Curiosity XR

Contextualizing Learning through Immersive Mixed Reality Experiences Beyond the Classroom

By Aaditya Vaze

A thesis exhibition presented to OCAD University in partial fulfillment of the requirements for the degree of Master of Design in Digital Futures.

Toronto, Ontario, Canada, April 2023

Abstract

The focus of education is shifting towards a learner-centered approach that highlights the importance of engagement, interaction, and personalization in learning. This thesis explores new technologies to facilitate immersive, self-directed, curiosity-driven learning experiences aimed at addressing these key factors. I explore the use of Mixed Reality (MR) to build a context-aware system that can support learners' curiosity and improve knowledge recall. I design and build “Curiosity XR,” an application for MR headsets using a research-through-design methodology. Curiosity XR is also a platform that enables educators to create contextual multi-modal interactive mini-lessons, and learners can engage with these lessons and other AI-assisted learning content. To evaluate my design, I conduct a user participant study followed by interviews. The participants' responses show higher levels of engagement, curiosity to learn more, and better visual retention of the learning content. I hope this work will inspire others in the MR community and advance the use of MR and AI hybrid designs for the future of curiosity-driven education.

Acknowledgments

Completing a Master's thesis is a significant accomplishment, and it would not have been possible without the support of many individuals. I would like to express my sincere gratitude to the following people for their invaluable contributions to my thesis project.

First and foremost, I would like to thank my parents for their unwavering love, support, and encouragement throughout my academic journey. My father has deeply inspired me to work on this project, he used to constantly take me and my mom on vacation trips and teach me about the things around us.

I am deeply grateful to my advisors, Dr. Alexis Morris and Dr. Ian Clarke, for their guidance, insights, and expertise in the field of academics. Their support and feedback were essential in the development of my thesis.

I would also like to express my gratitude to my partner, Renu Indapurkar, for her constant support and encouragement throughout the entire process. Her belief in me and my abilities provided me with the strength and confidence to complete this challenging project.

Finally, I would like to thank my thesis instructors, Dr. Emma Westecott and Dr. Cindy Poremba, and the Digital Futures cohort for their support. The collaborative and supportive atmosphere fostered by our cohort made this journey all the more enjoyable.

Table of Contents

Table of Figures

1. Introduction

1.1 Motivation

Since I was a child, I have always been curious about the world around me and eager to learn. I remember wondering about many things during my school days, but I didn't always get the answers I was looking for. However, everything changed when we got a computer at home, and I discovered the vast amount of information and knowledge it could provide. I felt like I had found a mentor and became more knowledgeable and confident every day.

As I grew older, I realized how computers and the internet are transforming the way we learn. Engaging with interactive content and watching videos seemed to be much more efficient than traditional learning methods. I realized that watching videos and engaging with interactive content makes the learning experience more enjoyable and effective, providing learners with more dynamic and engaging ways to learn.

My curiosity to learn how computers work and to design interactive experiences brought me to where I am today, pursuing my Master in Design at Digital Futures in Toronto. Recently, during a trip to Goa in India with my family, I saw sea creatures such as octopuses for the first time in my life. After watching them move and holding them in my hands, I realized that engaging content that lacks immersion, interactivity, multimodality, or active participation is just a projection of true knowledge and doesn't necessarily drive curiosity.

Having seen my mother teach students for 30 years and having experience teaching kids from K-12, I understand the importance of multimodality and dynamism in learning content. I learned that people have different learning preferences and interests, and there is a need for learning content that is immersive, dynamic, and self-driven, where technology can assist them with their curiosity.

I believe technology-assisted learning can be improved by providing tools that allow learners to take ownership of their learning process, adapt the learning experience to individual needs and preferences, and fully immerse the learner in the content with multiple senses. With my thesis, I hope to address these issues and create immersive and interactive learning experiences that drive curiosity and motivate learners.

1.2 The challenge of context and curiosity support in Education

Context and curiosity are two critical factors that influence learning and education. Providing appropriate historical, physical, practical, or cultural context to the content helps to make learning more meaningful and relevant while fostering curiosity and encouraging learners to explore and discover new ideas and concepts on their own (Brown et al., 1989). However, supporting context and curiosity in education can be challenging, as it requires a deep understanding of the learners and their needs, as well as the ability to design learning environments and experiences that can accommodate these factors.

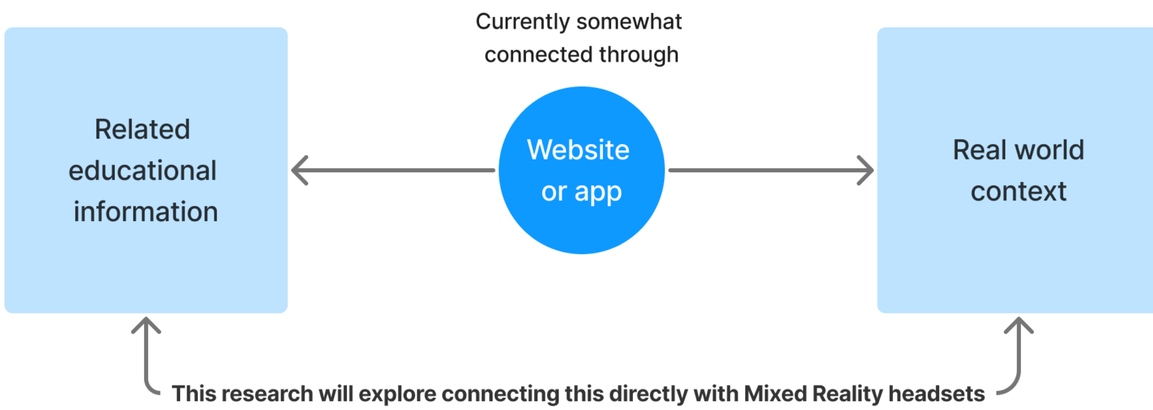

Figure 1: A Diagram to visualize the focus of this research to connect educational information to the real-world context using Mixed Reality headsets.

In the realm of formal education, there are instances that demonstrate the limitations of our ability to directly experience phenomena during the learning process. For example, when acquiring knowledge about nutrients, we do not interact with the soil's texture. Likewise, when learning about gravity, we do not physically experiment with a bouncing ball. Similarly, we do not traverse hills when studying topography, nor do we immerse ourselves in the water cycle while standing in the rain. Additionally, when learning calculus, we do not feel the expansion of a balloon. These examples emphasize the lack of direct experiential engagement with the subjects studied and the potential limitations of traditional teaching methods in delivering a comprehensive understanding of the material.

