using the application on mobile phones convinced them that the activities could also be done in situations where twenty to thirty people are required to cooperate without requiring an excessive number of visors, while the hands-on session convinced them of the potentiality of the application. They were also really pleased about the learning curve of the devices and activity. All of these factors led them to express the intention to test the activity in a future workshop. They also had several useful comments about functionalities they would like to see implemented. One of the said functionalities would be an automatic system that detects synonyms and, more in general, words that in different nomenclatures refer to the same concept, which is a situation that commonly presents itself, because more often than not people coming from widely different backgrounds participate in this kind of activities. Another feature they would like to have is the possibility to associate more content with the words in the cloud and access them by clicking on the corresponding sphere.

## 6.2. Politecnico di Milano Open Day

The second context in which I, with the assistance of the team leader of the ASP group, tested the application was the Open Day at Politecnico di Milano, which is an event where future students who are considering enrollment at the university can come and attend presentations on the available degrees, departments and the Politecnico in general, while also going around campus and visiting the booths hosted by current students and faculty members of degrees they are interested in. There they can interact with said hosts and catch a glimpse of their possible future if they decide to enroll. We were stationed at the "Ingegneria Informatica and Computer Science and Engineering" booth where we made interested future students try the Word Cloud. The goal of this experience was mainly to evaluate the usability of the application, rather than its utility in its main domain of application. The activity was in an almost complete state at that point. It was only missing the moving hold animation for the spheres.


Figure 6.1: Photo of the experience at the Open Day

As for the testing in itself, users were given a Hololens 2 to access the activity with. They had 10 minutes to perform various actions on the Word Cloud. After being instructed on how to interact with the MR device using their hand, they were left free for 5 minutes to add and vote words. After that, they were instructed to switch the activity to the second phase and were encouraged to interact directly with the spheres and to perform various actions on them, like merging two words or resizing some of them. The choice of the device was purely logistical since it was the easiest device to pass around, not having any external unit attached to it. Due to the long nature of the event (approximately 8 hours), we had three devices, two available to the interested people, while one was charging and ready to be swapped with out-of-power ones, as to have always two ready-to-use devices at all times. We also decided to make each user interact with a Word Cloud of its own, rather than making all of them take part in the same session. This was done to make each user more independent in the exploration of the activity, without being constrained to the phase that the activity is currently in inside the shared session, but allowing them to pass from the first phase to the second one when they were ready. It also facilitated for us the passage of the device between people, making it independent for each device, and giving us the possibility to rest the activity to a default state before giving it to the next tester. Speaking of this default state, it consisted of a Word Cloud in the first phase of the activity (adding and voting of words) with already some words inserted and voted. First, we explained to them how to interact using their hands, since for almost all of them it was the first time using this kind of device and participating in this kind of experience. Then we let them add new words to the cloud using either the keyboard, the dictation system, or both, and vote the word present. Then, we asked them to switch to the second phase and interact with the words themselves, moving resizing, and merging them. Each user tested the application for an average time of 10 minutes.

At the end of their experience, each user was asked to submit an online questionnaire on the usability and perceived usefulness of the activity to give us feedback on their experience. The first 10 questions were taken from the System Usability Scale (SUS) while the remaining 3 were designed by us to get more specific opinions on the Word Cloud itself. To the first 11 questions, participants could answer on a scale from one, equivalent to strongly disagreeing with the question, to five, equivalent to strongly agreeing. The second to last question instead had some customized recompiled answers, while the last one was openhanded. In total, we were able to gather 82 responses obtaining the following results:

