User centric Requirements for Enhancing XR Use Cases with Machine Learning Capabilities

Sueyoon Lee* Moonisa Ahsan† Irene Viola‡ Pablo Cesar§

*†‡§ Centrum Wiskunde & Informatica (CWI), Amsterdam, The Netherlands
§ Delft University of Technology (TU Delft), Delft, The Netherlands

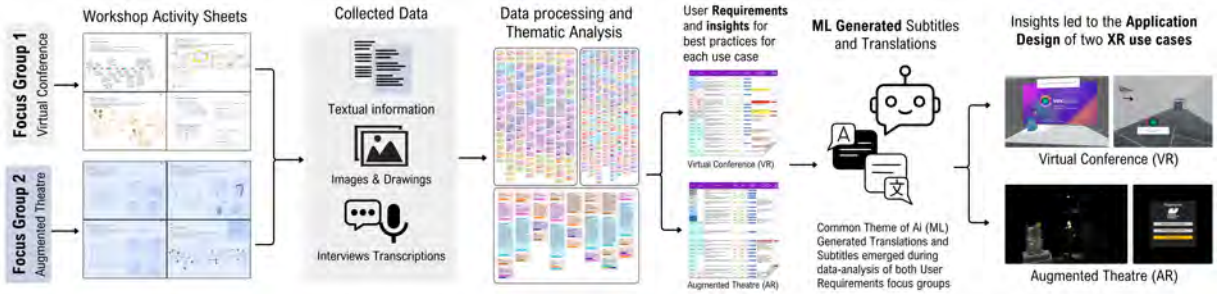

Figure 1: *Requirements Gathering Overview*: Starting from focus groups, finding common theme of Machine Generated Translations in both use cases and leading to the design decisions for the development of two XR Applications Virtual Conference and Augmented Theatre

Abstract

The combination of Extended Reality (XR) and Machine Learning (ML) will enable a new set of applications. This requires adopting a user-centric approach to address the evolving user needs. This paper addresses this gap by presenting findings from two independent focus groups specifically designed to gather user requirements for two use cases: (1) a VR Conference with an AI-enabled support agent and real-time translations, and (2) an AR Theatre featuring ML generated translation capabilities and voice-activated VFX. Both focus groups were designed using context-mapping principles. We engaged 6 experts in each of the focus groups. Participants took part in a combination of independent and group activities aimed at mapping their interaction timelines, identifying positive experiences, and highlighting pain points for each scenario. These activities were followed by open discussions in semi-structured interviews to share their experiences. The inputs were analysed using Thematic Analysis and resulted in set of user-centric requirements for both applications on Virtual Conference and Augmented Theatre respectively. Subtitles and Translations were the the most interesting and common findings in both cases. The results led to the design and development of both applications. By documenting user-centric requirements, these results contribute significantly to the evolving landscape of immersive technologies.

Index Terms: Virtual Reality, VR conference, Augmented Reality, AR theatre, Focus groups, User requirements, Use cases, Human-centric design.

*e-mail: sueyoon@cwi.nl
†e-mail: moonisa@cwi.nl
‡e-mail: irene@cwi.nl
§e-mail: p.s.cesar@cwi.nl

1 Introduction

As the boundaries between the real and virtual worlds merge [4], two scenarios, Virtual Reality (VR) Conference and Augmented Reality (AR) Theatre, emerge as innovative platforms where users can come together, share experiences, and interact meaningfully, even from a distance. By integrating AI-enabled support agents [14], real-time translation [11] and dialogue system capabilities[3], these immersive environments can address the diverse needs of participants, allowing for a richer user experience. This paper explores the requirements for next generation of Extended Reality (XR) applications with Machine Learning (ML) capabilities.

The requirements were collected in two focus groups, that were designed to understand and brainstorm (A) Ideas for the role and design of virtual agents and language translation at the VR conference and (B) Needs and wants of the users and the organizers on subtitles (Language translation) and VFX while watching a theatre play with AR glasses.

