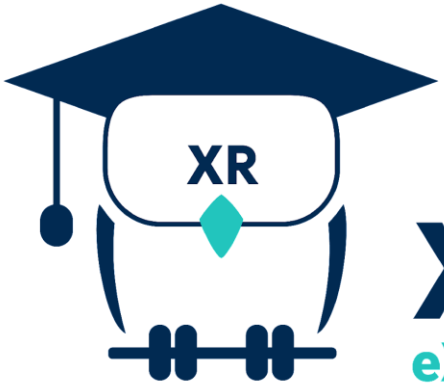
Ref. Ares(2025)11587491 - 24/12/2025

**Horizon Innovation Action**
HORIZON-CL4-2022-HUMAN-01-19




XR4ED
eXtended Reality for EDucation

# D5.4
# Evaluation and troubleshooting

**DISCLAIMER**
This document is issued within the frame and for the purpose of the XR4ED project. This project has received funding from the European Union's Horizon Europe Framework Programme under Grant Agreement No. 1010931159. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the European Commission. The dissemination of this document reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains. This deliverable is subject to final acceptance by the European Commission. This document and its content are the property of the XR4ED consortium. The content of all or parts of this document can be used and distributed provided that the XR4ED project and the document are properly referenced. Each XR4ED partner may use this document in conformity with the XR4ED Consortium Grant Agreement provisions.

Funded by
the European Union

# XR4ED Project Profile

| | |
|---|---|
| **Project title** | eXtended Reality for EDucation |
| **Project acronym** | XR4ED |
| **Grant agreement No.** | 1010931159 |
| **Start date of project** | 1 January 2023 |
| **Duration of project** | 36 months |
| **Project website** | http://www.xr4ed.eu/ |

## Partners

| | | | |
|---|---|---|---|
| CYENS CENTRE OF EXCELLENCE | CYENS | RESEARCH | CYPRUS |
| konnektable TECHNOLOGIES | KONNEKTABLE | SME | IRELAND |
| 21 Sci science 27 Co communication | SciCo | SME | GREECE |
| F6S | F6S | SME | IRELAND |
| INFINITIVITY DESIGN LABS | Infinity Design Labs | SME | FRANCE |
| KAINOTOMIA | KAINOTOMIA | SME | GREECE |
| NUROGAMES | NUROGAMES | SME | GERMANY |
| VASS complex made simple | VASS | INDUSTRY | Luxemburg |
| SIMAVI | Software Imagination & Vision SRL | SME | ROMANIA |
| UPC | Universitat Politecnica de Catalunya | RESEARCH | SPAIN |
| CELLOCK | CELLOCK Ltd | SME | CYPRUS |

Funded by the European Union

# Document information

| | | |
|---|---|---|
| **Related work package** | WP5 | |
| **Document reference** | D5.4 | |
| **Related deliverables** | D2.3, D5.2, D5.3 | |
| **Type** | | DEM - Demonstrator, Pilot, Prototype |
| | | DMP - Data Management Plan |
| | ✓ | R - Document, Report |
| **Status** | Final | |
| **Version** | 1 | |
| **Due date** | 31/12/2025 | |
| **Submission date** | 24/12/2025 | |
| **Lead partner** | UPC | |
| **Contributing partners** | ALL | |
| **Reviewers** | Fotis Liarokapis (CYENS), Maria Poveda (CYENS), Epameinondas Koutavelis (KT) | |
| **Keywords** | | |

| **Dissemination level** | |
|---|---|
| ✓ | PU: Public |
| | SEN: Sensitive |

| **List of contributors** | |
|---|---|
| Name | Partner (short name) |
| CYENS CoE | CYENS |

# Revision History

| Document History | | | |
|---|---|---|---|
| Version | Date | Change editors | Changes |
| 0.1 | 15/11/2025 | UPC | First Draft |
| 0.2 | 12/12/2025 | UPC | Second Draft |
| 0.3 | 18/12/2025 | KT | Review and sugestions |
| 0.4 | 23/12/2025 | UPC | Final draft |
| 1 | 24/12/2025 | CYENS | Review and submission |

| Quality control | | |
|---|---|---|
| Role | Partner (short name) | Approval date |
| Deliverable leader | UPC | 15/12/2025 |
| Quality manager | UPC | 23/12/2025 |
| Project coordinator | CYENS | 24/12/2025 |

# Table of Contents

XR4ED Project Profile ..... 2
Partners ..... 3
Document information ..... 4
Revision History ..... 5
List of Figures ..... 8
List of Tables ..... 8
List of Acronyms ..... 9
Executive Summary ..... 10
1 Introduction ..... 10
1.1 Overview ..... 10
1.2 Structure of the Document ..... 11
2 User Experience ..... 12
2.1 Avatar interaction metaphors ..... 12
2.1.1 Interaction Metaphors User study ..... 14
Participants ..... 14
2.2 LLM-powered Embodied Conversational Agents for AR ..... 18
2.2.1 Application Modes ..... 18
2.2.2 Qualitative Evaluation ..... 20
2.2.3 Limitation and Conclusions ..... 22
2.3 LLM-based Embodied Conversational Agents enhanced with bi-directional multi-modal communication for XR (BIRDS) ..... 24
2.3.1 BIRDS experiment design ..... 26
2.3.2 BIRDS user study ..... 27
2.4 Conclusions regarding user experience ..... 32
3 Evaluations ..... 33
3.1 Developers ..... 33
3.1.1 ExcursionXR ..... 33
3.1.2 MyVirtualTrip ..... 34

3.1.3 ICEVR: Immersive Climate Change Education Through Virtual Reality ..... 35
3.1.4 VERTEX ..... 36
3.1.5 HXT ..... 37
3.1.6 SURGIVIZ4D ..... 38
3.1.7 XPLORER ..... 38
3.1.8 SIGNCOMM ..... 39
3.1.9 LOOP LEARNING XR ..... 39
3.1.10 TUTORIST ..... 39
3.1.11 Conclusions ..... 40
3.2 Pilots ..... 41
3.2.1 XR2LEARN ..... 41
3.2.2 ECO-EXT: Ecopoly XR Edition ..... 41
3.2.3 XRAQUA ..... 42
3.2.4 SURE-VFT (Virtual Field Trips) ..... 43
3.2.5 GLEAM ..... 43
3.2.6 TEXIVOS ..... 44
3.2.7 AGRITWIX ..... 45
3.2.8 XR2CLASS ..... 45
3.2.9 VR4BEES ..... 46
3.2.10 HARMONY4ED ..... 47
3.2.11 Conclusions ..... 48
3.3 Final demos evaluation @ UnitedXR ..... 49
4 Troubleshooting ..... 55
4.1 Challenges ..... 57
4.2 Portal Hoper & World Builder documentation ..... 58
5 Conclusions ..... 62

# List of Figures

Figure 1. Enhancing Avatar reaching abilities through stretching-arm interaction metaphors ..... 13
Figure 2. Independent variables used to define the 5 interaction metaphors ..... 13
Figure 3. Smart watch interaction metaphor to switch between modes (on the left), and user experience preference questionnaire (on the right) ..... 15
Figure 4. One-way repeated measures ANOVA on ranks (Friedman test) of Embodiment and the corresponding post-hoc tests (Wilcoxon signed-rank test). adjusted p-value with Bonferroni correction. ..... 16
Figure 5. Perceived virtual hand location being accurate ..... 17
Figure 6. Two different modes of operation. (a) Teacher mode interpreting a PowerPoint presentation, (b) Two intelligent agents (students or teachers) exchanging ideas about the teaching material. ..... 19
Figure 7. Panoramic XR Intelligent Assistants which are located in a laboratory environment but shown as being in an outdoor environment. ..... 20
Figure 8. Immersive Tech Week 2023 Intelligent XR presentation. ..... 21
Figure 9. Bi-directional multimodal communication in VR, experiment setup ..... 25
Figure 10. Task 1: Object collection (left). Task 2: Placing objects (right) ..... 27
Figure 11. Participants interacting with our BIRDS system to evaluate user experience ..... 28
Figure 12. Quantitative results from Task 1 ..... 29
Figure 13. Quantitative results from Task 2 ..... 29
Figure 14. Information and consent form for the user evaluation (left) and demographics questions (right) ..... 51
Figure 15. Demo selection for the user evaluation ..... 52
Figure 16. List of questions for the user evaluation ..... 54
Figure 17. Forum dedicated to troubleshooting, development & updates, and Requests & feedback (capture taken on the 13th of December 2025) ..... 56
Figure 18. Detailed list of topics included in “Developments & Updates” ..... 56
Figure 19. Screen shot of the Troubleshooting forum ..... 57
Figure 20. XR4ED website to download the Portal Hopper and World Builder ..... 59
Figure 21. Documentation available for the World Builder ..... 60
Figure 22. Example of detailed documentation for getting started with the Portal Hopper ..... 61

# List of Tables

Table 1. Embodiment questionnaire used for the avatar interaction modes experiment ..... 15
Table 2. Summary of main finding on user experience ..... 23
Table 3. conditions for the BIRDS experiment ..... 26
Table 4. Demos agenda during UnitedXR ..... 50

# List of Acronyms

| Acronym | Description |
|---|---|
| API | Application Programming Interface |
| CA | Consortium Agreement |
| DoA | Description of Action |
| ECA | Embodied Conversational Agent |
| FS | Financial Statements |
| GA | Grant Agreement |
| HCI | Human Computer Interaction |
| HMD | Head Mounted Display |
| LLM | Large Language Model |
| PO | Project Coordinator |
| WP | Work Package |
| XR4ED | Extended Reality for Education |

# Executive Summary

The specific deliverable entitled “Evaluation and Troubleshooting” provides a comprehensive and structured overview of the evaluation activities carried out within the XR4ED project. In particular, it describes in detail the user studies conducted for several of the core modules developed throughout the project, as well as the evaluation processes implemented by the winning projects, addressing both developer-oriented and pilot-level perspectives. In addition, the deliverable thoroughly documents the troubleshooting methodology applied across the project lifecycle, which was designed to provide continuous technical and methodological support to the winning projects and to ensure the robustness and quality of the developed solutions.

Overall, the results of the final evaluation demonstrate a highly positive assessment of the project outcomes, particularly with regard to their potential impact on enhancing the learning experience for students through the use of extended reality technologies in education. The evaluations highlight the relevance, effectiveness, and innovation capacity of the developed solutions, confirming their alignment with the project’s objectives and with broader European priorities in digital education and immersive learning. The detailed outcomes of the evaluation of the winning projects presented during the final event are reported in Deliverable D5.3, which provides an in-depth analysis of the performance, usability, and perceived value of the winning projects.

Furthermore, the pilot experiments conducted throughout the entire duration of the project played a crucial role in generating valuable insights into the design and implementation of immersive educational experiences. These pilots enabled the consortium to identify best practices and key design principles for increasing learner engagement, strengthening the sense of presence, and improving overall user experience and learning performance. The knowledge gained through these iterative evaluation and experimentation processes represents an important contribution of the XR4ED project, supporting the development of more effective, engaging, and impactful XR-based educational applications in future initiatives.

# 1 Introduction

## 1.1 Overview

This deliverable corresponds to Task 5.4 which is about “User experience, Evaluation and Troubleshooting”. Since the goal of this project has been the development of modules that can facilitate the development of XR applications for education, it has been essential to evaluate the user experience for each module. This has been an iterative task which was carried out through the entire length of the project to evaluate each iteration of the modules with the purpose of detecting possible limitations at early stages of the modules and refine the design. The purpose of the present deliverable entitled

"Evaluation and troubleshooting", therefore is to describe in detail the findings done regarding user experience, the recommendation provided to the winning projects for the development of their projects, as well as helping them through their development with troubleshooting. Additionally, evaluations were performed for both the modules incorporated in the platform as well as for the final pilot projects.

From the beginning of the project, we ran smaller user studies with the purpose of detecting possible sources of problems or deciding between several possible designs of an interface or user interaction technique. Towards the end of this project, we performed more formal studies to evaluate for instance the impact of the tools in the learning process for the students. The evaluations performed during this project have also led to several novel research publications in the top conference in Human Computer Interaction, the ACM CHI conference on Human Factors in Computing Systems. With these contributions we not only strengthen the outcomes of this project, but help the international research community in XR and HCI with recommendations for the future development of applications with a special focus on education.

In this context, the present deliverable aims to fulfil the following main objectives:

- Understand the user experience of virtual tutors and interaction with self-avatars.
- Evaluations of the final projects delivered by the SMEs that were granted with cascade funding.
- Provide a comprehensive forum for troubleshooting as well as online biweekly meetings to provide support during the implementation of projects and integration of new modules in the platform.

## 1.2 Structure of the Document

The structure of this document is as follows:

- Section 2 provides several user studies that were conducted during the development of the avatar modules for the XR4ED platform.
- Section 3 presents the evaluations carried out for the 20 winning projects.
- Section 4 lists the troubleshooting activities that were carried out by the XR4ED consortium to support the SMEs developing the Pilots and Developers projects.
- Section 5 presents the conclusions for the evaluation and troubleshooting.

# 2 User Experience

User experience is essential to any development of XR applications, and particularly important for the case of education. User experience needs to provide a smooth flow to the events happening in XR, allowing the user to be engaged with the plot of the experience to enhance the benefits of learning in an immersive VR or AR application. From the beginning of the project, we put special emphasis on the evaluation of user experience through pilots developed by the technical partners. For example, in the case of the XR avatars module, we proposed novel techniques to improve the user experience when interacting with the virtual environment through an animated self-avatar. We then investigated Embodied Conversational Agents enhanced by LLMs (Large Language Models) to provide natural language conversations with the users of our platform to facilitate natural conversations in AR applications. Finally, we explore the combination of animations and LLM-based ECAs to provide natural bidirectional communication between the user and the ECA, so that besides natural language, we could also capture the user's gaze as an additional modality to improve and disambiguate the prompts send to the LLM, as well as having the ECA capable of using hand pointing to disambiguate instructions in a typical assembling scenario.

## 2.1 Avatar interaction metaphors

To bridge the gap in the existing differences between the user body and the avatar, we propose novel solutions to enlarge the arm of the avatar in order to facilitate the interaction with the virtual objects that were in hard-to-reach locations. This work involved the full development of avatars with enhanced IK, motion reconstruction from sparse data and extended features in the avatar configuration to extend the reach of the avatar.

Figure 1 shows a schematic representation of the five interaction metaphors that were used for the initial user studies, to stablish the most appropriate solution to include in the avatar module, with the independent variables being shown in Figure 2. From left to right: *FreeController* implies that the virtual controllers may be observed to move away from the avatar's hand due to the limitations in the avatar skeleton with respect to the human, *AttachedController* shows a common solution to the previous problem that consists of always rendering the virtual controllers on the avatar's hand, even if there is a mismatch in its position with respect to the real controller that the user is holding (which can lead to a visuo-proprioceptive incongruency), *Hand* is another common solution which avoids this visual mismatch by simply not rendering the controllers in VR (which leads to a visuo-haptic incongruency), *StretchController* renders the virtual controller always in the position of the real controller, and solves the mismatch by elongating the avatar's arm, and finally *StretchHand* follows the same idea of elongating the virtual hand to provide visuo-proprioceptive consistency, but visuo-haptic inconsistency. After proposing these five interaction metaphors we run a full user study to evaluate embodiment, user experience and task performance. Since our goal is to use the avatars in the context of a classroom, we specifically

planned tasks that could be needed in the classroom. The first one consisted on picking up boxes with letters and positioning them in specific locations ( we could use such an application for a spelling lecture for example), the second one was more dynamic, required to reach for balls that were being thrown towards the user ( similar to a gym exercises for handball for example), and the final one required painting over silhouette in a blackboard, as if we were practicing for handwriting exercises.