In some cases, experiential labs do offer a hands-on approach to learning, where students engage in practical activities to acquire knowledge and develop skills. This type of learning is often more effective than traditional formal education, which relies on lectures, textbooks, and exams. Although lab-based learning provides a more immersive experience, it faces challenges such as the need for expensive equipment, materials, and space to support interactive information and 3D models across multiple domains. For instance, while learning about a plant cell, we do not touch a plant or watch it grow, and we may not question its colorful illustration compared to the green plant in our backyard. Traditional methods of instruction may limit learners' curiosity and prevent them from exploring non-traditional questions driven by direct experience.

Apart from these challenges, traditional modes of learning also have other issues. Instruction material for Mathematics concepts, for example, often focuses on abstract representations, making it difficult for learners to find them interesting or relevant to explore further. Laboratory experiences, while useful in demonstrating theoretical concepts, may rely on predetermined examples that may not appeal to all students. Such static materials may not address the diverse practical applications that students may be curious about, thus limiting their curiosity and desire to learn.

Human nature drives us to be curious and learn as we experience the world, and although curiosity varies from individual to individual, everyone possesses it. As we grow older, the concepts we learn become more abstract, and the applications, context, and motivation fade making it challenging to remain lifelong explorers and learners. Current 2D display technology, such as laptops and mobiles, restricts us to limited senses, primarily utilizing our vision and some auditory senses. However, our other senses, such as tactile, spatial, and olfactory, often do not play a role in the learning process at all. By addressing these limitations and incorporating a more experiential approach to learning, we can foster a more engaging, immersive, and effective educational experience.

1.3 How can the challenge of curiosity and context support in education be addressed?

Emerging technologies such as AI [Artificial Intelligence] (McCarthy, 2007), MR [Mixed Reality] (Milgram, 1994), and IoT [Internet-of-things] are advancing rapidly and possess the potential to fulfill the requirement of fostering curiosity among learners and supporting self-directed learning (Reiners et al., 2021). Combining these technologies could enable new forms of context-aware agents (Hong et al., 2001) using cloud-based infrastructure, as well as new forms of real-time visualizations which can support learning frameworks. Combining sources such as ego-centric sensing, environment cameras, and other wearable sensors makes it possible to gather context and visualize or augment educational content through Mixed Reality head-mounted displays. Together, this may enable learning experiences where users can build and discover contextual knowledge about their environment and drive personalized lessons based on their interests. To enable these experiences, systems built using context-aware agents are needed. These agents could gather context to support users in learning, awareness, and knowledge management need to be designed.

1.4 Research Summary

1.4.1 Problem Statement

Spatial context and presence are essential components of a learning experience. They refer to how well learners can perceive, understand, and engage with information in a given environment. Current 2D display technology restricts learners to limited use of spatial context and presence, which hinders their ability to explore objects and learn through them effectively. Traditional modes of learning often focus on generic pre-designed instruction material and may not consider information relevant to learners that motivate them.

1.4.2 Hypothesis

This thesis hypothesizes that learners' curiosity and knowledge recall can be enhanced when they are immersed in a contextually-relevant environment while engaging with educational content, as opposed to conventional learning methods, such as reading books, viewing 2D videos, or hands-on experiences. Furthermore, MR has the potential to effectively cultivate an individual's curiosity, leading to a more engaging and efficient learning experience.

1.4.3 Goals

This research aims to explore the potential of Context-aware Mixed Reality (MR) agents to enhance curiosity and knowledge recall in educational contexts. Specifically, this study aims to investigate how a Context-aware MR agent can be developed to support interactive learning experiences that leverage contextual information acquired from mixed reality devices. To achieve this goal, the following questions will be addressed:

1.4.4 Research Questions

How can a Context-aware Mixed Reality agent support curiosity and knowledge recall for education?

Sub-Questions

SQ1: How can a Context-aware educational system acquire context from mixed reality devices to support educational lessons?

SQ2: How can acquired context be used to present an educational concept to the learner?

SQ3: How can we build a curiosity-support agent through Mixed Reality?

SQ4: How can educators/adaptive-context aware systems create/generate environment-based multimodal mixed-reality lessons to teach concepts to the learner?

SQ5: How can Mixed Reality use multimodal content to support interactive learning experiences in context?

SQ6: How can we evaluate if such learning experiences support the learner's curiosity and retention of knowledge?

1.4.5 Approach

To address these questions, I start by exploring the literature for Mixed Reality (MR), Context, and learning pedagogies. I look to understand related MR applications and frameworks for Context agents to stage this research. Later, I design techniques and build prototypes to leverage the potential of Mixed Reality which supports learning experiences. Finally, I evaluate the efficiency of the last prototype in supporting learners' curiosity and the ability to recall information.

1.4.6 Contributions

This research contributes to the field of Mixed Reality learning environments. It would help digital learning content have a stronger connection to the real world which we experience in our day-day lives. It would also help educators and interaction designers understand the modalities, benefits, and limitations of such environments. The research would also contribute to the body of knowledge in understanding how various interaction techniques can be used to make natural interfaces for designing lessons in Mixed Reality.

1.4.7 Scope and limitations

Due to the time limitation and the broader nature of the target audience for this project, this research focuses on innovative interaction design techniques and their impact on curiosity support and improvement of knowledge recall. This research does not talk about curriculum design or supporting educational systems but aims to look at the broader picture of the concept and focus on subjects such as Science, Maths, and Language which are easier to connect to the real world and can serve as a proof-of-concept for this technique.

When measuring recall, the user tests and research focuses on short-term recall of around a week and not long-term recall. When measuring curiosity support, it is a qualitative analysis and provides an addition to the body of knowledge through documenting anonymous experiences, elements of fun, and surprise which drive self-motivation. It does not measure curiosity in an academic context.

1.5 Chapter Overview

This chapter introduced the Curiosity XR project, the thesis, and the inspiration behind it. It also mentions the research goals, expected contributions, and scope.

Chapter 2 begins with a literature review talking about different contexts important to understand this thesis. It introduces Mixed Reality (MR) and talks about the role of education in MR. The discussion then continues to define the context and explain context-awareness and agent frameworks to support context-aware environments. The last section discusses different pedagogies that could be used to support curiosity and knowledge recall.