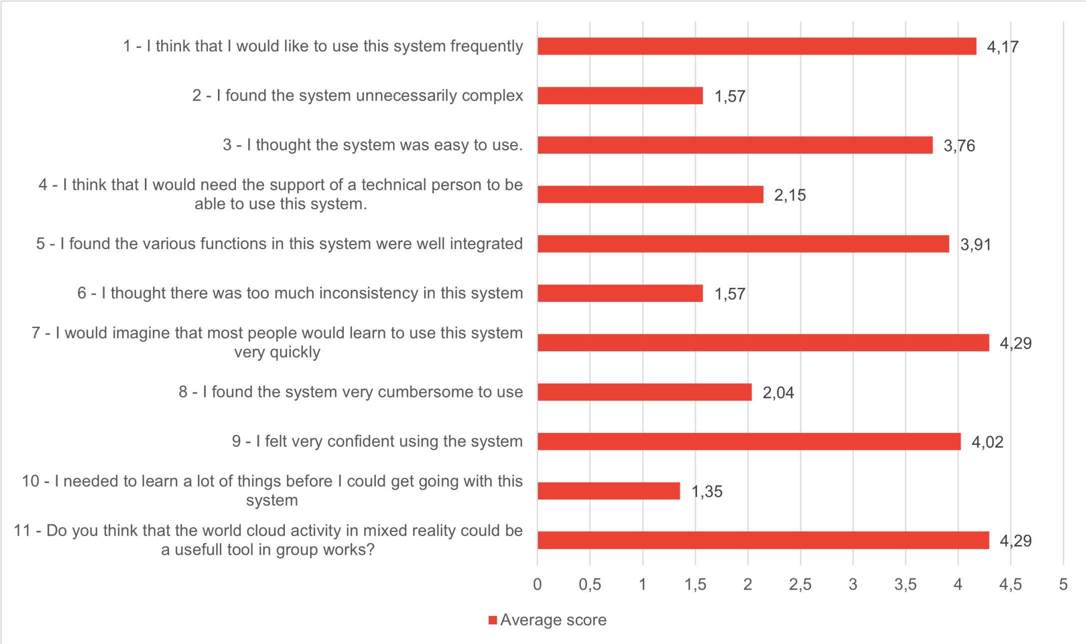


Figure 6.2: Open Day survey result: first 11 questions

Overall the feedback gathered from these questions was largely positive, although with some drawbacks being highlighted as well:

- the high score on the last question shows us that participants believe in the utility of the activity in its conceived use cases;
- the overly positive result in the first question shows us that participants are willing to use the activity often;
- the low perceived quantities of information needed to learn to use the system, makes it quite easy to pick up;
- The system however still seems to be cumbersome and not so easy to use at times, resulting in the belief that support from a technical person may be required at times. This, combined with the findings highlighted at the previous point, makes me believe that there may be still some usability issues related to the implementation of the method of interactions that need to be explored.

The second to last question asked users whether they thought that mixed reality brought advantages compared to the same activity in a 2D version on common devices like tablets, computers, or whiteboards:

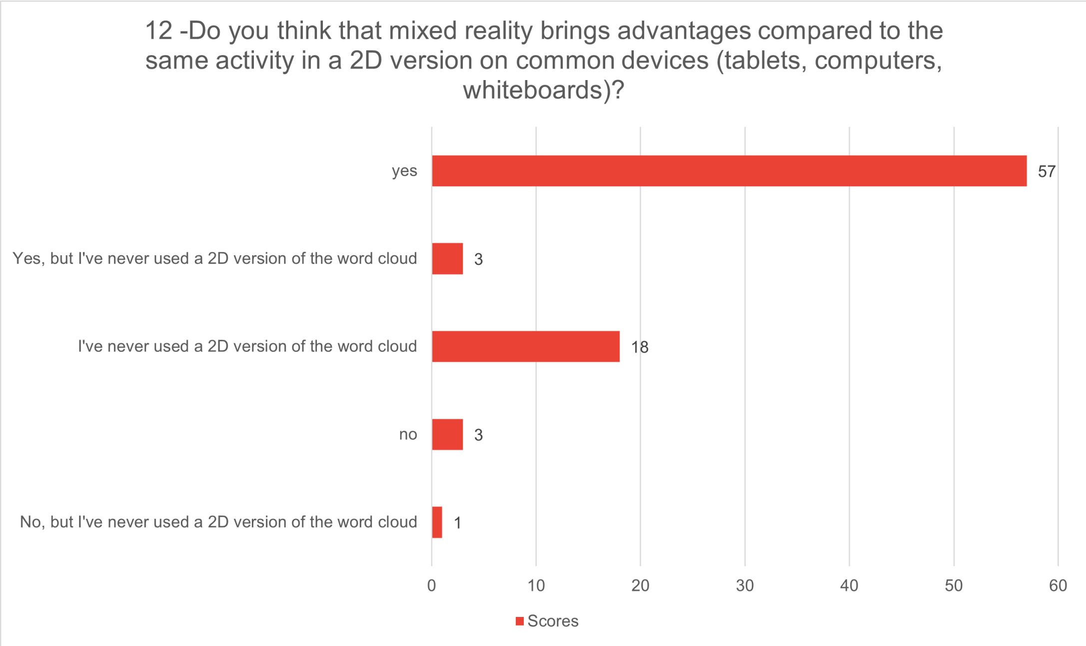


Figure 6.3: Open Day survey result: question 12

As can be seen, 73% of the opinions were favorable to the idea, while just 5% was against it.

Lastly, I would like to report what I could gather from the comments made in the last open question. Users were asked which advantages and/or critical issues they saw in the Word Cloud experience in mixed reality. From the answer, we could gather that:

- it was perceived as useful for work, also in situations with a lot of people involved and even in remote;
- the three-dimensional instead of two-dimensional ambient helps to improve ideas visualization and gives more freedom, especially during the discussion;
- the Integration with the external world is definitively an advantage;
- it resulted initially complicated to use, but easy to become accustomed to, especially if guided during the first use, even if not always necessary;
- the high involvement that the medium promotes was really appreciated;
- the intense sunlight caused some problems for some users.

Overall, the results were largely positive, especially when talking about the perceived usefulness of the activity and its design. A slightly worse performance under the usability aspects related to the implementation prompted me to improve on some aspects of it.

## 6.3. Liceo Classico Parini

The last situation in which I was able to perform an empirical evaluation of my work, was during a series of four encounters held at the high school "Liceo Classico Parini" in Milan. The practical parts of these encounters were handled by me, a group of employees of FifthIngenium, and the team leader of the group of students from ASP. The main goal of this sessions was to test the acceptance and the perceived usefulness by the Word Cloud intended users, i.e. students, in its real context of use. Usability aspects were investigated as well, and feedback was gathered on the devices and their use in a school environment. The Word Cloud at this point was in a complete state, presenting also some minor improvements on the usability born from the feedback I received during the Open Day (the testing session discussed in the previous section). A class of 24 students between 16 and 17 years old participated in the encounters. It is notable that these students do not follow a curriculum with a high density of STEM. subjects, but rather one focused on the study of literature (from ancient Greece and Roman literature to Italian and English), history, and philosophy.


Figure 6.4: Photo of the experience at the Liceo Parini

This testing session was organized as a full educational experience, in collaboration with the Italian, history, and English teachers of the students involved. The main theme of the experience was the birth of nation-states in Europe, based on four interviews with experts on the topic in English. The objective was to stimulate the learning of the subject and the ability to critically think about a topic that has relevant consequences on contemporary history. The teachers first introduced the topic to the students in the classroom, who then were asked to autonomously study the aforementioned documents. Then we proceeded with the four encounters, which were structured in the following manner:

- During the first encounter a presentation on the realm of extended reality, the devices that enable it, and its application, bringing Tinalp and the rest of the applications produced by FifthIngenium as an example, was given to the students. Then the students were allowed to try out the devices by letting them see and interact with some 3D models. Between this and the next encounter, students had a two-week time frame to study two interviews on the topics they would perform the activities in the following sessions.

- In the second encounter, we decided to divide the students into 4 groups in order to create sessions of at most 6 students, giving a more controlled environment and the possibility for everyone to meaningfully contribute to the experience. We also needed this division because we had available only 12 devices to let the students use. Two of the groups started in the classroom, where they learned how to use the web editor to upload and create content for Tinalp. The other two groups instead were given the devices and two sessions, one per group, were created. There the students tried three activities, all themed around the birth of nation-states in Europe: a card memory, where students had to associate each nation's coat of arms with their borders; a cluster activity where they had to identify whether a concept is of relevance for the definition of nation, state or both; a cube quiz where they had to associate the various historical events to the date they happened in. After 30 minutes the groups switched: the ones with the devices went into the classroom and vice-versa. After another 30 minutes, we finished by giving each student an online questionnaire to complete on their opinions towards the devices they've used.