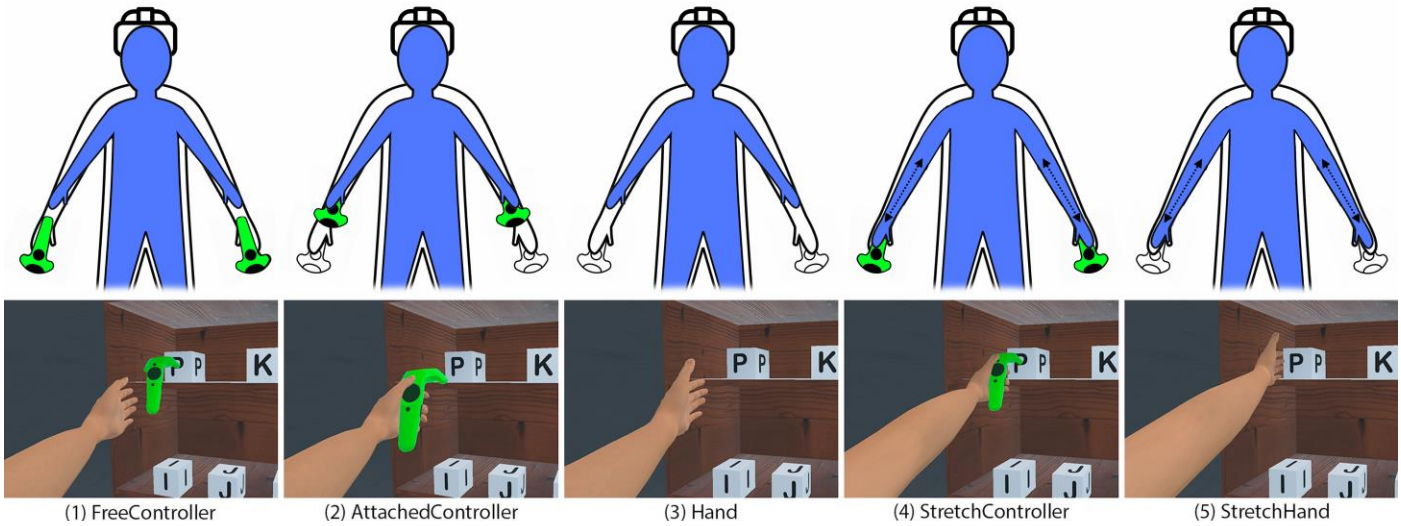



*Figure 1. Enhancing Avatar reaching abilities through stretching-arm interaction metaphors*

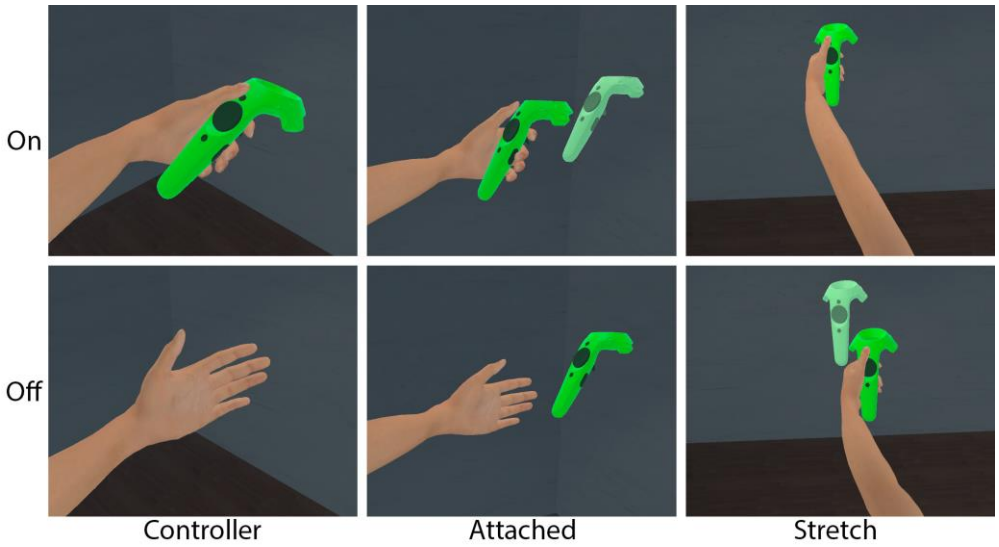



*Figure 2. Independent variables used to define the 5 interaction metaphors*

### 2.1.1 Interaction Metaphors User study

#### Participants

A total of 40 participants took part in the study: 34 were right-handed and 6 left-handed, with a gender distribution of 19 females and 21 males. Most participants were university students between the ages of 18 and 24. Participants were not compensated. Regarding gaming experience, 14 participants reported high, 13 medium, 10 low, and 4 no experience. Regarding VR experience, 6 reported high, 3 medium, 18 low, and 14 no experience. Nearly all participants were familiar with using computers, although only a few had prior exposure to VR technology. The Ethics Committee of the Universitat Politècnica de Catalunya issued a favourable opinion on the ethical aspects related to the research carried out in this project. The favourable approval of the application implies that the reviewed project complies with the criteria established by the institution's own expertise (which were included in deliverable D2.3).

#### Design

Participants were embodied in a virtual avatar animated using the AvatarGo library[1]. Participants could see and move through the virtual world as if they were controlling a virtual person with their own

movements, and focus on the interaction techniques rather than other embodiment issues. Furthermore, participants were given sufficient time to adjust and become accustomed to their new virtual avatars by following a series of standard embodiment instructions.

Throughout the first part of the experiment participants played in the virtual environment the three different tasks explained earlier in this document for each condition. The order of the conditions was established by a 10 x 5 balanced Latin square, and within each condition, the order in which the tasks were performed was randomized. After each condition, an embodiment questionnaire (see Table 1) was displayed inside the virtual environment. Questions were based on the work by Peck and Gonzalez-Franco [30] in which they propose four key embodiment components: appearance, response, ownership, and multi-sensory. We selected a subset of six questions to suit the specific needs of our experiment, with the aim of avoiding participant fatigue considering the duration of the experiment. While the chosen questions cover all the proposed four components of the SoE, they were particularly focused on those highly influencing ownership and multi-sensory aspects.

| Q# | question |
|---|---|
| **Q1** | I felt as if the movements of the virtual body were influencing my own movements. |
| **Q2** | I felt like I could control the virtual body as if it were my own body. |
| **Q3** | It seemed as if the touch I felt was caused by the controllers/block/brush touching the virtual body. |

[1] https://github.com/UPC-ViRVIG/AvatarGo

| | |
|---|---|
| **Q4** | It seemed as if my body was touching the controller/block/brush. |
| **Q5** | I felt as if my body was located where I saw the virtual body. |
| **Q6** | It seemed as if I felt the touch of the controller/block/brush in the location where I saw the virtual body touched. |

*Table 1. Embodiment questionnaire used for the avatar interaction modes experiment*

Following the three tasks across all conditions, we conducted an additional iteration of the conditions. During this phase, participants were queried about their perception of any differences between the positions of their real hands and those of the virtual hands or controllers.

To gather this information, a proprioceptive questionnaire was presented within the virtual environment, allowing participants to provide binary responses for each condition to the following question:

LQ1: Do you think your hand was in the position of your virtual hand?

At this stage, the experimenter explained the various interaction techniques to the participants. Following this, participants had the opportunity to freely explore and experience all the conditions while providing their feedback through a final preference questionnaire inside the VR (Figure 3 right). Notice that participants could change between different modes as many times as they wished using the VR controller's touchpad and a virtual watch (Figure 3 left), thus, they could easily compare between interaction modes and rate them based on preference. This questionnaire aimed to capture participants' personal preferences regarding the different interaction techniques used in the study.

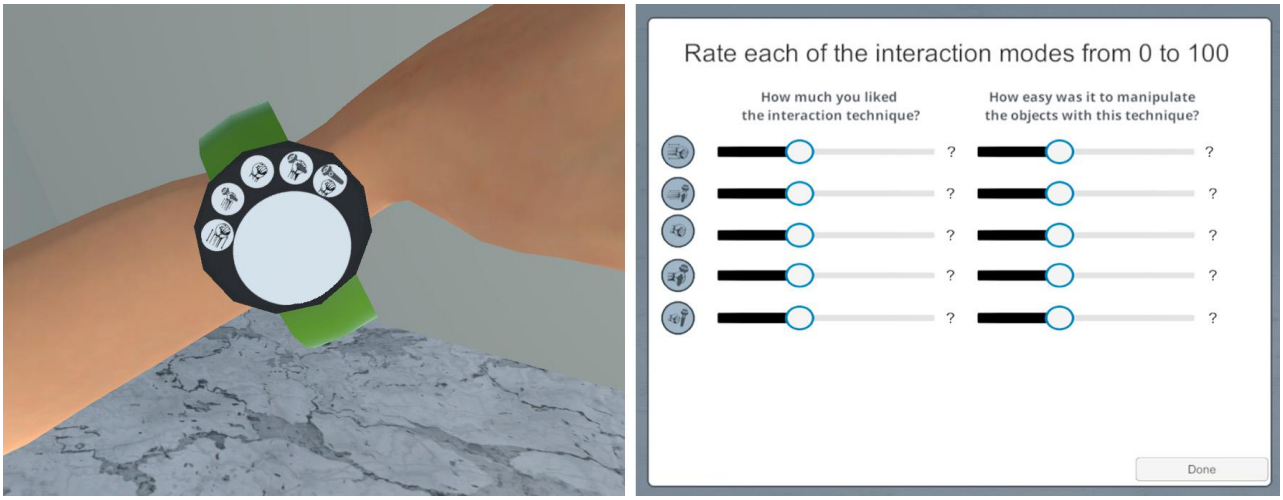


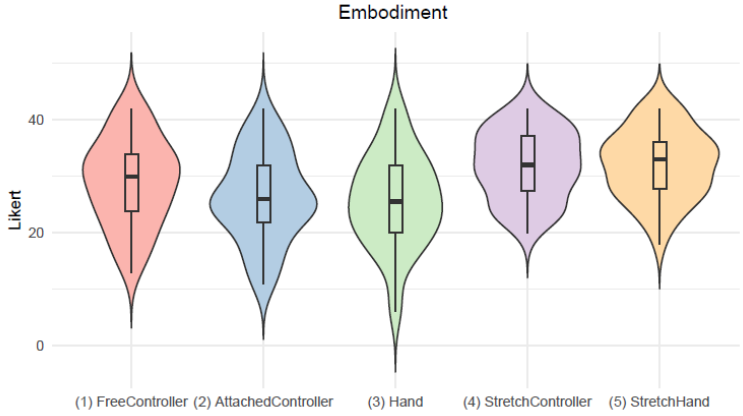

*Figure 3. Smart watch interaction metaphor to switch between modes (on the left), and user experience preference questionnaire (on the right)*

## Summary of results from the user study

In this report, we will provide an overview of the results obtained in this user study. Please refer to the full research publication of this work for the details of the statistical analysis:

Ponton, J.L., Keshavarz, R., Beacco, A. and **Pelechano, N.**, 2024, May. Stretch your reach: Studying Self-Avatar and Controller Misalignment in Virtual Reality Interaction. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (pp. 1-15).

Figure 16 displays the results of a Friedman test examining the impact of Mode on Embodiment. It reveals moderate differences among the groups, implying that certain interaction modes result in elevated Embodiment levels. In the post-hoc tests, we observe significant effects when contrasting methods that include Stretch with those that do not. However, no distinguishable difference is observed between the two methods that employ Stretch. Therefore, according to our results, both stretching arm techniques led to higher levels of embodiment.




| Effect | *p* | Effect Size | |
|---|---|---|---|
| *Mode* | **< .001** | 0.314 | moderate |
| 1 - 2 | **< .05** | 0.454 | moderate |
| 1 - 3 | 0.06 | 0.441 | moderate |
| 1 - 4 | **< .01** | 0.554 | large |
| 1 - 5 | **< .05** | 0.483 | moderate |
| 2 - 3 | 1.0 | 0.139 | small |
| 2 - 4 | **< .001** | 0.702 | large |
| 2 - 5 | **< .001** | 0.716 | large |
| 3 - 4 | **< .001** | 0.749 | large |
| 3 - 5 | **< .001** | 0.754 | large |
| 4 - 5 | 1.0 | 0.021 | small |

***Figure 4. One-way repeated measures ANOVA on ranks (Friedman test) of Embodiment and the corresponding post-hoc tests (Wilcoxon signed-rank test). adjusted p-value with Bonferroni correction.***

An interesting finding was related to the perceived hand location. As we can see in Figure 5, users had a higher feeling of the virtual hand being correctly located for the arm-stretching techniques, which was expected, since there was consistent visuo-proprioceptive congruency. However, participants rated above 50% for the case of the *attachedController*, which means that this mode could fool the participant to believe their hand was in a slightly different location. We believe that for virtual environments were direct manipulation is not needed, this perceptual position drift may be even stronger. Users also rated higher their preference for the interaction modes using the stretching-hand technique, therefore we strongly recommended this solution to enhance interaction in virtual reality when is mediated by having a self-avatar.

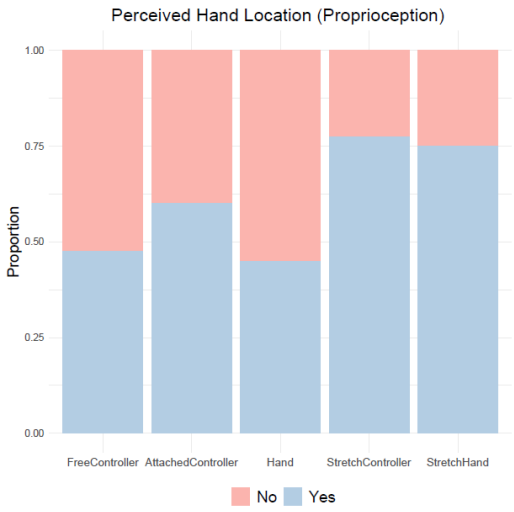



*Figure 5. Perceived virtual hand location being accurate*

Finally, regarding user performance during the three tasks, results suggest that for tasks with hard-to-reach objects, stretching arms provides the best performance without affecting embodiment. In our study, performance differences appeared mostly during the Cube manipulation task. In this task, users needed to carefully reach for objects located far away; thus, they could observe the differences between interaction metaphors.

Performance is higher when the end-effector (virtual hand or controller) is in the correct location, which occurs when we stretch the arm or use the flying controller. Similarly to Preference, modes *FreeController*, *StretchController*, and *StretchHand* would equally perform for pick-and-place tasks.

The Ball task required very rapid movements, in which the user was mainly focused on the next ball, that they could hardly observe differences, but even then, performance was worse for the *FreeController* mode.

For the Painting Task, the range of movements was mostly limited to a natural arm position, which did not often trigger the arm stretching or the flying controller. However, the few times these differences were observed led to users performing worse with the *FreeController* mode.

Therefore, for tasks that require accurate interaction with virtual objects, we recommend virtually stretching the arm to allow for correct end-effector positioning while respecting visual-tactile and proprioceptive coherence. For applications that require manipulation of objects within easy reach or that the interaction requires rapid movements without careful manipulation, the stretching technique would also be valid but not necessary. So, in these cases, having attached controllers would suffice.

## 2.2 LLM-powered Embodied Conversational Agents for AR

Advancements in technology that can reshape educational paradigms, with Extended Reality (XR) have a pivotal role. This work introduces an interactive XR intelligent assistant featuring a virtual teacher that interacts dynamically with PowerPoint presentations using OpenAI's ChatGPT API. The system incorporates Azure Cognitive Services for multilingual speech-to-text and text-to speech capabilities, custom lip-syncing solutions, eye gaze, head rotation, and gestures. Additionally, panoramic images can be used as a sky box, giving the illusion that the AI assistant is located at another location. The focus of the proposed AI assistant is not on replacing the teacher but on providing extra support with an intelligent interface that can be easily used by a non-technical educator. The content that is required is only the teaching material of the teacher and optional videos and panoramic images. Artificial Intelligence (AI) takes care of the interpretation of the teaching material whereas the presentation is conducted through XR. Teachers have the option to select between two generic avatars, one male and one female. Both include body animations, eye gaze, head motion, and lip-synchronization to increase realism. Findings from three pilots indicate that the proposed technology has a lot of potential to be used as an additional tool for enhancing the learning process. However, special care must be taken into privacy and ethical issues.

In this section, we will provide information about the AI avatars module for AR, and a summary of the findings on user experience obtained from the set of user studies performed by CYENS. The complete details of this module are included in the following research publication:

**Liarokapis, F.**, Milata, V. and Skola, F., 2024. Extended reality educational system with virtual teacher interaction for enhanced learning. *Multimodal Technologies and Interaction*, *8*(9), p.83.

### 2.2.1 Application Modes

The application supports two different modes of operation. In both cases, students can interact with the application and select the language of their preference. The first one refers to the virtual teacher delivering a presentation (Figure 6a). In this scenario, the learners can experience the virtual teacher interpreting the teaching material in real-time performance. In the second scenario (Figure 6b), the learner can participate in an intelligent conversation between two AI agents (which can be either colleague students or teachers) discussing the teaching material. In essence, ChatGPT is given a specific topic (the topic of the presentation), and the agents discuss a particular topic. The duration of the conversation is currently pre-specified to a few minutes, but this can be altered according to the needs of the learning scenarios.

An interesting feature of our application is that it is also possible to interrupt the AI conversation between the two agents and get engaged in the conversation itself. Based on this setting, a pilot study with six

participants explored a knowledge acquisition scenario within an AR environment featuring two ChatGPT-driven agents. The avatars, aware of users' gaze and mimicking realistic body language, gradually displayed social exclusion behaviors. While the results were not statistically significant, the data indicated an emotional shift post-interaction, with both positive and negative emotions decreasing, consistent with previous research on social exclusion.

*Figure 6. Two different modes of operation. (a) Teacher mode interpreting a PowerPoint presentation, (b) Two intelligent agents (students or teachers) exchanging ideas about the teaching material.*

Although the next wave of clever tools for perceiving the surroundings is panoramic imaging [45], it has not been used for creating educational XR environments. They provide a 360-degree view, which enhances the sense of immersion in virtual and augmented reality environments. Panoramic images are easy to capture by teachers and students and they can be used for creating immersive and realistic experiences. When incorporated into XR, they have the advantage of transporting the viewer to different places and times, such as historical sites, natural wonders, or fictional worlds. The XR system supports the activation of various panoramas, allowing users to immerse themselves fully in the educational content while obscuring physical surroundings. Figure 7 illustrates the use of panoramic imaging combined with XR by showcasing two XR intelligent assistants. This process is performed through the GUI of the XR application.