In Chapter 3, the related works in academia are introduced as applications that use similar techniques to address the context and curiosity-support gap in learning techniques. It starts with an overview and compares different applications in different contexts. The discussion illustrates its relation with this thesis and how these learnings could be used for designing and building prototypes.

Chapter 4 discusses the use of a mixed-method approach as a research methodology to conduct this project. It mentions the use of design through research to iterate on ideation, designing, building, and testing prototypes. It also revisits the concept from a speculative design perspective to discuss concepts for a learning framework.

Chapter 5 discusses the iterative prototypes conducted during the thesis. It talks about how these prototypes were designed through various stages from ideation to reflection. It talks about the architecture of the system which progressively develops as the iterations occurred to help design and develop the final prototype.

In Chapter 6, the final prototype: Curiosity XR is presented. It talks about ideation, process, outcomes, and evaluation of the final prototype. The evaluation is performed as a user-participant study.

In conclusion, Chapter 7 of this thesis revisits its objectives and contributions. The discussion finally highlights the future work and limitations of the scope of this thesis.

2. Literature Review

To discuss this thesis project of contextual-aware mixed reality learning environments this section provides the background knowledge of Mixed Reality, its role in Education, an introduction to context and context awareness, a theoretical framework for curiosity, and pedagogical techniques to support the tools built in this project.

The first section introduces Mixed Reality and talks about how the tactility of physical objects combined with the dynamism of the virtual world allows mixed reality to enrich learning experiences. It also analyzes the state-of-the-art learning techniques in Mixed Reality. The next section introduces the concept of context awareness which is about computationally acquiring and processing information about the user and the physical environment. The last section introduces a theoretical framework for curiosity and talks about pedagogies related to contextual learning.

2.1. Background for Mixed Reality

2.1.1 What is Mixed Reality?

Mixed reality (MR) is a term that refers to a range of technologies and experiences that combine elements of the physical and digital worlds. This can include experiences such as augmented reality (AR), where digital information is overlaid with the physical world, and virtual reality (VR), where users are fully immersed in a digital environment (Milgram, 1994). Mixed reality technologies are often used to create new ways for people to interact with computers and digital information. MR enhances the physical environment with virtual objects while also enhancing the virtual environment with data from the physical world. The term XR is often also used as a generic expression covering both AR and VR (Çöltekin et al., 2020).

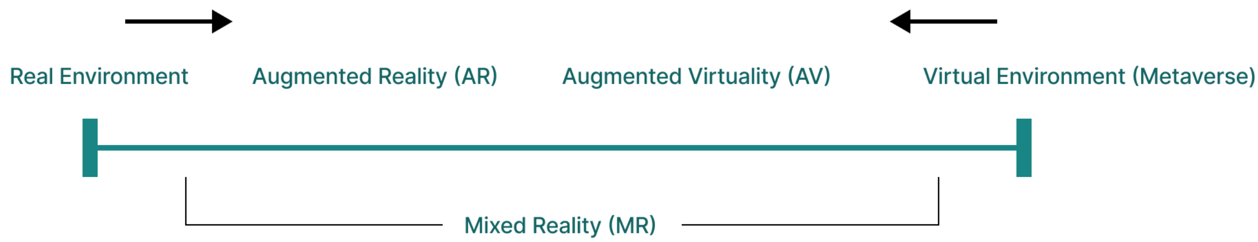

Figure 2. Reality–Virtuality Continuum adapted from Milgram & Kishino (1994)

MR is referred to with multiple existing notions. For this thesis, I focus on the notion; of "Strong AR" which considers MR as a "stronger" version of AR (Speicher et al., 2019). It is mainly characterized by an advanced environmental understanding as well as interactions, both of the users with virtual objects and the virtual objects with the environment.

2.1.2 Education in MR

MR has been applied to benefit several industries. Various academic papers talk about its application in fields such as healthcare (McCarthy et al., 2019), manufacturing (Egger et al., 2020), agriculture (Huuskonen et al., 2018), and maintenance (Siew et al., 2019); however, one of the most potential use cases implemented are in the field of education (Bacca et al., 2014).

Education in MR has been shown to improve the learning process and outcomes across various subjects for students. Recent research talks about how using sensory immersion, navigation, and information manipulation MR can improve student performance by stimulating learning with didactic materials (Hincapie et al., 2021). Another article presents the use of AR in mathematics to improve the understanding of vector geometry. The study proposes a modular augmented reality application that allows students to visualize geometric objects overlaid in the real environment, actively engaging them in the learning process. The app was found to be robust, easy to use, and effective in enhancing students' understanding of 3D space in a playful and didactic manner (Schutera et al., 2021). Kurubacak and Altinpulluk have reported that Augmented Reality (AR) has a plethora of advantages for students in the field of education including making courses enjoyable, lowering mental demands, boosting motivation and interest in the course, providing more opportunities to ask questions, encouraging interaction among students, creating new chances for personal learning, clarifying abstract ideas, and resulting in higher success rates (Kurubacak et al., 2017). With MR, students can interact with objects in both the virtual and real world, allowing for hands-on learning and increased motivation and engagement (Klopfer et al., 2008). Additionally, the ability to visualize abstract concepts makes learning more enjoyable and effective, even for complex topics (Sumadio et al., 2010), (Kurubacak et al., 2017).

Apart from the benefits of MR in education from the student's perspective, there have been numerous studies done on how technology can enhance the teaching experience as well. Researchers believe that MR is a particularly useful tool in this regard (Kaufmann et al., 2006). Kurubacak and Altinpulluk reported the benefits of AR for teachers including fostering creativity in students, ensuring active student participation in the course, and allowing students to progress at their own pace (Kurubacak et al., 2017).

These studies indicate that MR has a significant capacity to improve learning across various subjects, but limited attention has been given to an MR tool that can aid in the creation of educational content across these subjects. Previous studies have centered on mobile MR for education, while HMD(Head-mounted display)-based MR presents a possibility for a more engaging and immersive experience with the learning material.

2.2. Context and Context-awareness

2.2.1 What is context and context-awareness?

Literature has various sources to define context and context awareness. A definition for Context by Abowd et al states that "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves." (Abowd et al., 1999)

The concept of context-awareness was first introduced by Schilit et al. (1994), where the authors discussed the ability of mobile applications to detect and respond to changes in their surrounding environment. Later, the definition of context and context awareness was simplified and expanded upon by Abowd et al. (1999) and Dey et al. (2001). This definition states that context is any information that can describe the situation of an entity, whether that entity is a person, place, or object. The use of context awareness in mobile sensing and smart spaces was introduced by Rodden et al. (1998) and Essa (2000). Today, the vision of seamless interaction between smart devices and users has become a reality, and it is clear that this trend will only continue to grow in the future. The increasing availability of ubiquitous sensing has led to a growing demand for new applications and services that provide context-rich information anytime and anywhere.