- In the third encounter we divided the students again into 4 groups: the two in the classroom were tasked to first upload all the necessary content on the web editor to then create a mind map with a Hololens2 on any argument of their choosing. Meanwhile, the two groups with the devices had two tasks to complete. First, they had to use the Word Cloud to discuss the main concept of the topic of Nation-States. After that, they had to answer 10 questions on a survey inside Tinalp on the same topic. Again, after 45 minutes, the groups switched, and after another 45 minutes, we gave the students the same questionnaire as the second encounter.

- During the last encounter instead we simply asked for the students' feedback, both in the form of answers to another online questionnaire we gave them and in the form of some comments and observations they wanted to freely share with us.

The devices we decided to bring to let the students use were six Hololens 2, two Magic Leap 2, two Meta Quests Pro a smartphone, and a tablet. During all of the encounters, we encouraged the students to try out as many different devices as possible, to let them form an opinion on most of them, and to let us know which one they preferred and why.

First I will examine the results of the final questionnaire on the opinion of the students on the activity, where they were asked to score certain aspects of the experience on a scale from 1 to 5. The results out of the 21 answers we gathered were the following:

| Question | 1 (at all) | 2 | 3 | 4 | 5 (very much) | Mean | Mode |
|---|---|---|---|---|---|---|---|
| Word visualization usability | 0 | 4 | 8 | 8 | 1 | 3.3 | 3 |
| New word insertion usability | 1 | 4 | 8 | 8 | 0 | 3.1 | 3 |
| Menu usability | 2 | 7 | 8 | 3 | 1 | 2.7 | 3 |
| How much did you appreciate collaborating the word cloud creation in Mixed Reality? | 1 | 6 | 10 | 3 | 1 | 2.9 | 3 |
| How much did you appreciate collaborating with fellow students? | 0 | 1 | 5 | 7 | 8 | 4.0 | 5 |
| Do you feel that the Mixed Reality experience was a stimulus for learning content? | 2 | 8 | 8 | 1 | 2 | 2.7 | 3 |

Figure 6.5: Liceo Parini survey result

From these scores, I could extrapolate that:

- 42.9% of the students agreed that the words were clearly visible and readable, while only 19.0% disagreed;
- 38.1% had a positive experience with word insertion, and another 38% had a neutral one;
- most students had a negative(42.9%) or neutral(38.1%) experience with the menu;
- most students appreciated collaboration in MR, with an average score of 4.0, however, they felt slightly worse on collaboration in the context of the Word Cloud, scoring only 2.9 on average.
- 47.6% of students did not felt that MR was a stimulus for learning content
- the most common answer in almost all the questions was the average one(3).

The open question instead showed comments similar to the ones from the Open Day highlighting that:

- the activities are nice, but they are not immediately easy to use the first time, even if they are intuitive;

- sometimes objects were hard to grab or menus were hard to interact with, especially buttons;
- the third dimension improved upon the ability to remember the experience and the content;
- it took too much time to perform the activities.

Finally, from the survey on the devices, we gathered the following data:

| **How intuitive was the use of the worn device?** | **1** | **2** | **3** | **4** | **5** | **average** | **participants** |
|---|---|---|---|---|---|---|---|
| Microsoft Hololens 2 | 0% | 16% | 31% | 41% | 13% | 3.5 | 32 |
| Magic Leap 2 | 0% | 23% | 8% | 62% | 8% | 3.5 | 13 |
| Meta Quest Pro | 0% | 0% | 11% | 33% | 56% | 3.4 | 9 |

Figure 6.6: Liceo Parini device survey result

The students were also asked whether the experience with the device was comfortable, and they answered:

- 69% yes and 31% no for the Microsoft Hololens 2;
- 46% yes and 54% no for the Magic Leap 2;
- 44% yes and 56% no. for the Meta Quest Pro;

The device that performed the best was the Hololens2. To shed light on why that is, we can look at the answers that students gave to the open questions a the end of the survey and during the discussion on the final encounter. While talking about the Magic Leap 2 they said that after a while it caused headaches and eyestrain, meanwhile, when talking about the Meta Quest Pro, they said that it caused headaches and nausea and that it pressured their temples.