*Figure 7. Panoramic XR Intelligent Assistants which are located in a laboratory environment but shown as being in an outdoor environment.*

### 2.2.2 Qualitative Evaluation

This module has been qualitatively evaluated on three different occasions, with different audiences as described below. The sample of the participants covers an age range between 20 and 60 with a total of 65 participants (50 male and 15 female). On all occasions, the procedure followed was the same. All participants were prompted to ask questions and seek information about the presentations that were delivered to them. The title of the first presentation was “Introduction to Computer Graphics” and the second one was “Extended Reality”. The third presentation was about the XR4ED project entitled “An Overview of the XR4ED EU Project”. Qualitative data were then collected from participants’ responses. It is worth mentioning that this was just an informal evaluation with no data collection, anonymous feedback only, with no real risk. Verbal consent was obtained from all participants involved in the study.

**Teachers:** A group of 20 high-school teachers experienced the XR intelligent teacher in October 2023. Most of them did not have any previous experience with XR, which was expected. However, they also did not know what ChatGPT is about nor the latest developments on chatbots and large language models

(LLMs). On the positive side, they were all impressed with the fact that a potential student can ask interactive questions and get answers in real-time performance. Moreover, they appreciated the fact that multi-lingual support was provided, so the delivery of the material could be conducted in their language, even if the presentation was in the English language. They also liked the fact that a potential student could control the pace of the presentation. Finally, they really appreciated the fact that the application was operational and fully functional on mobile devices and tablets since the schools already have the necessary hardware. On the negative side, their main concern was the capabilities of the AI. Even if they found it interesting that ChatGPT could reply in real-time to several questions, they felt that their jobs could be in danger in the future. Several raised their concerns about ethical and privacy issues that come with AI as well as fake/wrong information. Some teachers also raised some concerns about the digital representation of the avatars, and they asked for better customization and more body language animations.

**Professionals:** During the Immersive Tech Week expo that took place in Rotterdam in November 2023, 35 professionals were exposed to the technology. Their background was very different from the teachers' group, and they mainly originated from industry (apart from three University Professors). They were all experts in XR and related domains (i.e., games) since the expo was focused on XR. A screen shot of the intelligent assistant that was demonstrated in the Immersive Tech Week expo is shown in Figure 8.

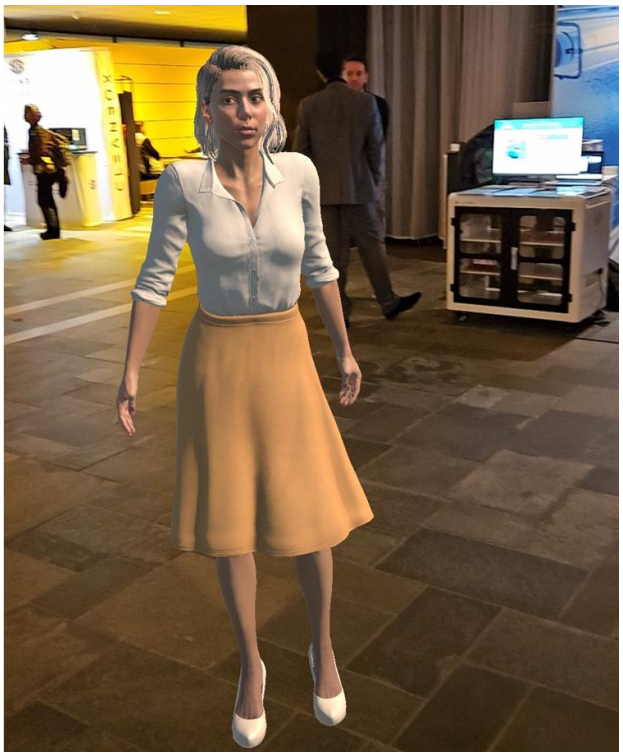


*Figure 8. Immersive Tech Week 2023 Intelligent XR presentation.*

The professionals indicated that the application has great potential at all levels of education. They really appreciated three aspects including (a) interactive responses, (b) multilingual support and (c) customizable in terms of content (i.e., can easily change panoramic images, PowerPoint presentations and avatars). They valued very highly the fact that the same application was ported on HMDs and mobile devices. Several of them asked if they could use the application in different application domains such as entertainment, business and marketing. The main concerns of the professionals were on the technical side of XR. Several asked for much more realistic virtual human representations. Unreal Engine supports the fast and high-fidelity creation of virtual humans, called MetaHuman Creator[2]. Several others wanted to have more variety in computer animations as well as computer graphics elements that can increase realism (i.e., real-time shadows). One person mentioned that it would be very interesting if there were collaborative aspects to the application so that multiple learners could access the same information at the same time. Finally, two participants were interested in having haptics so that they could physically interact with the avatars.

**Researchers:** During the Eurographics 2024 international conference that took place in April 2024, 10 researchers (ranging from post-docs to Professors) were exposed to the technology. All of them were experts in the domain of computer graphics including rendering, animation, computer vision and image processing. As expected, the main feedback that was received concerned the rendering quality of the intelligent assistant as well as the overall scene. Two participants complained that the body language animations are limited and they wanted more variety in the movements. One participant mentioned that there are no soft and hard shadows cast into the floor. While all participants liked the idea of using LLMs in conjunction with XR, they wanted to make use of AI to either generate animations or other relevant content (i.e., teaching material). Additionally, some mentioned customizing existing LLMs in order to be focused on education. On the positive side, the majority of researchers liked the features of the GUI which allows them to display the features of the application (see Section 5). Several participants rated positively the eye gaze, head movement, and lip synchronization. Additionally, they commented that realistic representations of avatars are not necessarily crucial in order to avoid the uncanny valley effect.

### 2.2.3 Limitation and Conclusions

The interactive XR intelligent assistant has the potential to transform education by providing immersive and interactive learning experiences. Results from the qualitative evaluation illustrate this. A summary of the results from 65 participants in three diverse fields is illustrated in Table 2. However, there are several limitations that can hinder the learning experience. One of the biggest barriers is the availability of high-quality software and hardware equipment. They are often expensive and depreciate within a period of three to five years, making it difficult for some educational institutions to afford them. Not all students have access to XR devices, which can create a digital divide. Additionally, some students may have physical or cognitive disabilities that make using XR challenging. Our solutions offer the choice between

[2] https:// www.unrealengine.com/en-US/metahuman

HMDs and smartphones which reduces the cost (since almost everyone owns a smartphone nowadays). Moreover, there is a limited amount of high-quality educational content available for XR educational experiences. Educators need proper training to effectively integrate XR into their teaching. Without adequate support, the technology may not be used to its full potential. Developing custom content can be time-consuming and costly. However, our solution is primarily based on content that can be easily generated by educators such as PowerPoint presentations and 360 panoramic images. In terms of avatars, the system has been connected to the XR4ED marketplace where teachers will be able to download and use related content. Finally, another potential limitation of the interactive XR intelligent assistant is highly immersive. This might lead to students becoming distracted from the learning objectives.

The feedback received regarding the avatars animations was overall very positive. However, avatars' gestures were so far limited to emphasizing speech without actually communicating content or helping in disambiguation of natural language, which led us to exploring such features for the avatar module as we describe in the next subsection.

| | Teachers (#20) | Professionals (#35) | Researchers (#10) |
|---|---|---|---|
| **Positive Feedback** | • Impressed by real-time interactive responses.<br>• Appreciated multi-lingual support<br>• Liked the ability to control the presentation pace<br>• Valued mobile and tablet compatibility | • Potential in all education levels.<br>• Valued interactive responses, multilingual support, and content customization<br>• Appreciated compatibility with HMDs and mobile devices | • Liked GUI features, eye gaze, head movement, and lip synchronization<br>• Noted that realistic avatars are not crucial to avoid the uncanny valley effect |
| **Negative Feedback** | • Worried about job security due to AI capabilities<br>• Ethical and privacy issues<br>• Concerns about fake information<br>• Desired better avatar customization and body language animations. | • Desired more realistic virtual human representations<br>• Wanted more variety in animations and graphics<br>• Suggested collaborative features and haptics for physical interaction | • Wanted more variety in body language animations<br>• Noted lack of soft and hard shadows<br>• Interested in using AI for generating animations and teaching material<br>• Suggested customizing LLMs for education |

*Table 2. Summary of main finding on user experience*

## 2.3 LLM-based Embodied Conversational Agents enhanced with bi-directional multi-modal communication for XR (BIRDS)

Virtual Reality (VR) enables users to experience fully immersive 3D environments in which they can explore and interact with digital objects in ways that closely resemble real-world interactions. This capability has expanded the range of applications across domains such as training, design, and education, where the combination of 3D visualization and interaction provides significant benefits. In face-to-face communication needed in education, natural gestures and non-verbal cues play a crucial role in clarifying verbal explanations and establishing spatial references. For example, in teaching 3D concepts (e.g., those found in anatomy or mathematics), students often rely on pointing to indicate objects when they lack the precise terminology. To be effective, VR interaction should strive to replicate the richness of human communication by supporting bi-directional exchanges augmented with non-verbal behaviors, including gaze and pointing.

Meanwhile the recent emergence of large language models (LLMs) has considerably encouraged the use of natural language in Human-Computer Interaction (HCI). However, current LLM-based interaction systems mostly rely on manually or verbally crafted prompts, which can be lengthy and effortful, or require users to specify large amounts of contextual information to narrow responses. To address these limitations, researchers have proposed leveraging multimodal signals (e.g., gaze) to enrich LLM prompts automatically with contextual cues. Similar approaches in augmented reality (AR) have demonstrated how gaze and other signals can disambiguate user intent and facilitate natural referencing. These examples illustrate the potential of multimodal input to reduce user effort in LLM-based interaction. However, in nearly all such systems, multimodal communication remains *unidirectional*, that is, the user provides enriched input, while the system responds only through text or audio.

Building on these developments, the integration of LLMs into virtual and mixed reality environments introduces new possibilities for embodied conversational agents (ECAs). Such agents enable richer and more human-like interactions by combining the open-ended language capabilities of LLMs with immersive, embodied experiences. With the introduction of LLM-powered ECAs, researchers have begun to deploy virtual humans capable of generating open-ended responses to user queries. These systems already support applications, for example, in teaching or psychological therapy. However, despite these advances, ECAs most often rely on voice alone by leaving out nonverbal output cues (e.g., gestures or pointing) that are central to human communication. Some systems do incorporate facial expressions and affective gestures to enhance empathy and social presence, yet these behaviors remain peripheral: they do not constitute communicative content and therefore fail to support functions such as disambiguating user intent or clarifying task-related concepts.

In this research study, we move beyond scripted or one-sided interaction by making exchanges with ECAs more closely to resemble human-to-human conversation. Nonverbal behaviors such as gaze, gesture, and pointing are essential to grounding and disambiguation in collaborative dialogue, providing context, emphasis, and nuance. While current ECAs can communicate verbally, achieving seamless *bidirectional*

*multimodal communication* (where both users and agents employ gaze, gestures, and speech) is critical for creating more effective, intuitive, and engaging interactions. To address this gap, we formulated two main Research Questions (RQs):

**RQ1**: How do gaze input and agent gestures, individually and in combination, affect task efficiency and accuracy in referential dialogue?

**RQ2:** How do these modalities influence user perceptions, including trust, user experience, and social presence?

We designed two tasks that required bi-directional communication and could benefit from non-verbal behavior, particularly in disambiguating spatial references (Figure 9 shows the experiment setup). In the first task, the user instructed an ECA to retrieve a specific object from a shelf. In the second, the ECA instructed the user on the spatial placement of objects within a shelf. Together, these tasks model interaction scenarios that resemble training or teaching contexts (e.g., assembling components, referring to objects whose characteristics, or relative positions are difficult to convey verbally). The purpose of this study is to evaluate user experience and student engagement when using our novel BIRDS framework (Please note that this research work is currently in the second review phase for the CHI 2026 conference, so it should be treated with confidentiality until we obtain the final acceptance notification, the work has been done as a collaboration between UPC and CYENS for the XR4ED project to investigate how to enhance user experience in education in XR):

Huang, K.; Ponton, J.L.; Constantinides, M.; Paschalidou, L.; Itatani, T.; Yin, T.**; Liarokapis, F. and Pelechano, N**., 2026. BIRDS: Bi-directional Interaction for Referential Dialogue in Spatial Contexts for Virtual Reality. Under review in *the 2026 CHI Conference on Human Factors in Computing Systems*.

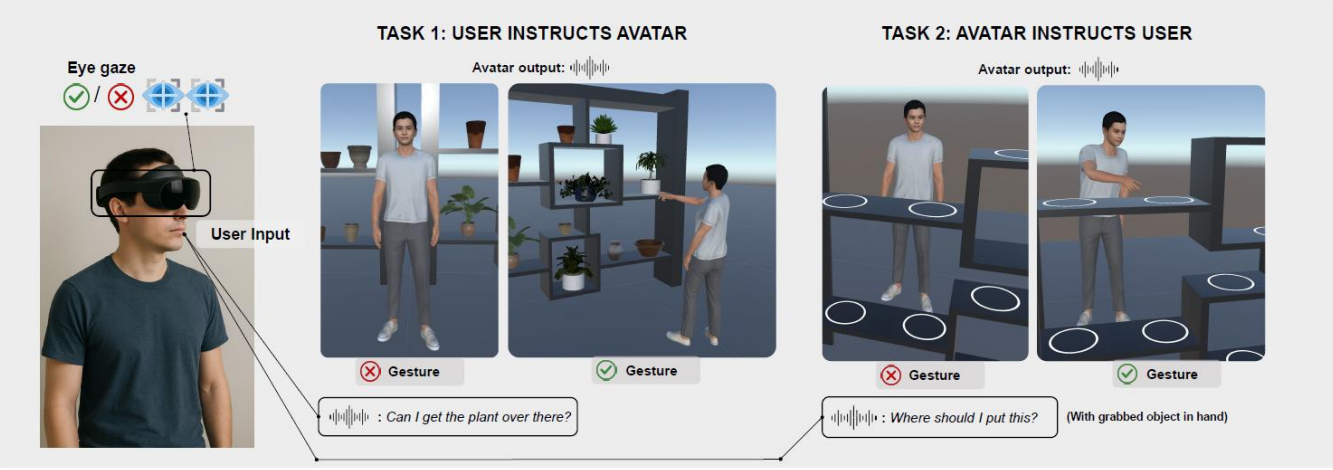



*Figure 9. Bi-directional multimodal communication in VR, experiment setup*

### 2.3.1 BIRDS experiment design

## Interaction Modes

To evaluate how gaze and deictic gesture contribute to referential dialogue with an LLM-based ECA, we employed a 2×2 within-subjects design crossing user-side gaze input (present/absent) with agent-side gesture output (present/absent) (Table 3). When gaze was absent (Baseline, Gesture Only), the ECA resolved referents using linguistic cues and dialogue history alone. When pointing was absent (Baseline, Gaze Only), referents were conveyed verbally only. Together, the four modes treat gaze (input) and gesture (output) as orthogonal factors that allowed us to evaluate their individual and combined effects in the two VR tasks.

| | Baseline | Gaze only | Gesture only | Gaze&Gesture |
|---|---|---|---|---|
| Gaze (user-side) | ✗ | ✓ | ✗ | ✓ |
| Gesture (agent-side) | ✗ | ✗ | ✓ | ✓ |

*Table 3. conditions for the BIRDS experiment*

The four conditions are thus:

**Baseline:** Interaction was purely verbal such that participants issued spoken queries or instructions, and the ECA responded with speech alone.

**Gaze Only:** Speech was augmented with real-time gaze captured by the headset. Information about objects detected by ray-casting was added to the LLM context. The ECA still responded with speech only.

**Gesture Only:** User input was speech only, but the ECA could accompany its replies with deictic pointing gestures toward target objects based on directives parsed from the LLM output, to make references explicit.

**Gaze and Gesture:** Bidirectional multimodality combined gaze-grounded input from the user with deictic pointing feedback from the ECA.

The tasks were inspired by educational settings where instructions may need to be provided to find virtual objects in a scenario or to correctly assemble objects in specific locations. More specifically we had the following two tasks:

**Task 1: Instructing the Avatar for Object Collection.** Participants verbally instruct an embodied conversational agent (ECA) to retrieve target items from two shelving units positioned behind the avatar (Figure 10 left). To induce referential ambiguity and place greater demands on spatial grounding, the shelves are stocked with visually similar pots and potted plants; near-duplicates (e.g., similar color or

size) are placed adjacently. This manipulation increases task difficulty by avoiding ceiling effects and improving sensitivity to differences across interaction conditions. A panel to the participant's left displays the current target item.

**Task 2: Placing Object under Instruction from Avatar.** The ECA instructs participants to place objects into designated locations (Figure 10 right). Participants stand between a table on one side—holding six graspable blocks, and a shelving unit on the other side that offers twelve candidate slots (twice the number of objects) to increase placement ambiguity and amplify demands on spatial grounding [28, 42]. The ECA is positioned behind the shelf and provides guidance.