Based on these focus groups, our work (see figure 1) provides two primary contributions:

1. The design of a new VR Conference application, based on the requirements presented in this article.

2. Insights that guide the development of the AR theatre Application

2 Related Work

The user motivations for attending immersive experiences are complex and multi-faceted [9], which is why effectively designing these spaces is challenging [17]. Users are often driven by the desire for immersive sensory experiences [15], this fosters the need for emotional connections and intellectual stimulation. To effectively address the motivations in a VR conference settings like the one in [21], it is essential to implement features that enhance user experience [7]. For instance, Ellis [4] describes virtual environments as places where users can experience a heightened sense of presence and immersion, which are crucial for participant engagement during a conference. Lee et al.[10] suggest that a well designed VR lobby can facilitate remote social experiences, making it vital

to incorporate spaces that encourage interaction and collaboration. Furthermore, the implementation of live captions and accessibility features can significantly enhance the experience for users with varying levels of technology proficiency, as noted by Pidathala et al. [16], Augllo and Matamala [1] and Teófilo et al. [22]. Rothe et al. [18] focuses on presenting subtitles with HMDs and goes deeper into analysing the subtitles placement, the identification of speakers, and their effect on the VR experience. Pei et al. [14] reports a dedicated dialogue agent for virtual support in XR environments.

Traditional theatre users engage in AR theatres for a variety of reasons, including entertainment, social interaction, and personal enrichment [8]. Therefore, the experience design of AR theatre involves a aesthetically nuanced understanding of audience motivations [12]. Rzayev et al. [19] discuss the importance of real-time translation in AR settings, emphasizing that the position of captions can significantly impact the overall viewer experience. Hughes et al. [20] discusses the placement of captions within head-worn augmented reality devices, focusing on design considerations for one-on-one conversations; Ouzts et al. [13] observed the effect of caption placement on information intake via eye tracking; Brown et al. [2] found that the general preference towards following the head immediately is better than other behaviours regarding ease of locating subtitles. However, the study only focuses on subtitle behaviour and did not extensively explore other design aspects that could impact user experiences, like font size or colour.

In both of the VR and AR scenarios, AI-enabled Machine Translation plays a significant role in enhancing user experiences. Zhang and Bach [25] explore the concept of context-aware machine translation in virtual worlds, demonstrating how AI can enhance communication in immersive environments by adapting to the context in which it is used. Maka et al [11] demonstrate the extended approaches for AI-driven context-aware language translations. Wolk [24] prototypes the incorporation of domain-specific neural machine translation into AR systems, demonstrating how tailored AI solutions can improve communication in augmented contexts.

3 Focus Group 1: VR Conference

We held a focus group workshop session to gain insights about the needs and preferences of users and conference organizers regarding language translation, also to generate ideas for language translation at the VR conference.

3.1 Participants and Setup

The focus group workshop was conducted with six (2f, 4m) participants from the use case owner company leading immersive events. They were from diverse professions (director, marketing manager, head of production, etc.), experience in conference (attendee, organizer), experience in AR (intermediate=3, advanced=2, expert=1), and native language (Dutch=2, Hindi=2, Spanish=1, English/French=1). One moderator and one assistant ran the workshop in English, and no preparatory tasks were required.

3.2 Materials

A *presentation slide deck* was shared along with the focus group to guide participants through the session and present key information, including relevant images. 7 pages of A4-size *activity sheets* have been prepared for participants to freely draw and write their ideas for each activity.

3.3 Procedure

Introduction. The moderator briefly introduced the study and provided an activity workbook sheet, a pen, and stickers. Participants, a moderator, and an assistant introduced themselves with information including name, job position, expertise in VR and native language. **Warm-up.** Participants were asked to recall their experience attending a VR conference, details on the date, location, the purpose of attending, what they enjoyed the most, and one thing that could have been improved. The participants then briefly shared their experiences as a group.

Part 1: My current conference experience: Here the activities were designed to understand the current VR conference experiences and to find a design space for adopting language translation solutions. The first activity (1-a) involved participants visualising their activities on a timeline from entering the VR venue to leaving it. They used green and red stickers to mark positive and negative moments, and blue stickers to indicate moments they thought could be improved. Then, participants paired up to discuss the similarities and differences in their timelines. For the second activity (1-b), the moderator introduced applications of real-time language translation and subtitles by presenting visual examples via slides. Participants reflected on their previous experience of using language translation and selected one situation to sketch and explain. They also answered questions about why they had interacted with it, how it helped solve their problem and the pain points they had encountered.