After this testing session, I decided to perform some minor improvements to the Word Cloud to directly address some of the problems raised by the testers in both this and the Open Day session. The main change that I made was the improvement in the interaction with the menu, by adjusting the distance between buttons and the hitboxes of several menu components. Hopefully, this is enough to solve the usability issues raised during the sessions, however, I believe that a little more testing is required.

# 7 | Conclusions

This thesis consisted in the design, implementation, and testing of an MR application to support collaborative brainstorming. It was part of a larger effort, involving two companies, a university lab, and a group of students enrolled in an honors program. My task was to design, develop, and test a Word Cloud in MR, allowing students, under the guidance of a teacher, to debate ideas. The main requirements were:

- create a way to represent concepts instead of objects to give the option to use XR for subjects outside of STEM and that do not study concrete objects;
- design and implement the activity in a way that gives the teacher back the role of curating the pedagogical aspect of the experience, thanks to the customization and flexibility of the activity;
- develop and deploy the solution to as many devices as possible using frameworks like Unity and MRTK to reduce the lack of dissemination of MR applications;
- reach a high degree of usability and accessibility through the technique explained at the end of section 4.2 and improve upon them after usability testing sessions;
- utilize animations and visual elements to improve feedback and make the virtual objects more look-alive.

The Word Cloud underwent a first mild testing by industry experts and two rounds of testing with users, one at the Polimi Open Day and the other, quite more substantial, in the frame of an actual educational experience at the Parini High School in Milano. The results seem to suggest that the introduction of MR elements in educational experiences can have a positive impact. In particular, especially the last evaluation sessions showed how an experience with innovative technology can be paired with traditional educational activities to boost interest and stimulate students.

The main lessons learned from this work are:

- since most users are still new to MR, applications for the technology should be even more intuitive than ones for traditional mediums (smartphones, computers, ...), to offset the challenges that users face when also trying to learn the new interaction paradigms

- it's important to strike the right balance between allowing and forbidding certain actions to users especially in a collaborative context with a lot of participants as to not make it feel like they are not contributing, while also maintaining some order
- the teacher or project leader's ability to exert a certain degree of control over the experience should be preserved
- the association between virtual objects and reality, and the possibility to interact with a full three-dimensional environment given by MR, improves the ability to represent, organize, and remember ideas

A publication titled "Supporting Project Work with Mixed Reality" [4], in which I am a co-author, was produced as a result of this project. It was presented as a full paper during the EDMEDIA conference in Vienna in 2023.

## 7.1. Future works

There are several directions from which to approach future works around what is presented in this thesis.

The Word Cloud itself could be further expanded by adding new functionalities including

- the possibility to reverse a merge of two words;
- a timer that is set by the trainer at the end of which the activity switches to the second phase;
- a view of the Word Cloud where the spheres are ordered based on the number of votes;
- the option to associate a description or other type of content (like pictures or models) to words in the cloud;
- a filter that recognizes words with the same meaning, to be used when adding new words to the cloud.

Of course, these new features should also be evaluated with new testing campaigns.

Another angle from which to tackle future development is the refinement of the design and the implementation of other activities from the MIRE project that are already planned like one for flow chart creation, an activity for debates, or "Loci Ciceroniani", which is an activity where the participants create a strong association between a concept represented through a virtual object and a real object or location to improve that concept memorization.

# Bibliography

[1] D. Allcoat, T. Hatchard, F. Azmat, K. Stansfield, D. Watson, and A. von Mühlenen. Education in the digital age: Learning experience in virtual and mixed realities. *Journal of Educational Computing Research*, 59(5):795–816, 2021.

[2] S. Bauro, C. Campanella, L. Corinaldesi, N. Di Blas, M. T. Faga, C. F. Arnau, G. Pinotti, and M. Valoriani. Mixed reality in education, for any school subject. In *EdMedia+ Innovate Learning*, pages 388–392. Association for the Advancement of Computing in Education (AACE), 2022.