Context-aware agents are artificial intelligence systems that are capable of understanding and adapting to the context of their environment (Hong et al., 2001). These agents use data from various sources, such as sensors and user input, to infer information about the context and make decisions based on that information. Context-aware systems, on the other hand, refer to larger systems or environments that incorporate context-aware agents. These systems can monitor and respond to various contextual factors and provide relevant services, applications, or interactions to the user (Hong et al., 2009).

Past research establishes a theoretical framework to provide a better understanding of context-aware systems. Perera et al. (2014) discuss the context life cycle and its four phases. First, the acquisition of context from various sources which could be physical sensors or virtual sensors (context acquisition). Second, modeling of the collected data in a useful manner(context modeling). Third, processing of the modeled data to derive high-level context information from low-level raw sensor data (context reasoning). Finally, the distribution of both high-level and low-level context to the consumers who are interested in context (context dissemination) (Perera et al., 2014). In another study, Hong et al (2009) talk about the various layers of context-aware systems. The four comprising layers of these systems are the network layer which supports context-aware systems and collects context information from sensors, the middleware layer which manages processes and stores context information, the application layer which provides users with appropriate services, and the user infrastructure layer which manages the interface to offer a suitable interface to users. These frameworks provide guidelines to build a well-structured tool and provide a deeper understanding of requirements to build a context-aware system.

There are several benefits of context-aware systems in human-centered computational environments. Researchers have talked about how context awareness adds to the system's capabilities to become more human-centered. “The ultimate objective of human-centered computing is that it will serve the benefit of users (acting as individuals or in teams) by empowering them, by improving their experience, by making them more productive and creative, and by integrating social and technical dimensions.” (Fischer, 2012) Another study mentions the ability of context-aware systems to provide adaptivity for computational systems in nature. It explains that adaptive systems are those that change themselves to better suit the needs of the user, while adaptable systems are those that allow the user to change them to better suit their needs. It argues that adaptable systems are better because they allow the user to keep control and because they are more likely to be compatible with other systems (Colman et al., 2014). The benefits of these systems have the potential to be realized for educational purposes and provide possibilities for a learner-centered context-aware system.

2.2.2 The role of context-awareness in Ubiquitous Computing

The term ubiquitous computing refers to computing in context. Ubiquitous computing is the next step in the evolution of technology, where computers are embedded into our daily interactions and environments (Lyytinen & Yoo, 2002). It consists of a large number of autonomous agents working together to build a smart and interactive physical environment [31]. Agents need to sense and reason about the current context of the environment for efficient functioning.

In recent years, there has been a shift towards creating a positive user experience, which can improve overall perception and satisfaction with using computer systems, particularly for tasks that require creativity or leisure (Schmidt, 2003). Using conventional interfaces, such as desktop computers with screen, keyboard, mouse, and speakers limits the design space for creating an experience; however, including the real environment as part of interaction for a computer system offers many interesting possibilities.

The role of context has gained great importance in the field of ubiquitous computing. “Context” is any information about the circumstances, objects, or conditions by which a user is surrounded that is considered relevant to the interaction between the user and the ubiquitous computing environment (Dey et al., 2001). Many applications have been built in ubiquitous computing environments that are context-aware and can adapt to different situations and be more receptive to users’ needs (Shafer et al., 2001) (Hong et al., 2021). Schmidt et al. describe the concept of Ubiquitous Computing and the importance of context in interaction with computers. They highlight that users' expectations and anticipations of a system's reaction are dependent on the situation and environment, as well as prior experience. They mention that the real world plays a crucial role in shaping these expectations as interaction with physical objects is experienced from a young age (Schmidt, 2003). Such contextual systems with ubiquitous computing principles can help create delightful experiences and enhance the possibilities of designing natural interactions.

Recent researchers are analyzing the requirements to build contextually aware environments. Dasgupta et al. (2020) talk about the importance of image recognition and speech recognition techniques in capturing context and analyzing environments. Real-time object detection is a computationally intensive process and would be a resource-heavy task on current MR devices as they already run at their full capacity to handle virtual object rendering. To solve these concerns, a few related works have used techniques to distribute computation for various applications such as object recognition in the workplace, dynamic identification of IoT objects within proximity, and complex manufacturing training processes (Dasgupta et al.,2020). AK Dey suggested a simplified design process to build context-aware applications. The process consists of three steps: (1) Specification, which involves specifying the problem, context-aware behaviors, and the required context, (2) Acquisition, which involves determining and installing the hardware or sensors to provide the required context, and (3) Action, which involves choosing and performing the context-aware behavior. The author refers to this system as a “Context toolkit” and used it to develop and analyze several applications (Dey, 2000).

Overall, these findings contribute to the larger body of work on context awareness and its role in ubiquitous computing. It also provides a strong foundation for future studies to build upon.

2.3 Learning theories to support curiosity and improve knowledge recall

2.3.1 Curiosity and curiosity-based learning

In the academic literature, there is a lack of consensus on the definition of curiosity. Some authors define it simply as the desire for knowledge, while others offer more nuanced definitions that consider multiple dimensions of the concept (Maw & Maw, 1961). Fitzgerald (1999) describes curiosity as the urge to investigate, to discover. In another definition, curiosity has been defined as a desire for information in the absence of extrinsic reward and has long been recognized as a crucial motivation driving educational attainment (Markey & Loewenstein, 2014).

Several research studies show that curiosity can have benefits on learning. Litman and Klahr (2012) show that curiosity leads to exploration and learning. Another study focusing on children shows that children who are more curious have higher teacher ratings of competence, motivation, attention, and persistence, and they ask more questions (Jirout & Klahr, 2012). Research on the brain and memory indicates that individuals who are more curious tend to have enhanced memory retention for information related to their curiosity, as well as improved memory recall for unrelated information encountered while they were in a state of curiosity.

These effects persist over a prolonged period. Researchers have also explored various techniques to promote curious behavior such as exploring an object's functions or using questions to solve tasks (Gruber et al., 2014). This body of research holds significant implications for educational practices, highlighting the crucial role of curiosity in the learning process.