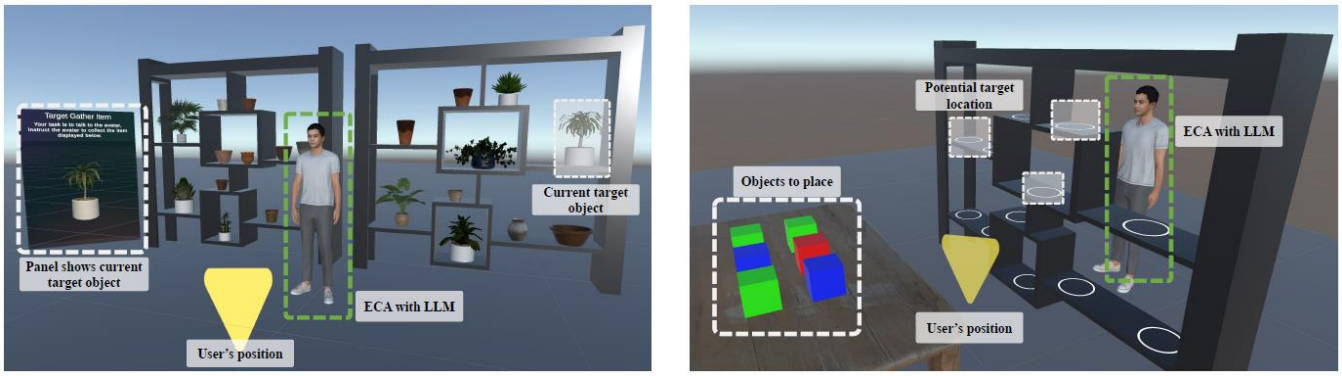



*Figure 10. Task 1: Object collection (left). Task 2: Placing objects (right)*

### 2.3.2 BIRDS user study

## Methodology

The evaluation study consisted of two phases: an in-VR interaction phase and a follow-up interview phase. These were designed to allow participants to engage with the *BIRDS* application. The VR sessions, including task completion and questionnaire responses, lasted between 60 and 90 minutes, depending on the participant. All sessions were conducted in a quiet and spacious room to ensure minimal distractions. In the second phase, each participant took part in a semi-structured interview lasting approximately 20 to 30 minutes. We employed a mixed-methods approach, combining quantitative data from questionnaires with qualitative insights from both interviews and log files recorded during the *BIRDS* VR interaction phase. The study was reviewed and approved by the ethics board of CYENS where the experiments took place.

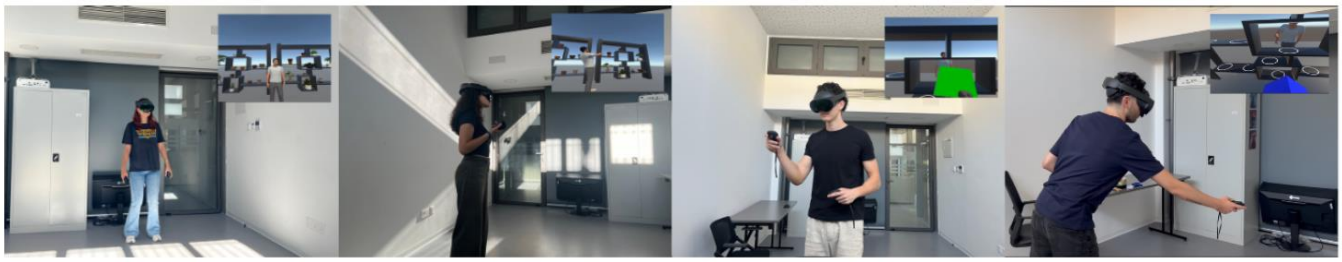


*Figure 11. Participants interacting with our BIRDS system to evaluate user experience*

## Procedure and Data Collection

Before the study, participants booked a two-hour time slot to accommodate both phases of the evaluation. Upon arrival, they completed a demographics survey. A researcher then introduced the VR headset and guided participants through how to use the system, creating a low-pressure environment for asking questions and getting familiar with the interface. Participants were given time to adapt and were instructed on using the VR controllers via an in-app tutorial integrated into the beginning of the BIRDS application.

Participants then proceeded with Task 1, which consisted of four conditions. Upon completing each condition, they removed the headset and completed a set of four questionnaires: (1) the Trust Questionnaire; (2) the User Experience Questionnaire (UEQ-S); (3) the Co-Presence scale; and (4) the Perceived Social Presence. The same structure was followed for Task 2. After completing both tasks, participants took part in a 20–30 minutes semi-structured interview. These interviews explored their overall impressions of the application, particularly the ECA, their thoughts on the use of such agents in broader contexts, and their personal reflections or revelations regarding the experience.

## Participants

We recruited 24 participants through personal networks and advertisement at CYENS. The sample included individuals from diverse backgrounds, such as bachelor's and master's students, accountants, CTOs, a stay-at-home mother, game engineers, programmers, marketing interns, unemployed individuals, and service workers, such as a waitress. All participants were healthy, had good knowledge of the English language, and reported no vision impairments that would interfere with gaze-tracking functionality. All participants resided in Europe. Ages ranged from 18 to 59 years (median = 25.33, std = 10.29); 12 participants identified as women and 12 as men. In terms of education, 2 held a master's degree, 9 a bachelor's degree, and 13 were undergraduate students. Participants reported varying levels of experience with VR: 2 participants used VR monthly, 10 had occasional experience, and 12 had never used VR before the study.

## Data Analysis

**Quantitative Evaluation:** We evaluated performance in each task and condition using six metrics: **Conversation Turns:** Total number of turns exchanged between the participant and the avatar. **Words Spoken:** Total word count from the participant's valid spoken responses. **Task Completion Time:** Time taken to collect the three objects (Task 1); time to correctly place the six objects (Task 2). **Task Accuracy:** Ratio of correct confirmations by the avatar over all selections (Task 1); ratio of correctly placed objects over all placements (Task 2). **Correlation Analysis:** Pairwise correlations between task accuracy and behavioral measures (turns, words, time), as well as among the behavioral measures themselves. **Subjective Measures** For each of the four post-experiment questionnaires, we compute the mean item score and use these per-questionnaire averages as subjective outcomes for both tasks.

For each metric, we first assessed normality using Shapiro–Wilk tests; if all condition-wise data satisfied normality, we used a repeated-measures ANOVA; otherwise, we applied a nonparametric Aligned Rank Transform (ART) ANOVA. Figure 12 and Figure 13 show the results of the quantitative metrics.

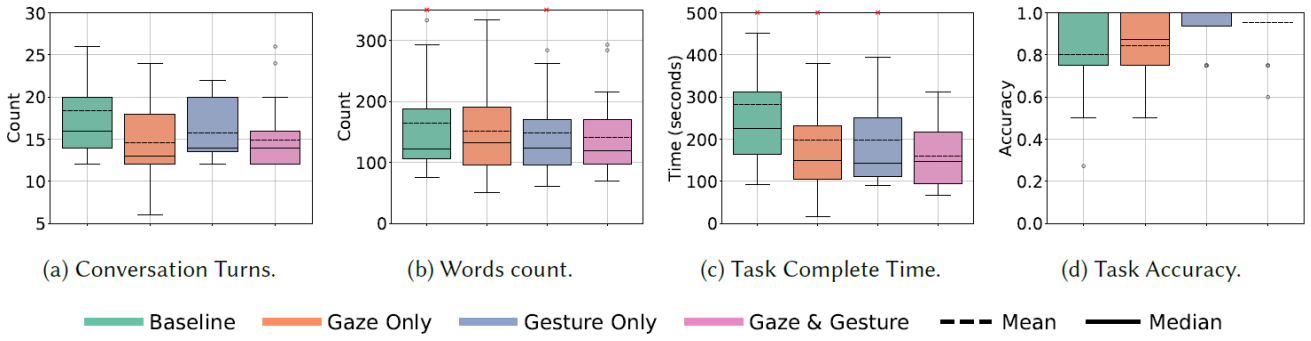



*Figure 12. Quantitative results from Task 1*

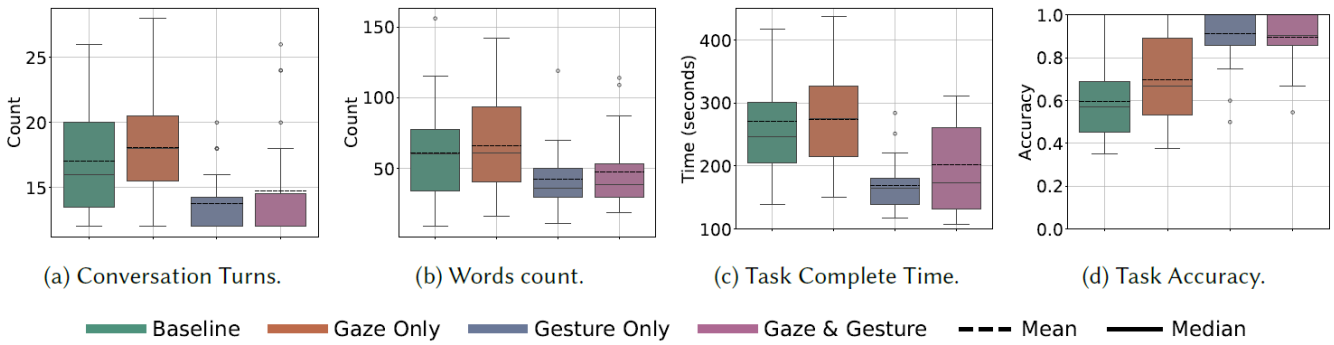



*Figure 13. Quantitative results from Task 2*

**Qualitative evaluation:** By analyzing the participants' interviews, we identified the most relevant qualitative insights from participant feedback, illustrating how different aspects of the agent's design and behaviour shaped user perceptions, interaction quality, and envisioned applications (we use P# to indicate the participant number that provided each comment):

1) *Comfort, Usability, Trust.* Most participants found interacting with the agent comfortable and easy, with 22 out of 24 describing the experience as relaxed, intuitive, and non-threatening after an initial familiarization period. For many, comfort was not just a default, but directly shaped by the interaction modality. As P4 remarked, *"I was a lot more comfortable with the agent when he could also point, and I also had my gaze. It was generally easier for me to complete the tasks, but yeah, as I said before, when there's only verbal communication, it was slightly more difficult"*. Both P19 and P21 similarly linked their comfort to the agent's capacity for gestures and gaze. P19 stated, *"it was awkward when I was moving around and he couldn't move his head to the direction I was going ... that felt kind of uncomfortable"*, while P21 noted, *"it felt comforting being in another environment that was almost real. It does feel safer and more natural in a way, although it is not the right word to use, but it is definitely closer to reality than voice assistance"*. Trust in the agent was also closely tied to these enriched modalities. Participants consistently reported feeling more confident in the agent when it provided gestures or attended to gaze, rather than relying on voice alone. As P21 explained, *"I trusted it more when it followed my gaze, like it was really paying attention to what I was doing"*. P13 described a clear difference: *"at the second task, it was a little bit difficult to understand some of the positions he was describing me, but when he was pointing at the position, it was clearer"*. P23 offered a similar sentiment, saying, *"in task one, I question it because when the agent had only voice, it was a bit confusing in order to understand what he was mentioning to me"*. Altogether, 13 participants explicitly discussed how richer conditions (gestures, gaze, or both) made them trust the agent more, particularly compared to voice-only interaction. Usability received high marks across the sample: 20 out of 24 participants described the system as quick to learn and intuitive to use, with the main challenges stemming from adapting to the VR hardware rather than the agent itself. In summary, participants overwhelmingly found the experience comfortable, trustworthy, and usable—especially when the agent combined gestures and gaze with its verbal instructions. The inclusion of these enhanced modalities not only fostered a stronger sense of ease and confidence but also established a foundation for reliable and satisfying collaboration.

*2) Tasks in Real Life and Future Use.* Participants described a wide range of potential applications for agents like the one in this study, with 19 participants envisioning concrete real-world scenarios where such systems could add value. Education emerged as the most frequently mentioned context, particularly for enhancing engagement and interactivity. P8 imagined immersive learning experiences, explaining that "*maybe in education it can be a fun way to get children to engage more with things that it's not very easy for schools to have access to...in the virtual world you can have basically everything*". Similarly, P24 highlighted the potential to support learning at university level, noting "*it would be more interactive...like a teacher talking to me and not me trying to understand alone and reading books*". Other participants described how the agent could serve as a guide in physical spaces. P9 proposed "*maybe in healthcare. . . or in schools,*

*maybe guide some students or people inside buildings they haven't been before*", while P1 suggested "*in a library...you can give a description of the book and it can guide you to the shelf where the book is*". Beyond education and navigation, participants also imagined roles in customer service, museums, tourism, assistive technologies for people with visual impairments, and professional environments such as warehouses or medical training.

**Main findings:** We will include in this report, the summary of the main findings from this experiment. Please note that the full statistical analysis can be found in the research paper referenced in this section.

Our main findings reveal three key insights: 1) agent gestures act as the primary driver of communicative efficiency by grounding dialogue; 2) user gaze plays a crucial, complementary role in refining understanding and achieving precision; and 3) these performance gains are directly linked to a richer subjective experience, fostering greater trust and perceived social presence.

1) Gestures employed by the agent emerged as the most effective mechanism for establishing conversational grounding and reducing cognitive load. In both tasks, the ability for the agent to point significantly improved performance. This effect was especially pronounced in Task 2, where the agent instructed the user. Here, a simple deictic gesture resolved spatial ambiguity far more effectively than complex verbal descriptions (e.g., "the second slot from the right on the middle shelf"), leading to dramatic reductions in time, words, and errors. When gestures were available, task accuracy became less dependent on the length and turn count of the conversation, indicating a fundamental shift to a more efficient communication strategy.

2) In contrast, user gaze functioned as a complementary, refining modality. While gaze input alone provided modest efficiency gains in Task 1, its most critical contribution was revealed in the significant interaction effect of Gaze and Gesture on accuracy in Task 2. This shows that the most effective communication was not just multimodal, but bidirectional. Agent gestures provided a clear directive, and user gaze offered a low-effort channel for the system to confirm the user's focus, enabling the highest level of precision.

3) Finally, these objective performance improvements translated directly into a richer subjective experience. Participants rated the agent as more trustworthy and found the experience more positive in the conditions where gestures and gaze were available. Our qualitative data confirms this link; participants described the agent as more natural when it used non-verbal cues. This increased the agent's perceived social presence. However, participants were also sensitive to technical limitations like animation quality and response latency. This underscores that while non-verbal modalities are powerful, their effectiveness is ultimately mediated by the quality of their implementation; for an ECA to be trusted, it must not only be capable but also perceived as responsive and coherent.

## 2.4 Conclusions regarding user experience

In this section we have described three user studies that were carried out as research goals to study how to improve user experience in educational applications, by having either a self-avatar to facilitate interaction with objects or as embodied conversational agents that can act as teachers. Our findings are key to advance the field of education in VR, since currently one of the biggest challenges in VR is having believable virtual humans to interact with. We have incorporated the self-avatar in the marketplace, and the ECAs for XR applications. The last development including non-verbal communication is still pending integration. The main reason being that only a few headsets allow to capture eye tracking, therefore we may want to explore the possibility of using controller or hand pointing instead, even if it has lower spatial accuracy.

# 3 Evaluations

Evaluations were included in all the winning projects, both for developers and pilots. Each project had its own KPIs to meet by testing their application through user studies. The detailed results are listed in the final reports that were submitted by the SMEs, so in this deliverable we include a summary of the main evaluation activities performed and the results obtained. Then we also include the final evaluation that was performed during the final event that took place during the UnitedXR[3] exhibition in December (8th-10th) in Brussels.

## 3.1 Developers

### 3.1.1 ExcursionXR

Testing and validation were carried out by bringing participants in contact with the platform and their extensions, observing their interaction, and translating findings into improvements while validating our core assumptions. They separated evaluation into two complementary groups:

- Developers (creation workflow): Can participants use the World Builder template and the Auto-Adaptive component model to produce a simple, spatially adaptive scene? Does the extension of the XR4ED platform afford the intended usability (component configuration, iteration speed, and export)?

- End users (runtime experience): Do participants perceive the promised technological benefits (free movement, interaction aligned with physical affordances, reduced motion sickness)? Where are process frictions (onboarding, scanning, re-scans) and how should they be improved?

Across 28 semi-guided tests — split between 9 developers and 19 end-users — the evaluation confirmed the core promise of AAMR while highlighting improvements that were subsequently implemented. All testers were of legal age, had varying degrees of previous experience with VR, lived in Germany, and spoke German or English.