[3] S. Bauro, C. Campanella, L. Corinaldesi, N. Di Blas, M. T. Faga, C. Franch-Arnau, G. Pinotti, and M. Valoriani. Mixed reality in education: Towards a new paradigm. In *Society for Information Technology & Teacher Education International Conference*, pages 1589–1596. Association for the Advancement of Computing in Education (AACE), 2022.

[4] S. Bauro, L. Bianchini, W. Brattelli, C. Campanella, G. Concialdi, L. Corinaldesi, M. Di Marco, N. Di Blas, M. T. Faga, A. Fasiello, et al. Supporting project work with mixed reality. In *EdMedia+ Innovate Learning*, pages 545–555. Association for the Advancement of Computing in Education (AACE), 2023.

[5] D. Beyoglu, C. Hursen, and A. Nasiboglu. Use of mixed reality applications in teaching of science. *Education and Information Technologies*, 25:4271–4286, 2020.

[6] M. Billinghurst and H. Kato. Collaborative augmented reality. *Communications of the ACM*, 45(7):64–70, 2002.

[7] P. Chen, X. Liu, W. Cheng, and R. Huang. A review of using augmented reality in education from 2011 to 2016. *Innovations in smart learning*, pages 13–18, 2017.

[8] M. Di Marco. Mixed reality in education: a new approach beyond 3d models, 2022.

[9] D. Fonseca, E. Redondo, and S. Villagrasa. Mixed-methods research: a new approach to evaluating the motivation and satisfaction of university students using advanced visual technologies. *Universal Access in the Information Society*, 14:311–332, 2015.

[10] J. Garzón. An overview of twenty-five years of augmented reality in education. *Multimodal Technologies and Interaction*, 5(7):37, 2021.

[11] M. Leap. Magic leap 2. `https://www.magicleap.com/magic-leap-2`.

[12] Meta. Meta quest tech specs. `https://www.meta.com/it/en/quest/quest-pro/tech-specs/`.

[13] Microsoft-Corporation. Hololens 2 hardware. `https://learn.microsoft.com/en-us/hololens/hololens2-hardware`, .

[14] Microsoft-Corporation. What is mixed reality? `https://docs.microsoft.com/en-gb/windows/mixed-reality/discover/mixed-reality`, .

[15] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[16] C. Pidel and P. Ackermann. Collaboration in virtual and augmented reality: a systematic overview. In *Augmented Reality, Virtual Reality, and Computer Graphics: 7th International Conference, AVR 2020, Lecce, Italy, September 7–10, 2020, Proceedings, Part I 7*, pages 141–156. Springer, 2020.

[17] M. Prince. Does active learning work? a review of the research. *Journal of engineering education*, 93(3):223–231, 2004.

[18] I. Radu. Augmented reality in education: a meta-review and cross-media analysis. *Personal and ubiquitous computing*, 18:1533–1543, 2014.

[19] I. XREAL. Xreal experience | immersive entertainment experience. `https://www.xreal.com/experience/?ar-space`, .

[20] I. XREAL. Xreal light. `https://www.xreal.com/light/`, .

# List of Figures

# Acronyms

**XR** Extended Reality
**AR** Augmented Reality
**VR** Virtual Reality
**MR** Mixed Reality
**AV** Augmented Virtuality
**VC** Virtuality Continuum
**MRTK** Mixed Reality Toolkit
**AI** Artificial Intelligence
**WoS** Web of Science
**MIRE** Mixed Reality in Education
**ASP** Alta Scuola Politecnica
**PM** Project Manager
**SUS** System Usability Scale
**STEM** Science, Technology, Engineering and Math

# Acknowledgements

I'd like to thank my family and friends, who stayed by my side during my academic years. I'd also like to thank Professor Gribaudo and Professor Di Blas for the opportunity to work on this project and the support shown throughout. I'd like to thank Matteo Valoriani and the whole team at FifthIngenium for their help shown from the first day. Finally, I would like to thank all the people and organizations involved in the MIRE project, especially the group of students from ASP, for letting me contribute to it.