2.3.2 Concreteness fading

Concreteness fading is a teaching technique where a given concept is introduced in decreasing levels of concreteness (Bruner et al., 1966). The first form is the physical or concrete model of the concept; the second one is a graphical model and the last is the abstract model. Suh et al. (2020) have presented techniques to design interfaces and interaction methods to support this learning technique. They talk about different types of concreteness representations namely in an embodiment, concept complexity, concept complexity, perceptual richness, and information.

Fyfe, E. R. & Nathan (2019) mentions the theory of concreteness fading as a promising approach to instructional design that aims to support learning and transfer by creating connections between multiple representations. This theory outlines a three-step progression from concrete to abstract representations of a concept, which is intended to facilitate initial learning and support transfer learning by starting with concrete experiences and gradually building up to more abstract representations, learners are better able to make connections between different contexts and apply their knowledge in new and unfamiliar situations. In this paper, the theoretical framework of concreteness fading is expanded by defining relevant terms, outlining the practical solutions it offers, and generating six testable hypotheses to motivate research and inform implementation by practitioners. The goal is to design and test instructional techniques that foster connections during learning, with the potential to contribute to an optimized version of concreteness fading that informs learning environments. Their work sheds light on the nature of learning and transfer and makes the concreteness fading theory more robust.

Kokkonen et al (2021) analyze the use of concreteness fading as a general instructional approach in mathematics, physics, biology, and chemistry. The authors find that the approach is more straightforward in mathematics, where the concrete representations used can be selected and constructed freely to support the understanding of the target concept. However, in physics, biology, and chemistry, concrete representations are specific instantiations of a particular phenomenon or principle and may only become meaningful to students after some prior knowledge has been acquired. The authors conclude that the use of concreteness fading in different domains is not analogous to each other and that different levels of representation in biology and chemistry provide complementary information, making the approach less generalizable. They suggest that these differences should be considered when attempting to use concreteness fading as an instructional approach.

This teaching strategy can be an effective way to promote deep learning and understanding in students. The approach involves gradually reducing the level of structure and support provided by the teacher, allowing students to internalize and apply their learning in increasingly complex and authentic situations. While concreteness-fading may require a more student-centered approach and a shift in the teacher's role, the results by Fyfe, E. R. & Nathan (2019) and Kokkonen et al (2021) suggest that this method can lead to greater student engagement and motivation, as well as improved learning outcomes.

2.3.3 Situated learning

Situated learning is a technique wherein students are placed in environments where they are immersed in an activity while using critical thinking skills (Anderson et al., 1996). Researchers have analyzed and applied these techniques to various concepts.

Several research studies show the importance and benefits of situated learning to improve knowledge recall. Miller and Gildea (1987) found that words are learned much more effectively when they are taught in the context of ordinary communication, rather than from dictionary definitions and exemplary sentences. This is because words are not islands, entirely unto themselves, and the context of an utterance provides extralinguistic help that is necessary to resolve ambiguity, polysemy, nuance, metaphor, and so forth. Another study mentions that there is substantial importance of activity in learning, and how activity provides an experience that is necessary for subsequent action. The essay argues that activity also produces representations that are lexicalized, meaning that they are dependent on context. The essay concludes by suggesting that knowledge similarly indexes the situation in which it arises and is used (Brown et al., 1989).

The body of research discussed in this section suggests that situated learning can lead to deeper, more meaningful, and authentic learning experiences, as well as increased engagement and motivation. Overall, the literature highlights the potential benefits of adopting a situated learning approach in educational contexts.

2.3.4 Constructivist learning theory

Researchers have analyzed the importance of constructivism in learning. (Bada et al., 2015) mention constructivist learning theory as a theory that emphasizes the role of the learner in constructing their own understanding and knowledge of the world. They mention that an important limitation of education is that knowledge cannot simply be transmitted from professors to students; rather, knowledge must be actively created in the minds of pupils. VonGlaserfeld (1989) identifies the learner as an active agent in the process of acquiring knowledge in this constructivist theory of learning (Bada et al., 2015).

Constructivism actually stimulates and taps into students' natural curiosity about the world and how things work. Instead of trying to invent the wheel from scratch, students try to comprehend how it works. They become involved by putting their prior knowledge and experience into practice, developing their ability to posit hypotheses, putting their theories to the test, and ultimately drawing conclusions from their findings (Bada et al., 2015). In another study, Fernando et al. (2017) investigate the constructivist teaching/learning theory and participatory teaching methods through a survey of 41 undergraduate students. The three claims of constructivist teaching/learning theory that learning is active, influenced by students' prior ideas, and socially and culturally rooted were supported by the student's responses. The survey also showed that participatory teaching methods, particularly question and answer, and group discussion, are popular and effective in improving the learning experience. The paper concludes by advocating for a combination of traditional lectures with participatory methods to enhance the learning experience, balancing subjectivity and objectivity in the classroom (Fernando et al., 2017).

In conclusion, constructivist learning theory emphasizes the importance of active, student-centered learning where individuals construct their own understanding through interaction and reflection, rather than simply receiving information from a teacher.

2.4 Summary

The literature suggests that mixed reality (MR) technologies, including augmented reality (AR) and virtual reality (VR), offer a range of benefits for enhancing the learning process across various subjects. However, limited attention has been given to MR tools that aid in the creation of educational content across various subjects, particularly in the context of a head-mounted display (HMD)-based MR, which presents a possibility for a more engaging and immersive experience with the learning material.

In the literature discussed, theoretical frameworks have been developed to provide a better understanding of context-aware systems, such as the context life cycle and the various layers of context-aware systems. There are several benefits of context-aware systems in human-centered computational environments, including improved user experience, increased productivity, and creativity. These systems could be beneficial for education and offer opportunities for a learner-focused, context-aware experience.

Overall, the literature suggests that different learning techniques such as Curiosity-based learning, Situated learning, Concreteness Fading, and Constructivist learning theory have the potential to lead to deeper, more meaningful learning experiences, increased engagement and motivation, and improved learning outcomes.

3. Related Works

This chapter introduces the existing literature on mixed reality (MR) technology in education and context-aware agents in fields including education and training. The first section reviews the applications of MR technology in education and compares them based on the Mixed Reality continuum, learning content metrics, and pedagogy frameworks. This section provides a better understanding of how these applications work and helps situate the prototypes explored in the later chapters within the context of the research questions. The second section discusses the literature on context-aware frameworks and their requirements and how they are applied in the context of education and training. This section provides insight into the importance of context-aware agents in various fields and their potential impact on education and training.