Participants reported a clear “wow” effect at the moment of transformation (mentioned by 15 end-user testers), followed by a sense of natural, safe locomotion in their real room. Being able to touch familiar furniture (passive haptics) increased confidence and immersion (mentioned by 10 end-user testers). Participants described the experience as safer and more believable because familiar objects remained visible/usable as anchors in space. Several testers contrasted the experience favorably with conventional VR or 360° content, noting reduced discomfort or motion sickness (mentioned by 5 end-user testers). Measured scan durations in our 19 end-user sessions ranged from 45 seconds to 11 minutes for a first scan (median ≈2 minutes; mean ≈3.4 minutes), with longer times linked to large rooms or missing familiarity with the controller buttons / first-time VR use. Repeat scans, when performed, typically

[3] https://unitedxr.eu/

completed in around 1-2 minutes. When rooms changed or mistakes occurred, a quick re-scan restored alignment without friction. The anticipated advantages of (i) free movement, (ii) passive haptics, and (iii) reduced motion sickness were thus highlighted as perceived benefits observed in the tests. Several educators linked these qualities to emerging use cases like micro-stations in history or geography, or low-setup "lab-like" explorations, stressing that the experience felt believable rather than purely audiovisual.

Developers found the template-first approach effective for a fast start. The AutoAdaptiveObject concept and the placement modes (Wall/Table/OnTable/Other) were repeatedly ranked as intuitive (mentioned by 8 developer testers), with parameter controls (offsets/spacing/tiling) enabling precise refinement. Measured times across nine developer sessions show where effort concentrated: "World Builder usage" had a median of around 95 minutes (range 20-120 min), "Project creation with the template" around 10 minutes (5-20 min), "Using the AAMR plugin" around 45 minutes (30-70 min), and "Iterative testing with the Portal Hopper" around 25 minutes (5–40 min). Overall, total time from first launch to a running AAMR application spanned roughly 65 to 230 minutes depending on prior familiarity and World Builder friction. One developer explicitly reported building a first app end-to-end in 2 hours. Common friction points, however, centered on World Builder usability rather than the AAMR functionalities itself (as described below). Some testers requested visual affordances during authoring (e.g., clearer wall-orientation indicators), inline tooltips/help, more building rules for advanced control (which we added later in the project) and a more visible hint for the sample-room start ("Press P"). In response, we refined property naming and expanded the documentation to shorten time-to-first-success.

In sum, the end-user tests validate the project's core proposition: rapid room capture, deterministic adaptation to the learner's space, and a resulting experience that supports confident free-roam movement with meaningful physical interaction. The iterative improvements to scan onboarding, clearer controller instructions, and restartability addressed first-use hurdles without architectural changes, while residual limitations relate primarily to platform constraints rather than to the AAMR implementation itself. On the developer side, the template and component system were confirmed as the right abstraction for broad adoption, with incremental requests for visual cues and advanced rules addressed through documentation, tutorials, parameter refinements, and in-editor hints.

### 3.1.2 MyVirtualTrip

Pilot Testing and Validation was performed through controlled experiments with 50 participants who validated personalization across learning styles (visual, auditory, kinesthetic) and cognitive load levels (low, medium, high). Regarding personalization, tests confirmed that tailoring content to user profiles significantly increased (more than 20%) engagement and efficiency.

All KPIs related to evaluations were successfully met:

1. Personalization Accuracy – Target: 90% | Achieved: 100% accuracy across 50 participants

2. User Engagement Increase – Target: 20% | Achieved: 22.7% increase, validated by paired t-test across 50 participants

3. Educator Satisfaction – Target: 80% | Achieved: 100% satisfaction among pilot educators.

Their results showed that learners benefited from shorter learning curves, higher retention, and greater motivation. Educators gained a **reliable authoring pipeline** with personalization built in, while the XR4ED ecosystem was extended with reusable personalization logic, paving the way for scalable adoption in education, vocational training, and beyond.

User engagement was measured as a composite metric combining (1) gaze-based attention, (2) task performance, and (3) participation level. Gaze data are captured for all learning-relevant objects (visual, auditory, kinesthetic), and time-on-relevant-objects (TRO) is computed as the proportion of session time spent fixating on these pedagogical AOIs. TRO has been shown to correlate strongly with learning gains and attention in immersive learning studies (Lai et al., 2013; Holmqvist et al., 2011; Makransky et al., 2019). Task performance is quantified through normalized task duration (actual/optimal), correctness of answers, and number of attempts, forming a performance score sensitive to both speed and accuracy. Participation is captured via tasks completed (%) and total user inputs. These components are combined into a weighted engagement index, allowing per-participant comparisons between a non-personalised baseline session and a personalised session. A ≥20% average increase in this engagement index across ≥50 participants (validated by paired t-test, $p < 0.05$) will verify the KPI.

### 3.1.3 ICEVR: Immersive Climate Change Education Through Virtual Reality

A total of five upper-secondary schools in Sweden have participated in testing, and early takeaways indicate clear objectives, strong curriculum fit (geography/biology), and interest in embedded knowledge checks.

The testing was structured as a synchronous digital session, followed by an asynchronous feedback form. The key evaluation objectives where: (1) to validate ICEVR's alignment with the Swedish curriculum; (2) to assess teachers' perceived value of immersive VR for climate education; and (3) to gather specific requests on lesson plan integration, assessment methods, and classroom use. The average answers received in the questionnaire regarding engagement and curriculum criteria were the following (scale from 0 to 10):

- ICEVR is an interesting way to work with climate-related topics: 9.6

- I see potential in using this type of VR material in my teaching: 10.0

- ICEVR could contribute to increased student engagement: 9.8

- ICEVR aligns well with core parts of the Swedish curriculum: 10.0

- I would like to use similar digital experiences in other subjects: 10.0

Overall, teachers saw strong potential for multidisciplinary use, especially in STEM and Civics.

### 3.1.4 VERTEX

Multiple evaluation activities were conducted to assess usability and learning potential:

- A pre-evaluation survey was designed and distributed to students.

- A structured user feedback study captured insights on app usability, functionality, and satisfaction.

- A survey assessed narrative, personalization, immersion, atmosphere, and overall impressions of three XR scenarios.

Results indicated high engagement, appreciation for personalized narratives, and strong user satisfaction.

They evaluated the educational efficacy of their modules and received students' user experience. More specifically, the evaluation of the VERTEX platform was conducted through a pilot phase with the Department of Informatics at the Athens University of Economics & Business (AUEB). This pilot involved 30 students from the MSc program in Digital Methods for the Humanities, who are ideal candidates for testing VERTEX's capabilities in creating immersive educational content. They chose two existing courses as tasks.

The evaluation was designed as a multi-phase process, covering the full cycle from requirements analysis to user testing and reflective feedback. They did pre-evaluation, in-task and post-task assessments, followed by a survey.

The main findings of their user study include:

- Device used: Identifying whether the app was accessed via smartphone, tablet, or other devices. *Results:* Smartphone (97%), Tablet (3%), Other (0%).

- Ease of Use: Rating how simple it was to start the reconstruction process, record videos, and view results. *Results:* Easy (60%), Fairly Easy (30%), Neutral (10%), Fairly Difficult (0%), Difficult (0%).

- Feature Performance: Evaluating the accuracy of the object identification tool, the video recording process and the visual exploration experience. Results: Object Identification: Yes, it worked perfectly (50%), Mostly, but I had to adjust it (50%), No, it didn't detect the correct object (0%), I'm not sure (0%); *Video Recording:* Easy (20%), Fairly Easy (40%), Neutral (20%), Fairly Difficult (20%), Difficult (0%); *3D Visual Exploration:* Easy (50%), Fairly Easy (50%), Neutral (0%), Fairly Difficult (0%), Difficult (0%).

- Reconstruction efficiency: Most students (60%) succeeded in 2-3 attempts, with half of the students needing 5-10 minutes to complete the reconstruction. Most students (90%) rated the quality as Good or Excellent.

- Educational Value: 70% answer yes, and 30% maybe when asked if VERTEX could be applied effectively in other academic contexts.

- Impact on Engagement: How did the integration of a participant-created 3D object affect your engagement with the story? Results were: Greatly enhanced engagement (40%), Enhanced engagement (43%), Neutral (17%), Slightly enhanced engagement (0%), No change (0%).

- Immersiveness of the Virtual Setting: How immersive was the virtual setting as an environment for the 3D experience? Very immersive (43%), Immersive (40%), Neutral (17%), Slightly immersive (0%), Not immersive (0%).

Therefore, the results of their user study provided high ratings regarding engagement, immersiveness, and education value. Regarding the improvement in trainee skills their responses show a clear perceived improvement of at least 20% in accuracy and environment understanding when training with Magos compared to controllers.

### 3.1.5 HXT

They performed user studies with 30 participants. To quantify efficiency gains, participants completed the same pre-take-off procedure twice—once using standard VR controllers and once using Magos gloves—under identical briefing, environment, and task conditions. Start and finish timestamps were captured automatically by the evaluation component for every run. The order of the two conditions was counterbalanced across participants to minimise learning-effect bias. Analysis of the timestamp logs shows that average completion time with Magos was approximately 29% lower than with controllers, exceeding the 25% reduction target and indicating a meaningful improvement in procedural fluency.

Then they assessed perceived skills/understanding via a structured post-session questionnaire administered immediately after both conditions to evaluate the improvement in Trainee skills. The instrument captured demographics and prior VR exposure, and used comparative items (e.g., immersion, perceived accuracy of actions, understanding of cockpit layout and flows), plus open-ended comments. Responses show a clear perceived improvement of at least 20% in accuracy and environment understanding when training with Magos compared to controllers. Qualitative feedback consistently cited more natural hand use, better spatial awareness of controls, and reduced cognitive load from not having to translate controller inputs into finger actions.

### 3.1.6 SURGIVIZ4D

More than 360 participants, including surgeons, medical students, and VR researchers and developers, tested the system during live demonstrations. Data collected was anonymous, limited to performance indicators, and independent from personal or demographic information. Each scenario generated a different set of measurable metrics, aligned with its pedagogical goals.

Overview of Participation

- Scenario 1 – Tumor Contouring: 83 users

- Scenario 2 – Tumor Localization with Ultrasound: 77 users

- Scenario 3 – Tool handling and targeted anatomy access: 200 users

Across all three scenarios, most users completed the task in under one minute. This reflects the rapid "demo format" required in public events, where individual users are typically allocated less than 2–3 minutes per station.

The most relevant results in terms of user experience were:

- 128/200 users successfully reached the tumor.

- 88/200 users touched a bone, an error representing a clinically relevant mistake.

- Repeated tries of both the access and navigating inside the organ were around 2-3 tries each.

Despite the higher procedural complexity, 64% success was achieved, even with many first-time VR users. Bone contact was frequent, showing that anatomical depth perception and needle guidance cues should be refined. This scenario provided the richest data for improving surgical realism and ergonomic feedback.

### 3.1.7 XPLORER

They performed testing and validation with 30 students at Handong Global University (Korea), 60 secondary-school students in France, and 6 teachers (3 in Korea, 3 in France) participated in testing and co-design feedback.

To evaluate user satisfaction, they run a pilot and collected feedback from a survey, which was completed by students from age 12 to 15. Results showed a 95% satisfaction rate, with all students rating the experience either 4 or 5 stars out of 5. They described Xplorer as fun, educational, easy to use, and appreciated exploring in 3D, interacting with AI, and taking quizzes. Suggestions included adding more levels and realistic transitions.

### 3.1.8 SIGNCOMM

User satisfaction has not been evaluated at the time of the final report. They simply reported that preliminary feedback from advisors indicates positive results in terms of usability and accuracy.

### 3.1.9 LOOP LEARNING XR

Testing and Validation has been conducted with a comprehensive testing campaign with 10 therapists and 20 students from partner educational institutions. The sessions evaluated usability, engagement, and learning outcomes. Feedback confirmed high levels of satisfaction and intuitive interaction, with users demonstrating measurable improvements in performing the targeted life skills.

For the therapists, they report an overall app satisfaction of 100%, meaning that all therapists would recommend it. They also obtain 100% yes to the indicator of “willing to use it in classes”. Ease of use got 4.62/5.0 on average with therapist considering the app “intuitive and clear, with only minor lags”. Interface clarity also received 100% of positive (Yes) answers, and usefulness for high-support students received an average of 4.38/5.

Regarding the student survey, they obtained the following results:

- Liked the app 27 Yes / 3 No (90 %) - “Peaceful and fun to work in VR.”

- Would use again in class 28 Yes / 2 No (93 %) - High enthusiasm after initial hesitation.

- Found VR learning useful 28 Yes / 2 No - Enjoyed learning by doing.

- Ease of use (1–5) 4.13 avg 21 - Needed teacher support in early sessions.

- Interface intuitive 18 Yes / 12 No - Minor difficulties with controls for some students.

Their main takeaways from the participants answers were that the application had a high engagement and acceptance by both students and educators. The Measured improvement in independent task completion (≈ +30 pp). The found Strong evidence of transfer from VR to real life skills. And finally, the average usability rating was >4 / 5 across groups.

### 3.1.10 TUTORIST

To measure user experience, they did Pre-pilot questionnaires (Likert scales for satisfaction/engagement) and focus groups captured user experiences, including avatar realism and multilingual Q&A usability. Analytics tracked metrics like session duration and completion rates via www.tutorist.tech dashboards. The pilots they ran validated the multilingual capabilities, with positive feedback on avatar realism, response accuracy (90%+ satisfaction), and ease of use. Quantitative metrics underscored the platform's effectiveness in boosting engagement. They also report that stakeholder satisfaction was 65%, completion rates increased by 20%, student overall rating on learning quality increased by 25%, student engagement increased by 50%, and student completion rate increased by 30%.

### 3.1.11 Conclusions

The integration of the developed modules into the XR4ED platform was successfully achieved by all developers. An iterative evaluation approach was consistently applied, allowing prototypes to be tested, refined, and progressively aligned with the platform's requirements. This process contributed to improving both functionality and usability prior to full integration, while enabling developers to identify and address friction points early in the workflow. Component-based templates, personalization features, and immersive content creation tools proved effective in accelerating development cycles, supporting scalable deployment, and enabling reproducible XR experiences.

Most pilots conducted user studies within real educational settings, involving both students and educators, which provided valuable feedback on usability, engagement, and learning outcomes. End-users consistently reported high levels of satisfaction, immersion, and motivation. Projects such as ExcursionXR and XPLORER highlighted intuitive interaction, engaging 3D experiences, and strong "wow" effects. MyVirtualTrip demonstrated that content personalization significantly enhances engagement and learning efficiency, while ICEVR and VERTEX confirmed alignment with curricular objectives and improvements in comprehension and skill acquisition. HXT and SURGIVIZ4D illustrated measurable gains in procedural efficiency and task performance, and specialized applications such as LOOP LEARNING XR, TUTORIST, and HARMONY4ED showcased the platform's capacity to support life skills, multilingual learning, and therapeutic applications. Preliminary feedback from SIGNCOMM also suggested positive outcomes regarding usability and accuracy.

Overall, these findings demonstrate the effectiveness of the implemented solutions and confirm the potential of the XR4ED platform to support engaging, immersive, and teacher-friendly XR-based educational experiences. By combining developer productivity with learner-centered design, XR4ED has proven capable of enhancing motivation, accelerating skill development, and supporting inclusive education across a wide range of subjects, learners, and educational contexts.

## 3.2 Pilots

### 3.2.1 XR2LEARN

Task based learning activities were developed to evaluate the performance of the tool. Out of 127 reached participants, 107 participants have successfully completed the training, corresponding to 84%. Overall satisfaction with the training 87%, according to the satisfaction questionnaires. As regarding to the comparing toward traditional methods, in the Pilot 2: Agriculture – 93% responded positively, while for the Piloting 1: Onboarding – 92%, prefer to use XR experience toward traditional methods. Knowledge retention was effectively demonstrated through the XR-based learning experiences. Participants consistently reported high levels of understanding, autonomy, and engagement across the training phases, confirming strong mnemonic and transversal learning outcomes. We estimated an average evaluation of the knowledge retention of the 85%.

Through the question in the pre and post questionnaires, "Would you recommend this training to other?" The initial answer within the Pre-evaluation questionnaire was around the 73%, but measurements though the post – assessment show up the increase up to 92%.

Comparison between the pre- and post-assessment questionnaires clearly demonstrates a marked improvement in learners' confidence, clarity, and autonomy. Before the XR onboarding, participants expressed curiosity but moderate self-assurance in applying new knowledge. After completing the immersive training, all respondents reported feeling "very much" more confident and rated the experience as "extremely engaging" and "clear and intuitive." These results indicate that participants transitioned from initial understanding to practical competence more rapidly, confirming that XR-based experiential learning significantly shortened the path from knowledge to skill. All the students started with a high level of knowledge about the subject. However, they still learned many new concepts about waste separation, company values and learned how to navigate the virtual office. This shows an extremely fast time to proficiency (expected value > 30% reached).

Comparison between pre- and post-assessment data, students demonstrated high engagement (average 4.2/5) and clearer understanding of practical applications compared to traditional lectures. Students also reported average 80% of the clearness.

Finally, engagement for each pilot was evaluated with a 4.5 out of 5 for Pilot 1 Onboarding and 4.2 out of 5 for Pilot 2 Agriculture. Thus, showing that engagement was rated as 87% positive on average.

### 3.2.2 ECO-EXT: Ecopoly XR Edition

An evaluation protocol was defined for running the pilots. With this protocol, the team conducted a total of 11 pilots with school classes, student and volunteer groups. These ranged from primary and secondary

schools, vocational education, university students from various backgrounds including technology, design, economics, psychology, environmental sciences, education and business, to volunteers from an environmental NGO. They reached a total of 90 players, in age range 11 to 37.