Figure 2: (a) Participants (use case owners) brainstorming, discussing and generating ideas for the VR Conference environment and (b) voting for the best concepts; after the workshop briefing by the moderators.

Part 2: Bringing language translation to VR conference: This part was designed to help the participants brainstorm ideas on how to apply language translation to solve issues raised in Part 1. The moderator presented a specific imaginary scenario of the VR conference. The participants were asked to envision themselves in a situation where they arrived at a conference venue and the speaker started giving a presentation in French, a language they did not understand. Subsequently, they were instructed to generate ideas for interaction methods and interfaces for a language translation and subtitle model on a brainstorming sheet. The sheet consisted of eight empty boxes for sketching an imaginary scenario or screen interface. Each participant shared and explained their ideas and voted for their top two choices using stickers (Figure 2).

Part 3: Designing a future VR conference: This part aimed to foster collaboration and collectively design the optimal future VR conference. To stimulate creativity, we utilized the 'Braindrawing' technique [23] for brainstorming the future VR conference experience. Participants were divided into two groups and tasked with conceptualizing an ideal VR conference with real-time language translation and subtitles on a blank sheet of paper. These papers were then swapped between the groups to expand upon the initial concept. After another round of exchanges, the papers were returned to their original teams for further refinement of the concept. Ultimately, both teams shared and presented the final ideal scenario.

4 Focus Group 2: AR Theatre

We conducted the other focus group workshop session to gain insights into the requirements and preferences of theatre enthusiasts and festival organizers regarding live language translation and subtitles during theatre performances. We also generated ideas for the user interface (UI) and interaction approach between the audience and the AR glasses.

Table 1: Participants (use case owners) in Virtual Conference and Augmented Theatre Focus Groups.

User	Job	Focus Group	Conference and Theatre going experience	Experience with VR/AR Technologies	Native Language(s)
01	Marketing Manager	VR Conference	Organizer, Attendee	Intermediate	Dutch
02	Director	VR Conference	Organizer, Attendee	Expert	Dutch
03	Head of Production	VR Conference	Organizer, Attendee	Intermediate	Spanish
04	Digital Marketing Specialist	VR Conference	Organizer, Attendee	Intermediate	Hindi
05	Head of Partnerships and Business Development	VR Conference	Organizer, Attendee	Advanced	English / French
06	Marketing Coordinator	VR Conference	Organizer, Attendee	Advanced	Hindi
07	Retired School Teacher	AR Theatre	Intermediate	Novice	Greek
08	Project Manager	AR Theatre	Intermediate, Attendee	Novice	Greek
09	Creative Labourer	AR Theatre	Expert, Attendee	Novice	Italian
10	Actress, Director	AR Theatre	Expert, Attendee	Novice	Greek
11	Composer, Soundtrack Artist, Curator of Mixed Media Projects	AR Theatre	Expert, Attendee	Intermediate	Greek
12	Musician, Guitar Teacher	AR Theatre	Intermediate, Attendee	Novice	Greek

Figure 3: (a) The moderator introduces the goals to the participants (use case owners) at focus group session, (b) a potential stage onsite at theatre venue.

4.1 Participants and Setup

The focus group workshop was conducted with six (4f, 2m) participants from the use case owner company organizing theatre festivals. We collected their professions (director, project manager, composer, retired school teacher, etc.), experience in theatre (intermediate=3, expert=3), experience in AR (novice=5, intermediate=1), and native language (Greek=5, Italian=1).

4.2 Material

A *presentation slide* deck was accompanied along with the focus group. What set apart the first focus group slide was that taking into account the participants' limited AR experience, supplementary videos and information about AR were included. 6 pages of A4-size *activity sheets* similar to Focus Group 1 material (3.2).

4.3 Procedure

Introduction. The moderator gave the participants a short introduction to the study and provided an activity workbook sheet, pen, and stickers. Participants, a moderator, and an assistant shortly introduced themselves, sharing their names, positions, expertise in theatre, expertise in AR, and native language. **Warm-up.** Participants were requested to recollect a time when they attended a theatre performance and then choose their most memorable one. They also reflected on the details such as the date, location, language of the play, what they found most enjoyable, and one aspect that could have been enhanced, and then shared briefly as a group.