3.1 MR Education Applications

In this section, I will explore the existing literature on the use of mixed reality (MR) technology in education. This section reviews various aspects of the applications which are related to the prototypes explored in the later chapters. I have summarized what these applications are used for, and how they work and make a comparison table based on the Mixed Reality continuum (Milgram, 1994), learning content metrics, and pedagogy frameworks. Using these parameters to analyze related works helped me better situate my prototypes in the context of the research questions.

3.1.1 Applications

a) Serendipitous Language Learning in Mixed Reality (Vazquez et al., 2017)

This research work talks about the benefits of situated learning outside the classroom where the student is in control of the experience through exploration. They present a prototype for language learning called “WordSense: Vocabulary learning in the wild”. The user wearing a Hololens is able to look at the objects around them and find vocabulary help in the contextual affinity of the object. For eg., a user looks at the cup and finds options to see example sentences, hear the pronunciation, etc.

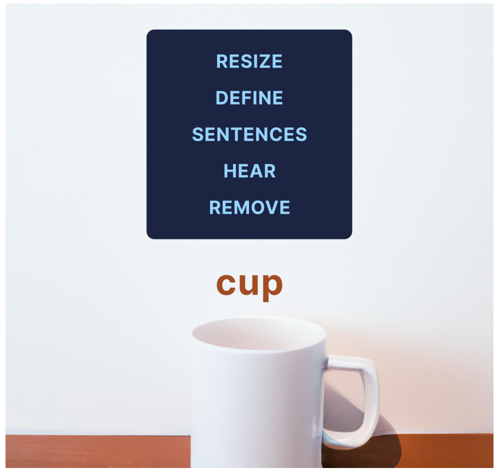

Figure 3: Vocabulary help guide contextual affinity built for Serendipitous Language Learning in Mixed Reality. Recreated image for reference (Vazquez et al., 2017)

b) NeverMind: Using Augmented Reality for Memorization (Rosello et al., 2016)

The prototype built in this research work uses the technique of memory palace. The hypothesis is that using a virtual memory palace with AR can improve memorization. They introduce the concept of the memory palace method, "Given a concept or word you want to memorize, you come up with a visual mental symbol for this concept that will help you remember it by association. Then, take this imaginary image and mentally place it in an architectural scene. Finally, to recall your concept, you imagine the scene you mentally created and the concept you want to remember will effortlessly emerge."

They ran experiments to measure recall accuracy for memorizing Superbowl champions with 14 participants. They compared their technique with the Paper-based learning task and the results indicated positive and recall accuracy had significant improvements when measured a day and a week after the participants learned it.

c) Enabling the Use of Real-World Objects to Improve Learning (Niemann et al., 2010)

This paper presents the use of RWOs (real-world objects) in the MACE (Multimedia Annotation and Corpus Engineering) system, which connects related learning materials used in architecture education. RWOs allow learners to explore new paths while using the MACE portal. The paper describes how RWOs are automatically generated and how they are used in MACE. The authors plan to further research the identification of RWOs and to include more repositories in MACE.

3.1.2 Table comparison of Related works

The above-related works were compared based on different parameters in a table (Table 1). Firstly, whether the prototype is mobile-based or HMD-based, or neither. Secondly, the multimodality of the prototype was marked with tactile, auditory, and visual support for learning content. Thirdly, the prototypes were marked for eye-gaze or hand interactions support. Fourthly, the relation of learning content with the environment context was compared. The table also describes the nature of learning content instruction, if it's predetermined or is dynamically based on certain factors such as the user's surroundings or interests. Finally, it also describes the learning pedagogy frameworks used.

Related Work	Mobile-based / HMD-based	Multimodality	Eye gaze / Hand interactions	Relation of learning content with Environment Context	Predetermined / Dynamic Instruction	Learning pedagogy frameworks
Serendipitous Language Learning in Mixed Reality	✅ Hololens 2 Understands objects around the user and uses contextual affinity to display learning content.	✅ Tactile ✅ Auditory ✅ Visual	❌ Eye gaze ❌ Hand interactions	Relates real-world object to language learning content	Dynamic based on the objects in the surrounding	Constructivist learning theory Situated Learning
NeverMind: Using Augmented Reality for Memorization	✅Mobile-based Understands objects around the user and uses contextual affinity to display learning content	✅ Tactile ❌ Auditory ✅ Visual	❌Eye gaze ❌Hand interactions	Relates contextual virtual images to learning content	Predetermined content	Situated Learning Memory palace
Enabling the Use of Real World Objects to Improve Learning	❌ None	✅ Tactile ❌ Auditory ✅ Visual	❌ Eye Gaze ❌ Hand interactions	Relates real-world architectural buildings to architecture subject content	Dynamic content based on user's queries	Learning by examples

Table 1: Comparison of Related Works based on various parameters

Analysis of these prototypes inspired me to direct aspects of the prototypes which I built later. Firstly, these related works use several pedagogies and I decided to focus on different learning frameworks in each of my prototypes. Secondly, none of these works used gaze or hand interactions to incorporate contextual learning, I wanted to explore the use of hand interactions. Finally, these applications have used different techniques to provide contextually-relevant learning content to the learners, which inspired me to experiment with different approaches during the interactive prototyping.

3.2 Context-aware agent framework for learning systems

Context-aware agents have become increasingly important in various fields, including industrial and training domains. In this section, I discuss literature about different requirements to build context-aware frameworks and how they are applied in the context of education and training.

3.2.1 Frameworks

Researchers have built frameworks to support context-aware applications. In one paper, Ashok (2020) suggests three categories of requirements and conditions for designing such applications. Firstly, the user interaction should be dynamically created based on the structure of the task rather than hard-coded. Secondly, the services used by the application should be specified abstractly, and the system should dynamically compose, adapt, and load application components based on available resources. Finally, the application should be able to adapt to the resources available and handle task context transfer seamlessly between environments. In certain cases, users should also be given control over multiple choices.

Researchers have also built frameworks to support specific use cases. Yaghmaie (2011) proposed a system framework for improving the learning process quality based on multi-agent concepts, where agents with properties such as autonomy, pre-activity, pro-activity, and co-operability are used to tailor content to learners' needs. Four types of agents are used in the system: context management, content selector, content organizer, and content presenter. These agents retrieve the current learner on the model, decide on appropriate course topics and content packages, and deliver them to the LMS. The use of agents in learning adaptation has become more common and they simulate the human side of learning more naturally than other computer-based methods. Dasgupta et al. (2020) propose a mixed reality (MR) system to create context-aware spaces that can be applied in industrial and training domains. They suggest that their system can be used in industrial assembly operations to improve the assembly process and reduce human error, and can also reduce the user's cognitive load. Their system uses machine learning-based object detection in real-time to assist with object manipulation tasks and generates audio-visual cues for the user to follow.