The average SUS score of the Ecopoly XR Edition app was 76.3, putting it in the 70-80% percentile bracket. The app and game are easy to understand and use for all age groups. Furthermore, the digitization reduced the play time of a single session from two hours to 48 minutes. This makes a session much easier to fit into a class hour.

The player questionnaire evaluated the user experience with scores (in a 0 to 5 score) were: “I think the app was a useful addition” being 4.4, and “I had enough time to see the game elements in the app” being 4.4.

The app and game were well received across all age categories. They found that the sweet spot for the current game setup is primary and secondary schools up to age 15. For older groups, and in settings where more time can be made available, players suggested more complex game mechanics. Extensions along these lines allow for a re-playable game, and a deeper engagement and exploration of the subject matter for the players. The feedback from the classroom teachers was positive, suggesting the use of the current game in citizenship education, and education around political and social processes.

### 3.2.3 XRAQUA

The pilot programme took place across 3 schools, engaging ≈ 150 students (ages 13–17) and 16 teachers. Workshops and XR lessons were organised from March to October 2025, ensuring continuous interaction between students and their aquaponics systems.

The key activities performed included:

• Workshops: over 20 sessions held, combining hands-on training, VR use, and parameter analysis.

• Daily engagement: students dedicated on average one hour per day to aquaponics-related tasks.

• Knowledge improvement: average diagnostic test = 50.03%; final quiz = 97.33% → +94.6% gain.

• STEM motivation: 85% of participants reported higher interest in STEM; top areas were Environmental Science (72%), Biology (65%), Chemistry (54%), Digital Tools (41%), and Mathematics (29%).

• Confidence: 86% felt more confident applying STEM in practice; 91% would recommend the programme to other schools.

• Teacher satisfaction: Around 90 % of participating teachers expressed positive feedback and confirmed curricular alignment. The project was used across Biology, Physics-Chemistry, and Mathematics curricula, aligning with KPIs 3–5 and 7–9.

These outcomes confirm that the combination of physical experimentation, XR simulation, and online learning results in measurable cognitive gains and increased motivation.

### 3.2.4 SURE-VFT (Virtual Field Trips)

Developed a total of 3 virtual field trips and evaluated the pilots gathering feedback from 54 participants. The main achievements in terms of evaluation include:

- Likelihood of recommending the SURE-VFT results: 75.7%

- Satisfaction levels of students, parents, educators, and administrators with the educational experience: 73.5%

- Percentage of students who complete assigned activities on time or per task: 35%

- How well students recall or apply learned knowledge over time: 71.7%

- Level of involvement and participation in educational activities and discussions: 79.5%

- Number of users registered in the XR4ED platform for SURE VFT: 319

They organized one training session (demo applications) for tutors, with 9 participants, and one training session (demo applications) for learners, integrated into lectures during the Spring semester 2024-2025 and Winter semester 2025-2026, as part of the following courses:

- Remote Sensing (RS course (spring semester 2025): 88 students enrolled

- Digital Photogrammetry (spring semester 24/25): 8 students enrolled

- Digital Terrain Models (spring semester 24/25): 42 students enrolled

- Digital Photogrammetry (winter semester 25/26): 58 students enrolled

### 3.2.5 GLEAM

They completed two rounds of user testing with validation of accessibility, gameplay balance, and learning outcomes. They collected around 100 questionnaires and interviews during the 12 months duration of the project

A fully playable mixed-reality game was developed for mobile devices, integrating spatial audio, haptic feedback, and adaptive interfaces. The prototype was successfully tested with visually impaired and sighted players in both Germany and Italy, confirming the game's inclusivity and functionality.

Institutions' involvement: Based on the positive reception of the game idea, partnerships were established with 5 museums and 2 universities, integrating mythological and cultural content into the game narrative. (MART, Oriente Occidente, Albisola cooperative, Museo del Cinema, Tactile Vision cooperative, DBSV in Germany).

Museums, primary and high schools, university students' feedback: Over 80 participants were directly involved in co-design workshops, accessibility discussions, and playtesting sessions. Feedback from these sessions informed multiple design iterations and accessibility improvements. The educational framework was validated through university lectures and testing sessions at HTW-Berlin and the University of Turin.

Professionals' feedback, including teachers and blind educators: Professionals' feedback, including insights from industry leaders (Ubisoft, Inclusive Gaming), teachers, museum educators, and blind education specialists, confirmed the game's strong pedagogical value, accessibility design, and potential for use in inclusive learning environments.

### 3.2.6 TEXIVOS

They validated their pilots with 58 end-users (students/teachers) by requesting them to answer questionnaires. They achieved a completion rate of 100%, and a self-reported improvement in learning outcomes of 75% (with 70% of stakeholders assigning a value of 4 in a Likert scale where 1 means low and 5 extremely satisfied). The main results from their evaluation include:

- Experience with VR/XR during the course: Participants described their experience with VR/XR during the course as very positive. The hands-on activities, like using VR headsets and doing simulations, were particularly immersive and engaging. While some initially experienced minor discomfort (dizziness, disorientation), these issues were quickly solved as they became more familiar with technology. Many emphasized how the course "demystified VR/XR", making it more accessible and less intimidating. The practical approach based on direct experience was widely appreciated.

- Engagement: The impact of VR/XR on teaching and student engagement was a major focus of the interviews, and the feedback was again very positive. Teachers observed a marked increase in student engagement, particularly among younger learners, when VR/XR was added into lessons. Technology was seen as a powerful tool for visualizing complex concepts (reconstructing virtual environments, simulating real-world scenarios, etc.). Confirming the answers to the previous question, participants highlighted the importance in improving comprehension.

- Suggestions for improving digital skills: When asked about improving digital skills, participants demonstrated a mood going from neutral to positive. While there was a consensus about practical, experiential training, the collected suggestions were quite wide and not very focalized - hands-on workshops, webinars, and video tutorials. Teachers also asked for support and access to resources, such as online forums or mentorship programs. On one aspect all the participants agreed: the need for institutional investment in technology and infrastructure for sustaining long-term digital skill development.

- Barriers to digitalization in construction education: This topic revealed a neutral mood, with participants expressing a mix of optimism and concern. Some were uncertain about how to overcome challenges like limited access to technology or even resistance to change. However, others believed that simulations, practical tools, and targeted training could help reduce the digital divide.

### 3.2.7 AGRITWIX

The final version was tested through seven pilot activities: one at the University of Bologna within the course Precision and Sustainable Agriculture, two in Higher Technical Institutes for Agri-food Innovation, and four in an agricultural high school in the Emilia-Romagna region (Cesena). In total, more than 130 students and teachers used the tool and provided structured feedback through questionnaires. The first set of seventy responses, collected during the high-school pilots, was analyzed in depth to improve clarity of instructions, interface usability, and performance. The later pilots confirmed the effectiveness of these adjustments and the educational potential of the tool.

The results of the pilots show that students appreciated the opportunity to learn agronomic concepts in an interactive, visual, and experimental way. Many respondents highlighted the originality of combining technology and agriculture, the clarity of the irrigation decision process, and the usefulness of visualizing how water management affects plant development. Suggestions for improvement focused mainly on graphical performance, loading times, and the inclusion of short tutorials or additional crops, all of which provide valuable guidance for the next development cycle.

All technical and educational objectives were achieved. Participants found AgriTwix educational, engaging, and innovative. Many emphasized its value for teaching sustainable irrigation.

### 3.2.8 XR2CLASS

This project deployed 2 XR learning modules (Databases & Gamification Principles), and conducted 3 pilot sessions with a total of 121 participating students, using the XR2CLASS platform in the university's XR laboratory.

During the evaluation process, they reported the following KPIs:

- Measure on active user interaction with the XR platform, indicating the platform's effectiveness and user interest, aligning with XR4ED's focus on engaging educational experiences. This goal was achieved through interactive and immersive content designed to captivate users. Quality Verification: Monitored through analytics tools tracking user activity and engagement metrics. They recorded an average session duration of 31.5 minutes and activity completion rate above 90%, confirming strong engagement.

- Measured an average 27% improvement in post-assessment test scores, demonstrating enhanced skill acquisition and understanding of database concepts. Quality Verification was evaluated using pre- and post-assessment tests to measure improvements in user competency.

- Collected user satisfaction feedback averaging 87%, with qualitative comments emphasizing clarity, interactivity, and increased motivation through gamified features.

- User satisfaction and the likelihood of recommending the platform to others, which is important for assessing market acceptance. Enhanced user experience designed to increase advocacy and satisfaction. Quality Verification was collected through post-use survey responses. They reported a 24 % increase in user satisfaction.

- The effectiveness of the learning modules in engaging users to complete educational activities. Was achieved through compelling gamification and interactive content. For quality verification they tracked using the platform's internal analytics tools to monitor completion rates. Reported results shown 90% of activities completed across modules compared to the baseline of 67.5%.

In conclusion, from a pedagogical perspective, the pilot confirmed that immersive visualization and gamified progression substantially improved engagement and learning outcomes. Students reported that the 3D representation of relational tables and data flows made abstract database concepts more intuitive and helped them correct logical errors faster. The real-time feedback delivered by the avatar and on-screen messages supported self-paced exploration and reduced cognitive fatigue during problem-solving. Participants also noted that competitive leader board and badges encouraged persistence and created a positive sense of achievement, particularly in longer tasks. Observational data reinforced these impressions: most participants remained focused throughout the sessions and voluntarily repeated exercises to improve their performance.

### 3.2.9 VR4BEES

The VR4BEES consortium successfully developed a comprehensive suite of multimedia and educational outputs, designed for both classroom and online learning environments. These resources promote interactive, bilingual learning, and support educators in teaching environmental topics through immersive

technologies. The developed and published 35 educational and VR videos. They ran 4 pilot sessions with 143 students and several educators.

Student questionnaires consisted of 8 questions consisting of 3 Likert-scale, 2 Yes/No, and 3 open-ended questions. The aim was to assess students' overall experience, learning outcomes, and engagement in a simple and age-appropriate way.

Educator questionnaire consisted of 10 questions: 6 Likert-scale, 1 Yes/No, and 3 open-ended questions. These were designed to capture teachers' perceptions of usability, pedagogical value, student interaction, and areas for improvement.

Across the four pilots, a total of 63 student feedback forms were submitted. Most students reported a highly positive overall experience, with 91.9% rating the VR activity as very good or excellent. Results also demonstrated a strong Net Promoter Score, as 89.2% stated they would recommend the activity to their friends, and 75.7% expressed interest in participating in similar sessions in the future, indicating strong willingness for continued engagement.

Knowledge retention showed notable success, with 78.4% confirming they learned something new about bees. Open-ended responses demonstrated that students retained key scientific facts such as the existence of over 20,000 bee species, the significance of pollination for food production, the bee dance as a form of communication, and the process of pollen collection using their body hairs. Some students also mentioned feeling less fear toward bees after understanding their role in the ecosystem.

Student engagement was exceptionally strong. Most students cited the VR glasses as their favourite aspect of the session, describing the experience as fun, exciting, and interactive. Several students also expressed interest in the XR4ED platform, noting the resemblance to video games that combine entertainment with learning.

Four educator feedback forms were collected in Cyprus. The responses confirmed that the VR component was easy to use and enjoyable, even for those with limited technical expertise. Teachers highlighted that students remained highly attentive and actively engaged throughout the session, and demonstrated clear comprehension and memory of key information immediately after the activity.

Educators affirmed the educational value of VR, pointing out that the immersive environment supported understanding and maintained motivation. Additionally, they emphasized the quality of the structured learning approach that transitioned from theory to interactive assessment and finally to virtual exploration.

### 3.2.10 HARMONY4ED

They developed a total of 4 modules, and each module was tested with educators, therapists, and vulnerable students, confirming its therapeutic value and usability. The pilot phase demonstrated statistically significant improvements in task completion (+81.6% on average) and student engagement (+97%), validating the solution's impact on learning effectiveness and motivation.

Among the KPIs that they measure we can find:

- Student engagement: 97% increase. Baseline data collected during workshops indicated that the average level of student engagement with VR-based activities in regular educational settings was 1.68 on a 0–5 scale, reflecting limited or occasional use. After introducing the HarmonyVR platform, the mean score rose to 3.32, representing a 97% improvement in engagement and applicability. Participants emphasized that the application was easy to use, visually appealing, and enjoyable, with an overall user satisfaction rating of 4.7/5. The increase in engagement demonstrates the effectiveness of music-based gamified learning and confirms the high motivational potential of immersive XR tools in working with individuals on the autism spectrum.

- Completion Rate of Educational Tasks: 81.6% average improvement. Evaluation sessions were conducted with a group of students who completed multiple game iterations in different scenarios. Statistical analysis using the Wilcoxon rank test confirmed significant improvements in task completion across all modules ($p \leq 0.05$). Scenario 1 (Math Equations): +110.75%, $p = 0.029$; Scenario 2 (Memory & Social Skills): +37.12%, $p = 0.003$; Scenario 3 (Focus & Listening): +97.00%, $p = 0.018$. These results confirm that the application not only enhances cognitive and sensory performance but also promotes better focus, memory, and accuracy among users with ASD.

- Training completion: 20 trained professionals. A complete training cycle was delivered to 20 educators, therapists, and university lecturers. The sessions combined theoretical and practical modules on the use of VR and music therapy principles in educational and therapeutic contexts. Participants acquired competencies enabling them to implement the HarmonyVR tool in everyday work with children on the spectrum. The training ensured effective knowledge transfer and laid the foundation for long-term adoption of XR-based learning in special education.

### 3.2.11 Conclusions

The ten pilot projects demonstrated strong evidence of the effectiveness, engagement, and educational value of XR-based learning tools across diverse subjects, age groups, and educational contexts. All projects successfully deployed immersive modules and conducted structured pilot evaluations, confirming both usability and pedagogical impact.

Across the projects, participant engagement was consistently high. Student satisfaction rates ranged from 73% to over 90%, while teachers and educators reported positive experiences in integrating XR tools into curricula. Notably, XR-based training consistently improved learning outcomes, with measured knowledge retention gains up to 94.6% (XRAQUA) and task completion improvements exceeding 80% (HARMONY4ED). These results highlight the accelerated transition from conceptual understanding to practical competence compared to traditional teaching methods.

The projects also showcased the versatility of XR in addressing diverse educational needs: from STEM and agriculture (XR2LEARN, XRAQUA, AGRITWIX) to cultural heritage and accessibility (GLEAM), environmental education (ECO-EXT, VR4BEES), database and gamification concepts (XR2CLASS), construction and VR skill development (TEXIVOS), virtual field trips (SURE-VFT), and therapeutic applications for vulnerable learners (HARMONY4ED). In each case, immersive experiences enhanced engagement, motivation, and autonomy, while providing measurable cognitive and practical benefits.

User feedback confirmed the clarity, intuitiveness, and interactivity of XR platforms, with repeated suggestions emphasizing the value of hands-on, experiential learning. Several projects also demonstrated inclusivity and accessibility, enabling participation from visually impaired learners (GLEAM) and supporting neurodiverse students (HARMONY4ED). Teacher and institutional feedback highlighted curricular alignment, ease of integration, and the potential for XR tools to complement conventional methods while enabling innovative learning pathways.

Overall, the aggregated results indicate that XR4ED solutions not only improve learning effectiveness and motivation but also foster rapid skill acquisition, sustained engagement, and positive attitudes toward technology-enhanced education. These outcomes strongly support the adoption of XR-based approaches in formal and informal educational contexts, demonstrating their potential to enhance both pedagogical quality and learner experience across Europe.

## 3.3 Final demos evaluation @ UnitedXR

For the final event held in December 2025 in Brussels, XR4ED was represented with a large exhibition booth at the UnitedXR Conference. Table 4 shows the schedule for all the demos that were shown during the conference, from both developers and pilot projects.

| XR4ED @ UnitedXR, Exhibition in Brussels | | | | | | |
|---|---|---|---|---|---|---|
| **09/12/2025** | ***Developers*** | **Table 1** | **Table 2** | **Table 3** | **Table 4** | **Table 5** |

| | | | | | | |
|---|---|---|---|---|---|---|
| **TUESDAY morning** | **11:00-12:30** | XPLORER | VERTEX | ICEVR | TUTORIST | LEARNING LOOP |
| | **12:30-14:00** | SURGIVIZ4D | SIGNCOMM | HXT | MYVIRTUALTRIP | EXCURIONXR |
| ***BREAK*** | | | | | | |
| | ***Pilots*** | **Table 1** | **Table 2** | **Table 3** | **Table 4** | **Table 5** |
| **TUESDAY afternoon** | **14:00-15:30** | GLEAM | AGRITWIX | VR4BEES | TEXIVOS | ECO-EXT |
| | **15:30-17:00** | XR2CLASS | XRAQUA | XR2LEARN | HARMONY4ED | SURE-VFT |
| **10/12/2025** | ***Developers*** | **Table 1** | **Table 2** | **Table 3** | **Table 4** | **Table 5** |
| **WEDNESDAY morning** | **10:00-11:15** | SIGNCOMM | SURGIVIZ4D | HXT | ICEVR | VERTEX |
| | **11:15-12:30** | TUTORIST | LEARNING LOOP | MYVIRTUALTRIP | EXCURIONXR | XPLORER |
| ***BREAK*** | | | | | | |
| | ***Pilots*** | **Table 1** | **Table 2** | **Table 3** | **Table 4** | **Table 5** |
| **WEDNESDAY afternoon** | **12:30-13:45** | HARMONY4ED | XR2CLASS | AGRITWIX | GLEAM | XRAQUA |
| | **13:45-15:00** | VR4BEES | TEXIVOS | ECO-EXT | SURE-VFT | XR2LEARN |

*Table 4. Demos agenda during UnitedXR*

Over the two-day exhibition, the 20 winning projects were given the opportunity to present their final demonstrations to the public and to collect direct feedback from event attendees. To support this process, a structured evaluation form was made available online to participants via QR codes displayed at the exhibition booths, providing immediate access to the questionnaire.