Part 1: My current theatre experience: Here the activities were designed to understand the current theatre experiences of the audiences and to find a design space for integrating AR solutions. In the first activity (1-a), the participants were instructed to write down or draw the activities they engaged in during their visit to a theatre, starting from the moment they entered until the moment they left the experience. They used green and red stickers to mark positive and negative moments, and brown stickers to indicate moments they believed could be enhanced. Once everyone finished creating their timelines, the participants paired up to share and discuss the similarities and disparities found within their respective timelines. Next (1-b), the participants were asked to provide a brief reflection on any language or communication issues they encountered during the theatre play. They were requested to explain the situation in detail, including why it posed a problem, whether they took any action to resolve the issue and suggest alternative methods to address such problems in the future.

Part 2: Bringing language translation into the theatre: These activities were designed to help participants learn about AR and brainstorm ideas on how to apply AR technology to solve issues raised in Part 1. The moderator first introduced the concept of AR and played the video [1] to help the understanding of AR possibilities with visual examples. Next, the participants were asked to imagine themselves wearing AR glasses while seated in the theatre, a scenario where two actors on the stage began speaking in Korean, a language that the participants did not understand. On the activity sheet, the participants brainstormed ideas for using AR glasses to aid in translation, focusing on designing the user interface for displaying subtitles and determining the interaction method. Once they finished their sketches, the moderator captured a photo of each idea and shared it on the screen. This allowed every participant to clearly see the drawings and contribute ideas as a group.

5 Evaluation and Results

5.1 Data Collection

Data collection included worksheets detailing users' expertise, experiences, preferences, and interactions with given VR/AR scenarios; timeline drawings visualizing their experience; and visual markers indicating user choices. Audio recordings of discussions were transcribed using *Dovetail* [2] for further analysis (see Table 2).

5.2 Data Analysis

We used Thematic Analysis [6, 5] to analyse the data and identified patterns that turned into tags and requirement categories mentioned in Table 2.

[1] The Future of Augmented Reality: 10 Awesome Use cases (Link) https://www.youtube.com/watch?v=WxzcD04rwc8

[2] Dovetail: https://dovetailapp.com/

Category	Type	Outputs and Results
Inputs	Data Collection	- *Text:* Worksheets with user introduction, level of expertise, past experiences, and preferences. - *Images and Drawings:* Timeline drawings visualizing experiences; visual markers (stickers, notes). - *Audio:* Recordings of discussion sessions later transcription text.
Method	Data Analysis	*Method:* Thematic Analysis [6, 5]. Two researchers independently reviewed and labelled transcription texts, organized labels into themes, and established user requirements. Identified patterns and relationships were turned into user tags and requirements categories.
Results	User Journey Tags	- *Positive Moments:* Memorable instances in AR/VR experiences. - *Negative Moments:* Pain points or cumbersome instances in AR/VR experiences. - *Could be Improved:* Instances that could enhance the user experience.
	User Requirements Tags	- *New Functions/Features*: Suggestions for new functionalities. - *Behavior/Action Patterns*: Agreed actionable practices. - *Requirements*: Technical or functional system requirements. - *Added Value*: Features enhancing overall experience. - *Raised Issues/Concerns*: Problems to address. - *Interface Suggestions:* UI improvement ideas. - *Interaction*: Practices for intuitive exploration.
Requirements	VR Conference User Requirements Categories	- *Communication*: using ML Generated subtitles and translations in multiple languages - *Virtual Conference Venue*: With entrance, lobby, trade shows, social areas and a conference room. - *Trade Show Booths*: Spatial booths designed for business interaction and engagement. - *Participant Representation*: Participants are represented as virtual avatars with interaction gestures.
	AR Theatre User Requirement Categories	- *Translation:* Using ML Generated subtitles and translations for the Greek play - *Audio Controls*: Users should have the option to control the volume. - *Program Controls*: Users can turn on/off the menu using interactive buttons. - *Active Speaker Highlighting*: Subtitles should highlight the active speaker. - *Subtitle Display*: Relative to the context using formats with default bottom placement. - *Eye-Strain Minimization*: Subtitles should be placed at a viewing angle level to avoid eye strain.