4. Methodology

My research focuses on building a framework to support contextual educational lessons in MR. The goal of the framework is to create engaging learning experiences. Such a framework could help improve curiosity support for learners and help improve the ability to recall concepts learned. I call this tool “Curiosity XR”.

4.1 Overview

I used a Mixed-methods approach which focuses on the Research-through-design (Zimmerman et al., 2010) and the Interaction Design process approach (Sharp et al., 2019). During the process, I worked on defining, ideating, designing, building, and evaluating five prototypes. This iterative prototyping approach (Figure 4) is adapted from the interaction design process by Sharp et al. (2019). Helen et al. 2010 describe an efficient approach to designing prototypes for HCI (Human-computer interaction) and mention a life cycle model. Using an adapted form of this approach, my iterative prototyping method involved these steps for each of the five prototypes: Defining use cases and pain points, ideating storyboards, building prototypes, analyzing and reflecting on them, and proceeding to the next prototype. Finally, I built a final prototype “Curiosity XR” as a learning system framework that could be used for creating and experiencing in-context mixed reality learning content.

4.2 Process

This project also involves collecting both quantitative and qualitative data such that combining two approaches provides a more complete understanding of the research problem than either approach alone (Creswell 2014). In this work, qualitative approaches include participants' user study responses while quantitative approaches include analyzing the MR prototypes and data collected during the experience testing.

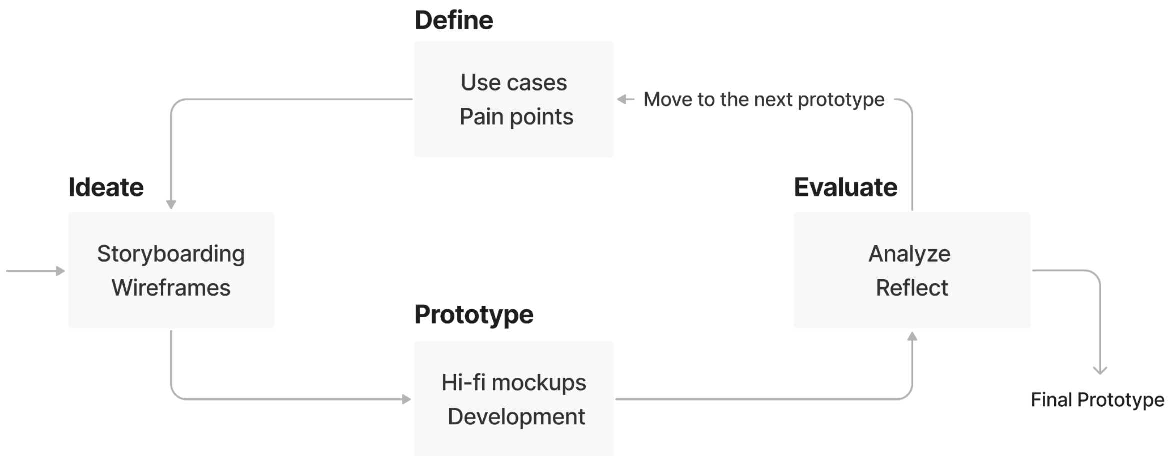

Figure 4: Outline of the adapted Interaction design process for the iterative prototypes (Sharp et al., 2019)

In the sections of the next chapter, I present the entire process in detail for each of the five prototypes (Figure 6) by explaining the ideation, design, development, system architecture, and reflection of the iterative prototypes.

The five prototypes were designed to explore the concept of my thesis in parts. For Prototype 1, I explored a tool for language learning through the real-world objects around the user wearing Vuzix Blade smart glasses (Vuzix Corporation, 2018). Prototype 2 tested the capabilities of MR as visualization support through paper sketches and doodles using an MR HMD; Hololens 2 (Microsoft, 2019). Prototype 3 was designed to explore the concept of providing real-time dynamic learning content to the user wearing Hololens 2 as well through a Voice User Interface (VUI). Prototype 4 focuses on a tool to provide automated learning content anchored to real-world objects through Hololens 2. Prototype 5 was built to support template-based lesson creation to support educators to create contextual MR interactive learning content using Hololens 2.

SQ1: Acquiring context from mixed reality devices to support educational lessons

SQ2: Using acquired implicit and explicit context to present educational concept to the learner

SQ3: Building a curiosity-support agent through Mixed Reality

SQ4: Building Systems for educators/(adaptive-context aware systems) to create/generate environment-based multimodal mixed-reality lessons

SQ5: Using MR multimodal content to support interactive learning experiences in context

SQ6: Evaluating such learning experiences to support the learner's curiosity and retention of knowledge