The primary objective of the form was to collect user evaluations for each demonstrated project. The questionnaire began with a brief explanation of the data collection purpose and clearly stated that participation was voluntary and anonymous (see Figure 14 left). Demographic information was then collected to characterize the profile of attendees who tested the demonstrations (see Figure 14 right). Participants were subsequently asked to indicate the specific demo they had experienced (see Figure 15).

The final section of the questionnaire comprised a set of questions designed to assess user experience, engagement, and sense of presence, followed by an open-ended question aimed at capturing additional

qualitative feedback and recommendations (see Figure 16). Attendees were able to complete multiple questionnaires corresponding to each demo they had tested.

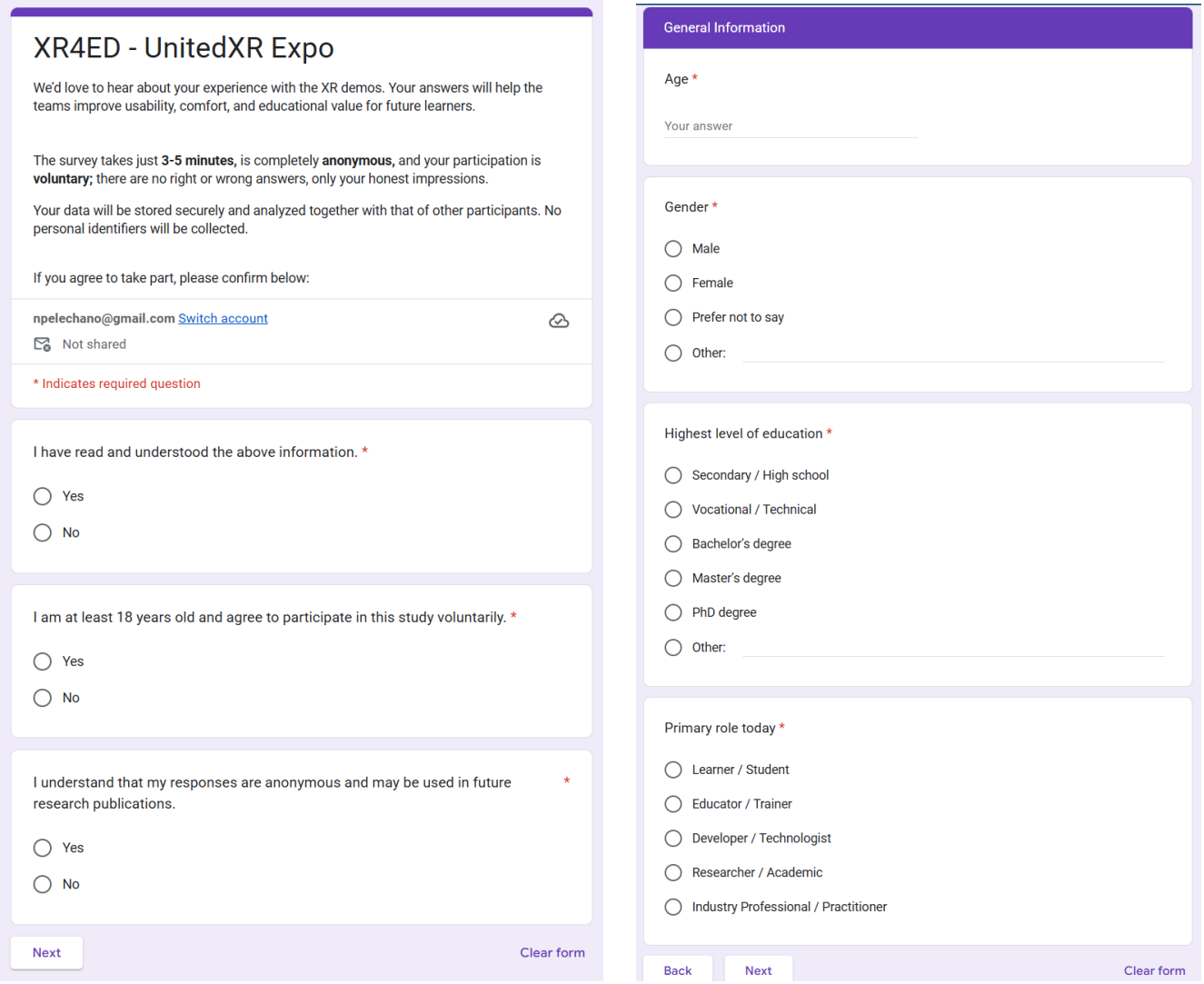



*Figure 14. Information and consent form for the user evaluation (left) and demographics questions (right)*

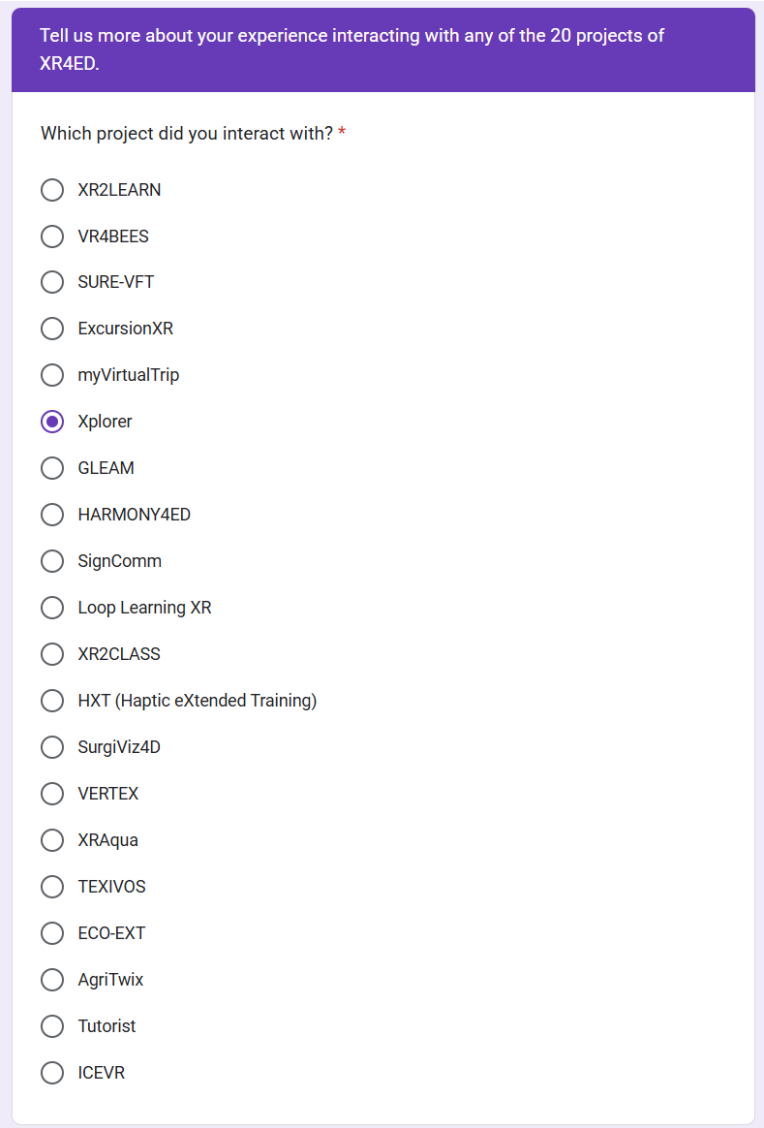



*Figure 15. Demo selection for the user evaluation*

This system's capabilities meet my needs. *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

This system is easy to use. *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

How likely is it that a person could live for 500 years? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Not at all likely | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Extremely likely |

Mental Demand: How mentally demanding was the task? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 = Very Low / Easy | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | 10 = Very High / Demanding |

Physical Demand: How physically demanding was the task? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 = Very Low / Easy | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | 10 = Very High / Demanding |

Temporal Demand: How hurried or rushed was the experience? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 = Very Low / Easy | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | 10 = Very High / Demanding |

Performance: How successful were you in achieving your goals? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 = Not successful | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | 10 = Very successful |

Effort: How hard did you have to work to accomplish your level of performance? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 = Very low effort | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | 10 = Very high effort |

Frustration: How insecure, discouraged, irritated, or annoyed did you feel? *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 = Very low frustration | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | ○ | 10 = Very high frustration |

I felt present in the virtual/augmented environment. *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

The environment felt like a place I was really in. *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

The environment felt like a place I was really in. *

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Strongly Disagree | ○ | ○ | ○ | ○ | ○ | ○ | ○ | Strongly Agree |

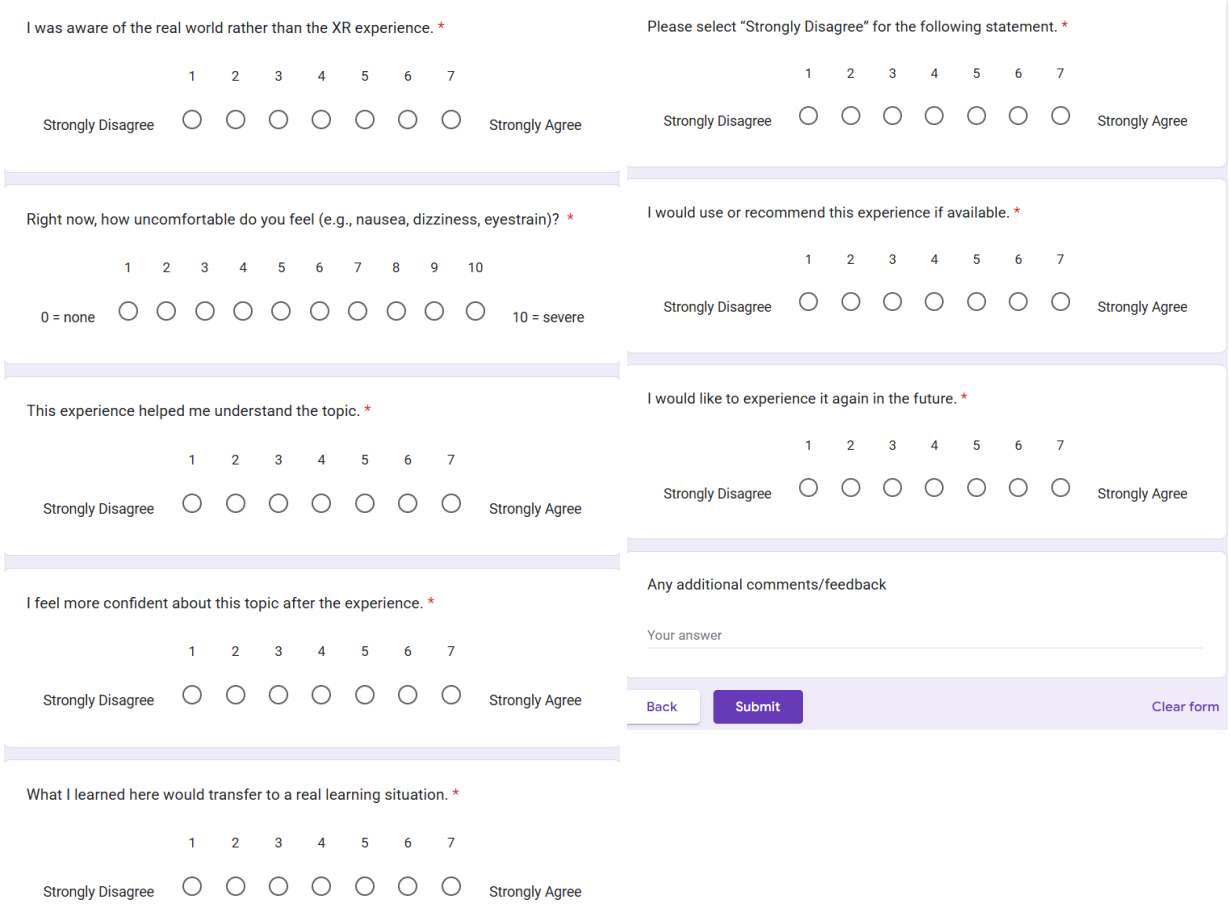



*Figure 16. List of questions for the user evaluation*

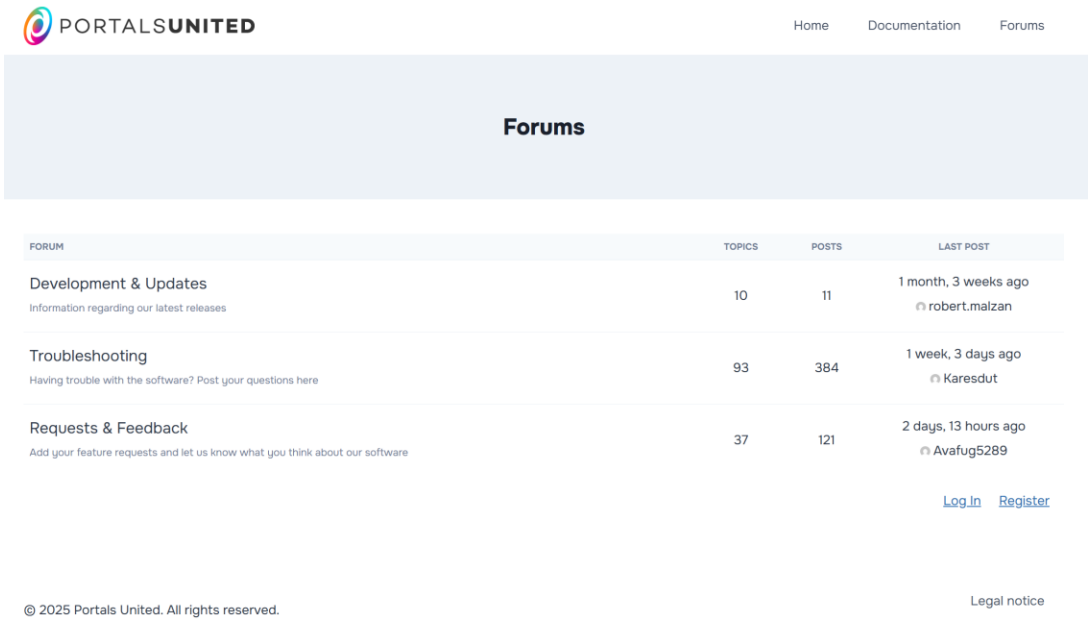
Responses to these questionnaires are have been collected and analysed. The detailed results have been included in deliverable D5.3. 125 responses have been gathered from participants attending the UnitedXR exhibition. These responses are particularly valuable, as most attendees were experts in eXtended Reality (XR). The results obtained highlight the high acceptability of the projects developed for education. Overall, the results were highly positive in terms of engagement, presence, user experience and value in education. The resulting projects also received high scores when asked whether they would recommend them to other teachers to use them in their courses.

# 4 Troubleshooting

Three dedicated online forums[4] were established to support and inform the winning projects on matters related to *development & updates*, *troubleshooting*, and *requests & feedback*. As illustrated in Figure 17, the XR4ED consortium provided a centralized communication online channel through which both developer and pilot project winners could access timely updates and raise questions concerning platform integration and usage or detected errors.

Forum activity metrics indicate a high level of interaction between the SMEs responsible for implementing the winning projects and the technical partners. Specifically, the *Development and Updates* section comprised 10 topics with 11 posts, *Troubleshooting* included 93 topics and 384 posts, and *Requests and Feedback* accounted for 37 topics and 121 posts. These figures reflect the intensive support efforts undertaken by the technical partners to facilitate the successful implementation of the funded projects.

In addition to maintaining active engagement through the online forum, the consortium organized biweekly meetings with the participating SMEs. These meetings enabled more direct, responsive, and personalized communication, further strengthening the feedback loop and accelerating issue resolution.




[4] https://www.portalsunited.com/support-forums/

*Figure 17. Forum dedicated to troubleshooting, development & updates, and Requests & feedback (capture taken on the 13th of December 2025)*

Figure 18 presents a detailed overview of the topics listed under the *Developments & Updates* section, illustrating that, on average, a new release of the World Builder and/or Portal Hopper was delivered on a monthly basis. Figure 19 presents a screenshot of the *Troubleshooting* forum, where both pilot- and developer-led projects could create new topics or contribute to existing discussion threads. This forum was actively used by the technical partners to rapidly disseminate solutions and guidance, ensuring that all SMEs involved in the development of the winning projects had timely access to relevant support information.