Table 2: Overview of the Inputs (Data Collection), Method (Analysis), Results and Requirements (Categories) for both applications.

5.3 Results

We observed that the most interesting and common findings in both focus groups were around the *Subtitles and Translation* generated using ML capabilities (see Table 2 for details and further results), with need for AI-enabled real-time translation of text and voice, supported by a voice-activated VFX system. The users should be able to communicate and translate both text and voice. Essential controls include muting, volume adjustments, and toggling interactive options for subtitles and translations. In multi-speaker settings, subtitles should automatically emphasize the active speaker, especially the one the user is focusing on, using intuitive placements such as speech bubbles or text strips. To avoid eye strain, subtitles should be placed in level to viewing angle and also ensure not to mask or block important parts of the display or screen. For HMDs, positioning subtitles at the eye level reduces the screen obstruction and eye strain. The subtitles and translation experience should promote inclusivity and accessibility, enabling attendees from around the world to fully participate in the conference and engage with conference content in a meaningful way.

5.4 Application Interface Screenshots

The results led to the design and development of both applications VR Conference (see figure 4) and the AR Theatre (see figure 5); both applications are user centric, scientifically validated through both qualitative and quantitative evaluation, and hosted by industry partners for real-world deployment.

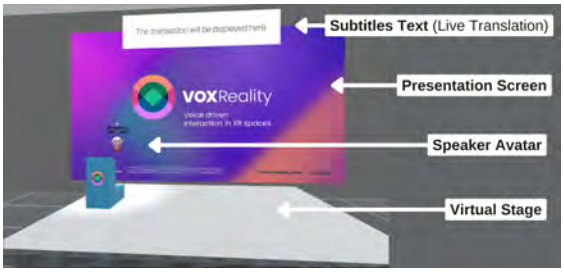

Figure 4: VR Conference Application Interface Screenshot

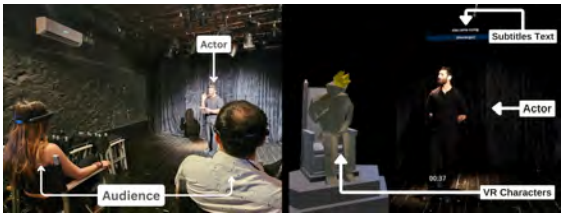

Figure 5: AR Theatre (a) real-time and (b) application screenshot

6 Conclusion

This paper has explored the evolving area of XR when combined with ML. In particular, based on user-centric technologies, it has a number of requirements for future applications like VR Conference and AR Theatre. Working with the domain experts, the result is two working applications which, in the future, will be evaluated in experiments and focus groups. The integration of machine translation and intuitive user interface design stands as a testament to the transformative potential of immersive technologies within AI-supported applications. The work reported in this article informed the design and development of two novel applications: VR Conference and AR Theatre by documenting user-centric requirements.

References

[1] B. Agulló and A. Matamala. (sub) titles in cinematic virtual reality: a descriptive study. *Onomázein*, (53), 2021. 2

[2] A. Brown, R. Jones, M. Crabb, J. Sandford, M. Brooks, M. Armstrong, and C. Jay. Dynamic subtitles: the user experience. In *Proceedings of the ACM international conference on interactive experiences for TV and online video*, pp. 103–112, 2015. 2

[3] W. Deng, J. Pei, Z. Ren, Z. Chen, and P. Ren. Intent-calibrated self-training for answer selection in open-domain dialogues. *Transactions of the Association for Computational Linguistics*, 11:1232–1249, 2023. 1

[4] S. R. Ellis. What are virtual environments? *IEEE Computer Graphics and Applications*, 14(1):17–22, 1994. 1

[5] J. Fereday and E. Muir-Cochrane. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1):80–92, 2006. 3

[6] G. Guest, K. M. MacQueen, and E. E. Namey. *Applied thematic analysis*. sage publications, 2011. 3

[7] S. N. Gunkel, H. M. Stokking, M. J. Prins, N. Van Der Stap, F. B. t. Haar, and O. A. Niamut. Virtual reality conferencing: Multi-user immersive vr experiences on the web. In *Proceedings of the 9th ACM Multimedia Systems Conference*, pp. 498–501, 2018. 1

[8] R. Hu, C. Wang, T. Zhang, T. Nguyen, V. Shapoval, and L. Zhai. Applying augmented reality (ar) technologies in theatrical performances in theme parks: A transcendent experience perspective. *Tourism Management Perspectives*, 40:100889, 2021. 2

[9] V. Komianos. Immersive applications in museums: an analysis of the use of xr technologies and the provided functionality based on systematic literature review. *JOIV: International Journal on Informatics Visualization*, 6(1):60–73, 2022. 1

[10] S. Lee, A. Striner, and P. Cesar. Designing a vr lobby for remote opera social experiences. In *Proceedings of the 2022 ACM International Conference on Interactive Media Experiences*, pp. 293–298, 2022. 1

[11] P. Maka, Y. C. Semerci, J. Scholtes, and G. Spanakis. Sequence shortening for context-aware machine translation. *arXiv preprint arXiv:2402.01416*, 2024. 1, 2

[12] M. R. Marner, S. Haren, M. Gardiner, and B. H. Thomas. Exploring interactivity and augmented reality in theater: A case study of half real. In *2012 IEEE International Symposium on Mixed and Augmented Reality-Arts, Media, and Humanities (ISMAR-AMH)*, pp. 81–86. IEEE, 2012. 2

[13] A. D. Ouzts, N. E. Snell, P. Maini, and A. T. Duchowski. Determining optimal caption placement using eye tracking. In *Proceedings of the 31st ACM international conference on Design of communication*, pp. 189–190, 2013. 2

[14] J. Pei, I. Viola, H. Huang, J. Wang, M. Ahsan, F. Ye, J. Yiming, Y. Sai, D. Wang, Z. Chen, et al. Autonomous workflow for multimodal fine-grained training assistants towards mixed reality. *arXiv preprint arXiv:2405.13034*, 2024. 1, 2

[15] O. Petit, C. Velasco, and C. Spence. Digital sensory marketing: Integrating new technologies into multisensory online experience. *Journal of Interactive Marketing*, 45(1):42–61, 2019. 1

[16] P. Pidathala, D. Franz, J. Waller, R. Kushalnagar, and C. Vogler. Live captions in virtual reality (vr). *arXiv preprint arXiv:2210.15072*, 2022. 2

[17] E. M. Raybourn, W. A. Stubblefield, M. Trumbo, A. Jones, J. Whetzel, and N. Fabian. Information design for xr immersive environments: Challenges and opportunities. In *Virtual, Augmented and Mixed Reality. Multimodal Interaction: 11th International Conference, VAMR 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part I 21*, pp. 153–164. Springer, 2019. 1

[18] S. Rothe, K. Tran, and H. Hußmann. Dynamic subtitles in cinematic virtual reality. In *Proceedings of the 2018 ACM International Conference on Interactive Experiences for TV and Online Video*, pp. 209–214, 2018. 2

[19] R. Rzayev, S. Hartl, V. Wittmann, V. Schwind, and N. Henze. Effects of position of real-time translation on ar glasses. In *Proceedings of Mensch und Computer 2020*, pp. 251–257. 2020. 2

[20] C. Schipper and B. Brinkman. Caption placement on an augmented reality head worn device. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 365–366, 2017. 2

[21] F. Steinicke, N. Lehmann-Willenbrock, and A. L. Meinecke. A first pilot study to compare virtual group meetings using video conferences and (immersive) virtual reality. In *Proceedings of the 2020 ACM Symposium on Spatial User Interaction*, pp. 1–2, 2020. 1

[22] M. Teófilo, J. Nascimento, J. Santos, Y. Albuquerque, A. L. Souza, and D. Nogueira. Bringing basic accessibility features to virtual reality context. In *2016 IEEE Virtual Reality (VR)*, pp. 293–294. IEEE, 2016. 2

[23] A. Van Boeijen, J. Daalhuizen, and J. Zijlstra. *Delft Design Guide: Perspectives, models, approaches, methods*. BIS Publishers, 2020. 2

[24] K. Wołk. Incorporating domain-specific neural machine translation into augmented reality systems. 2020. 2

[25] Y. Zhang and N. Bach. Virtual babel: Towards context-aware machine translation in virtual worlds. In *Proceedings of Machine Translation Summit XII: Posters*, 2009. 2