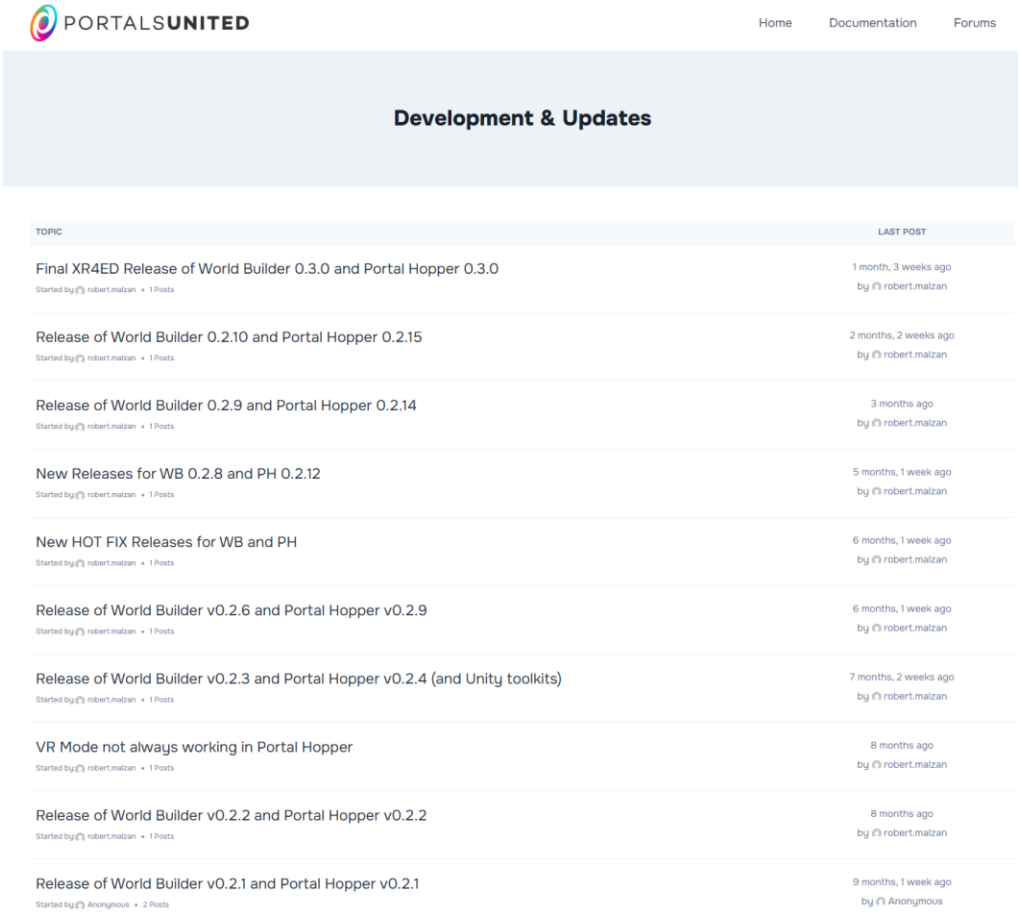



*Figure 18. Detailed list of topics included in "Developments & Updates"*

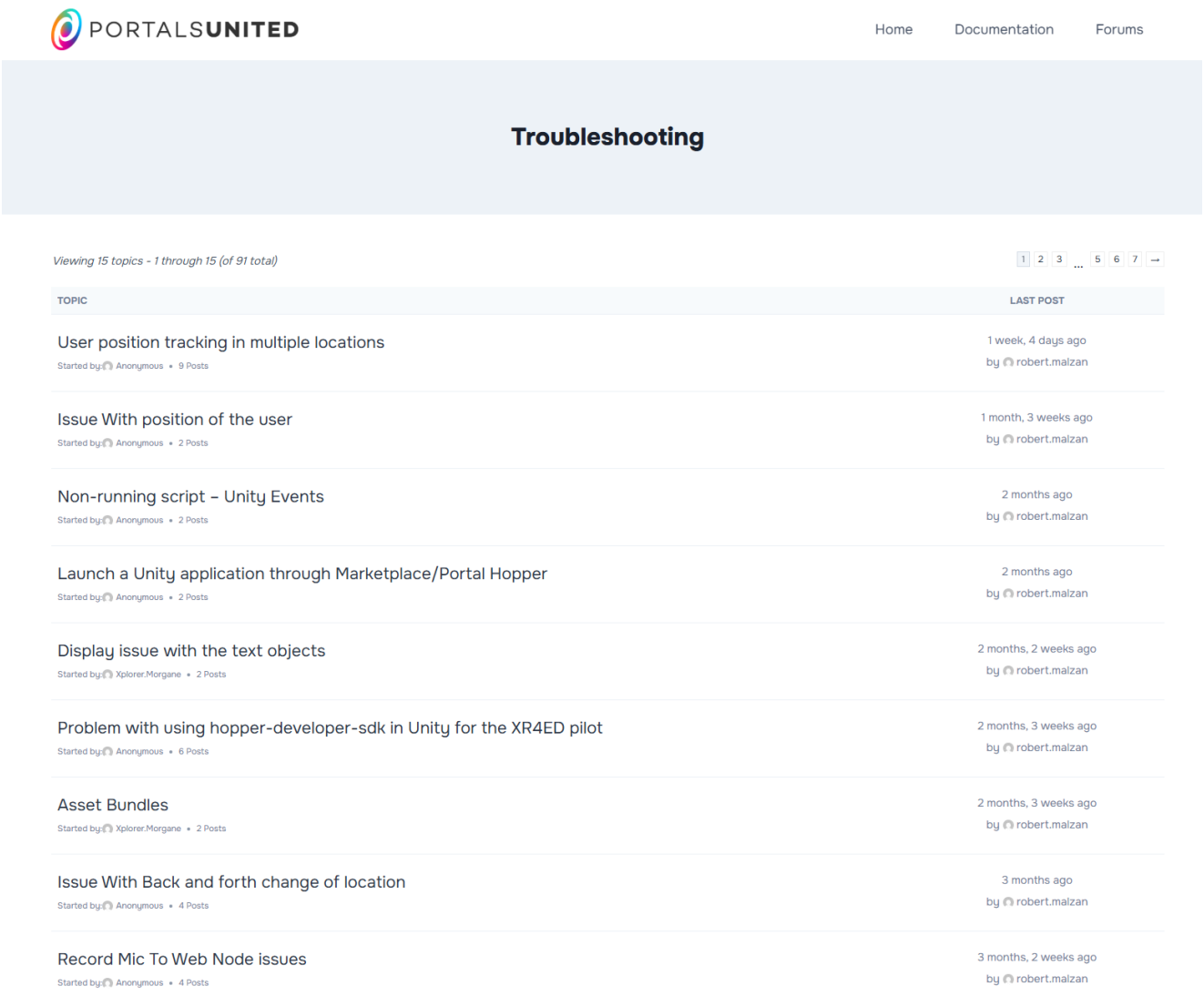



*Figure 19. Screen shot of the Troubleshooting forum*

## 4.1 Challenges

The most significant challenge encountered during the XR4ED project was the temporal overlap between platform implementation, the development of new modules by technical partners, and the simultaneous testing activities carried out within the pilot projects. While this parallel progression introduced coordination complexities, it also enabled an iterative refinement process that contributed positively to the robustness and overall quality of the platform.

However, this iterative development approach occasionally resulted in delays for the Pilot project winners, as progress on their developments was contingent upon the technical partner having sufficient time to implement required platform extensions, updates and improvements. In contrast, projects led by developers were less affected, as they were generally able to adopt alternative scheduling strategies and continue advancing other components of their modules in parallel. Pilot-led projects, by comparison, experienced a more pronounced impact on their planned timelines.

Despite these challenges, the platform was successfully completed and achieved a high level of maturity. This outcome was largely attributable to the systematic troubleshooting and iterative enhancement process undertaken during the final year of the project.

## 4.2 Portal Hoper & World Builder documentation

XR4ED made available a dedicated webpage for downloading both the Portal Hopper and the World Builder (see Figure 20), accompanied by comprehensive documentation and detailed installation instructions, as illustrated in Figure 21 and Figure 22. This webpage was actively maintained throughout the project's duration, with particularly frequent updates during the troubleshooting phase. During this period, additional information, documentation, and updates were provided in response to requests from project developers as they encountered issues in the development of their new modules and final demonstrations.

**Portal Hopper Installation/Update instructions**

Step 1: Download Portal Hopper (here)
Step 2: Run installer
Step 3: Select any destination location *
Step 4: At first start, allow firewall access *

**World Builder Installation/Update instructions**

Step 1: Download World Builder (here)
Step 2: Run installer
Step 3: Select any destination location *
Step 4: At first start, allow firewall access *

* are only required for first install

Please also check our documentation and our installation guide for more information.

**Downloads**

November 17, 2025
**Version Portal Hopper 0.3.1** Download Details

November 17, 2025
**Version World Builder 0.3.1** Download Details

*Figure 20. XR4ED website to download the Portal Hopper and World Builder*

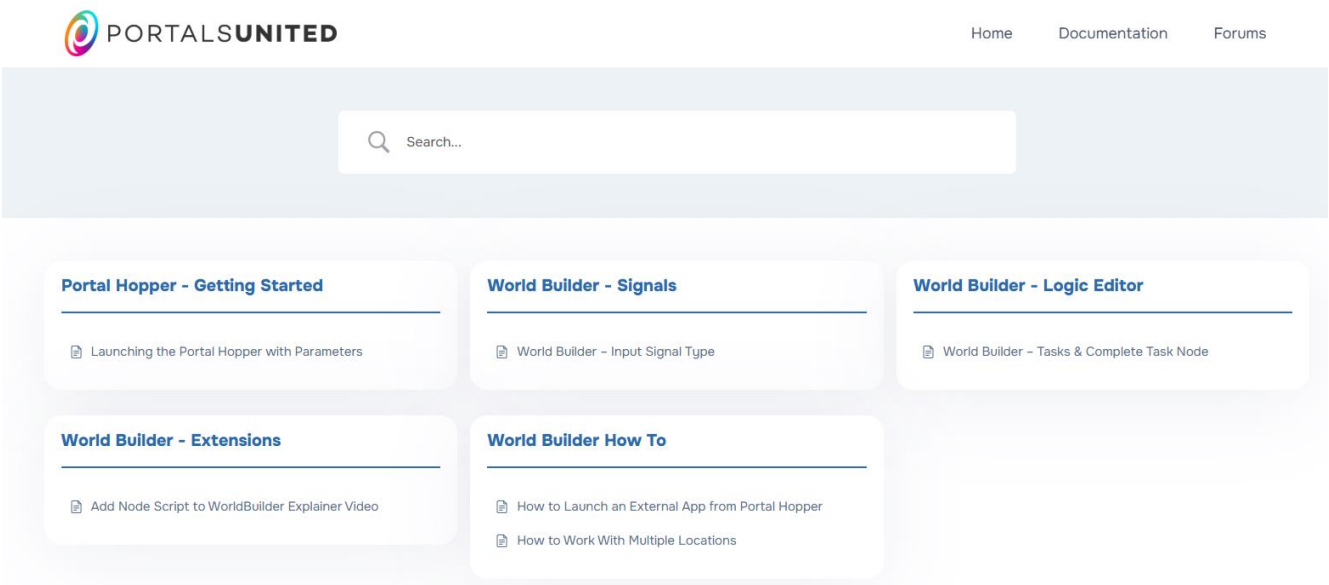



*Figure 21. Documentation available for the World Builder*

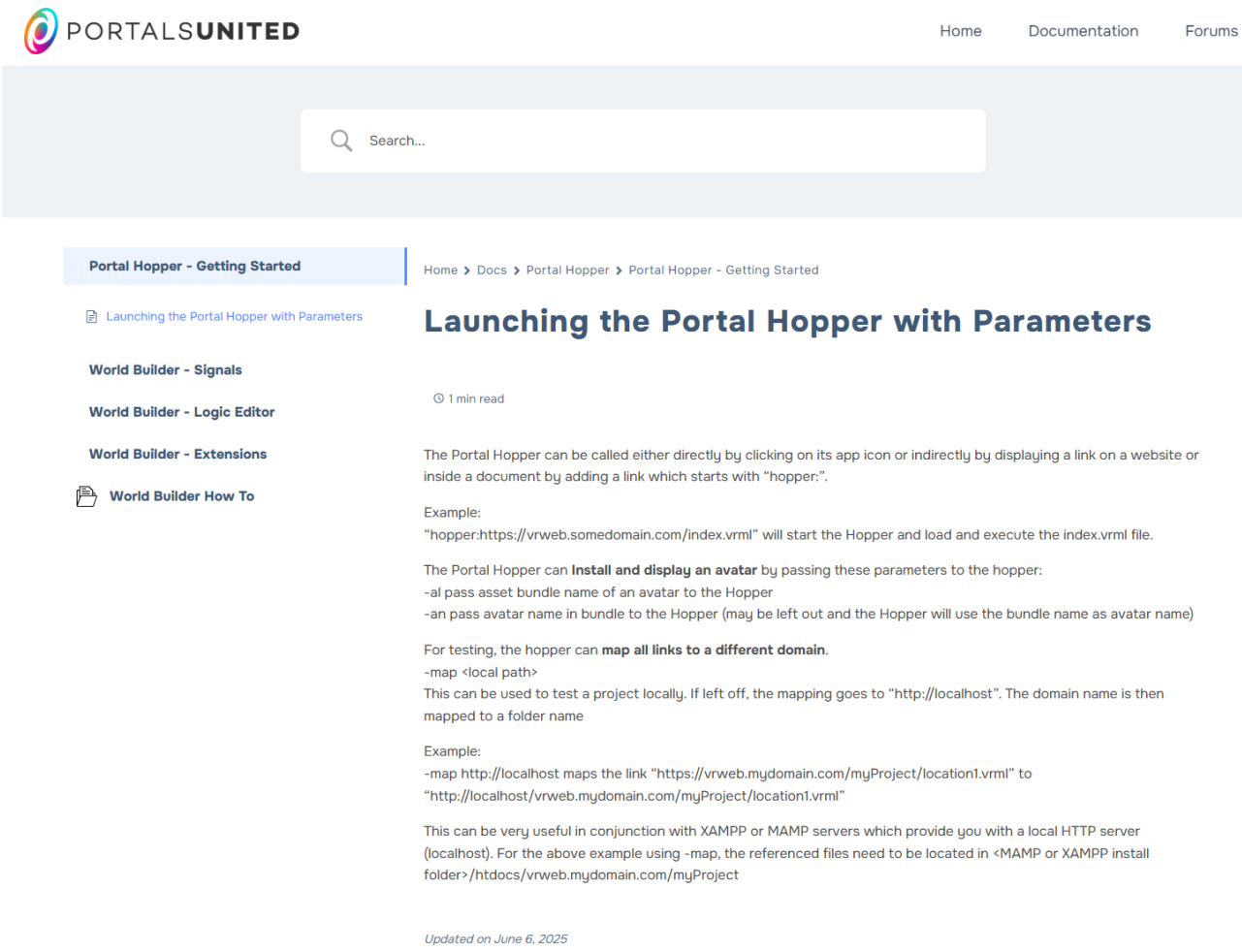



*Figure 22. Example of detailed documentation for getting started with the Portal Hopper*

# 5 Conclusions

The work presented in this deliverable confirms the effectiveness of the evaluation and support framework adopted within the XR4ED project and its key role in ensuring the quality, reliability, and maturity of the developed outcomes. Through a combination of systematic evaluation activities and continuous troubleshooting support, the project successfully accompanied the winning projects throughout their development processes. This approach enabled the timely identification and resolution of technical challenges, the refinement of design and implementation choices, and the overall strengthening of the robustness and usability of the implemented solutions.

Beyond the assessment of individual modules and pilot activities, the work described in this deliverable contributed to a broader understanding of how extended reality technologies can be effectively applied in educational contexts. Insights gathered across multiple iterations and use cases highlight the importance of user-centred design, iterative testing, and close collaboration between developers, educators, and end users. These factors proved essential in maximising user engagement, enhancing the sense of presence, and supporting meaningful and effective learning experiences. In addition to the evaluation activities carried out by the XR4ED consortium, each winning project conducted its own internal evaluations and user tests involving teachers and students. The winning projects were also highly active in dissemination efforts, including participation in events and the preparation of scientific publications.

Furthermore, the thorough user studies conducted during the development of the XR4ED platform modules by the consortium's technical partners resulted in research publications presented at leading international conferences in Virtual Reality and Human Computer Interaction. These contributions support the advancement of research in the field of immersive technologies for education and promote openness through the dissemination of research findings and, where applicable, open-source code has been made available to the wider research community.

Finally, the evaluation results and lessons learned provide a solid foundation for the future exploitation, replication, and scalability of the project outcomes. The methodologies, practices, and insights identified within XR4ED can be transferred to other XR-based educational initiatives, supporting the wider adoption of immersive technologies across European education and training systems. As such, this deliverable not only consolidates the achievements of the XR4ED project, but also contributes valuable knowledge that can inform future research, innovation, and policy actions in the field of immersive digital learning.