Figure 5: Research sub-questions for reference for Figure 6

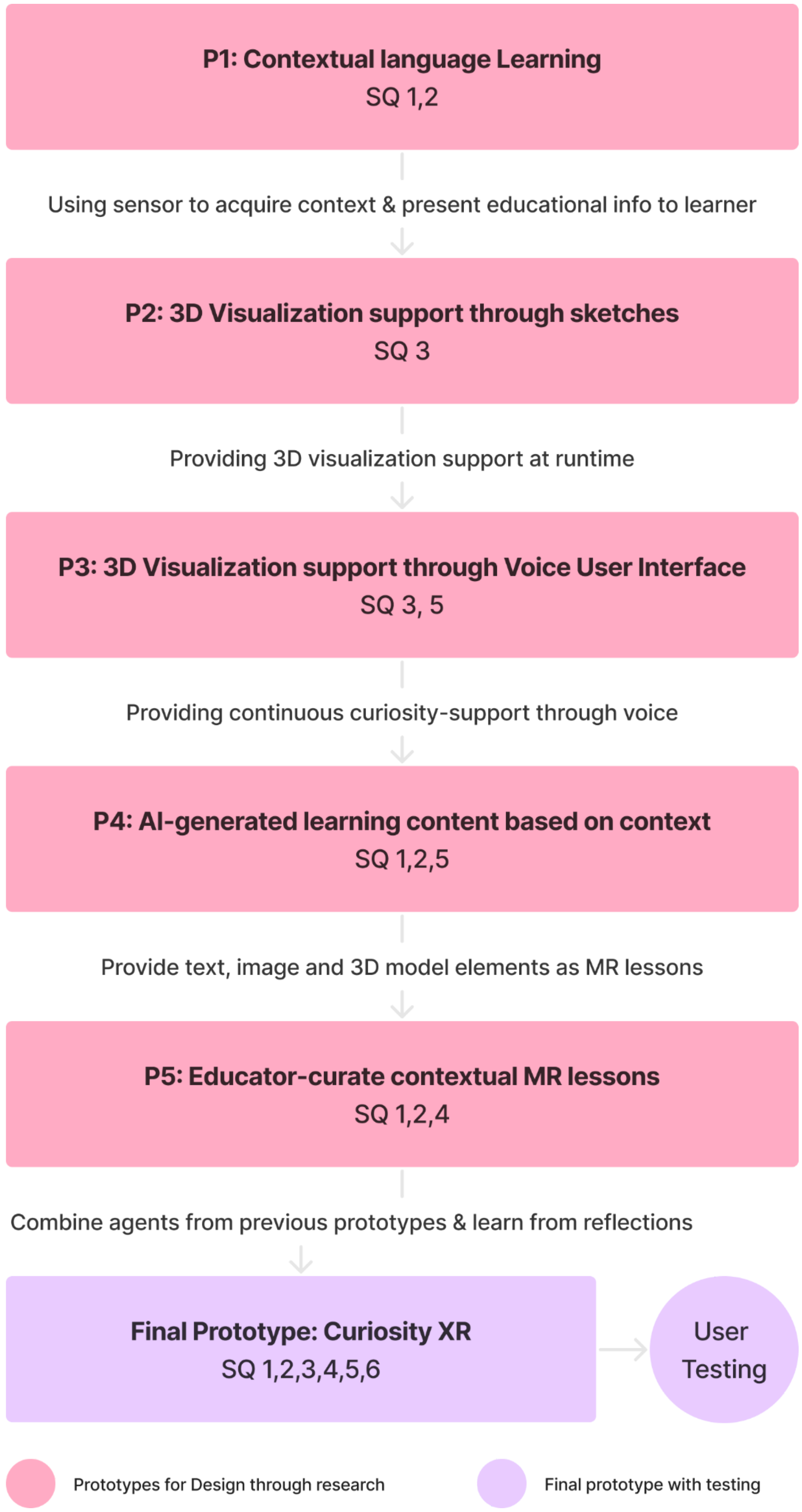

Figure 6: Methodology flow for this thesis (SQ indicating the Sub-questions relevant to the prototype, see Figure 5 for reference)

The final prototype "Curiosity XR" builds on top of the components of the previous prototypes and utilizes user context to support learning in a single MR application. Curiosity XR was earlier built on Hololens 2 (Microsoft, 2019) and later transitioned to target Quest 2 (Meta, 2019) and Quest Pro (Meta, 2022).

To evaluate the first five prototypes, various parameters were analyzed with reference to the literature and compared through a radar plot visualization (Figure 7) based on the following parameters in relation to the literature. This evaluation is based on my subjective understanding of the literature. The final prototype is then tested through a user study that involves participants. The design knowledge is then generated in the form of prototypes, its analysis, and the participant study findings for the final prototype.

4.3 Design criteria for measuring the prototypes

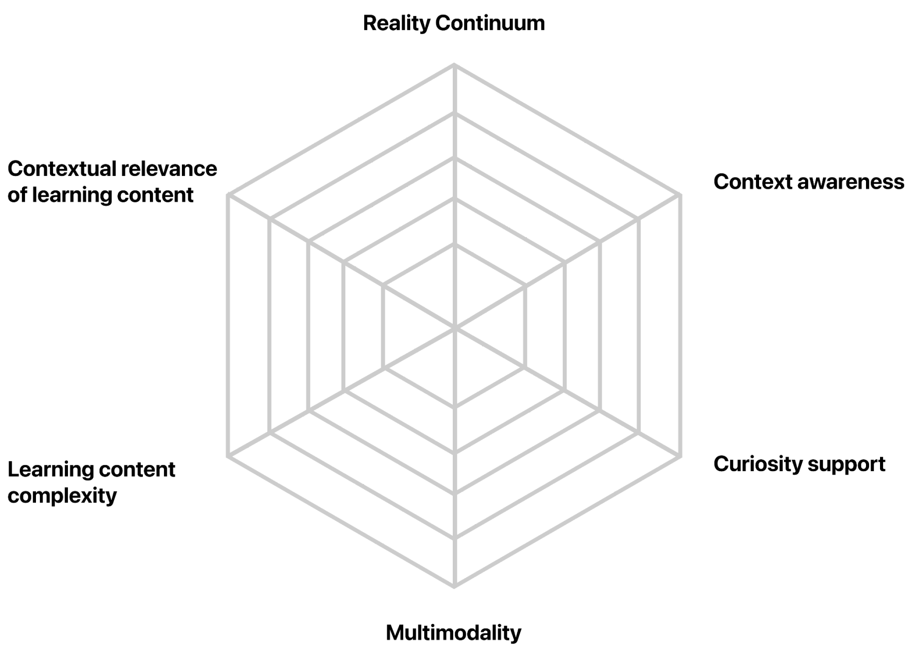

Figure 7: Radar plot visualization template for evaluating prototypes

Reality Continuum: Scale of 1-5. Based on the reality-continuum scale (Figure 2) (Milgrim 1994). 1 indicates the Real environment and 5 indicates the Virtual environment.

Context awareness: Scale of 1-5. This scale is based on the ability of the system to analyze the environmental context and user context. 1 indicates the system has no context awareness, while 5 indicates that the system is completely aware of the object instances in the environment, the 3D spatial position of each object, the user's speech, the user's actions and intents, the user's history with past interactions with the prototype, and preferences for learning.

Curiosity support: Scale of 1-5. This scale is based on the ability of the system to support the learner's curiosity. 1 indicates the system isn't able to support the learner's curiosity in any form. 5 indicates the system is able to support curiosity in any form MR AI-assisted technology can assist the learner including answering questions, understanding user intent, and providing visual and auditory support when required.

Learning content complexity: Scale of 1-5. 1 indicates simple plain static learning context, while 5 indicates dynamic multimedia learning content including text, speech, video, image, and 3D models.

Contextual relevance of learning content: Scale of 1-5. 1 indicates low relevance of the learning content, i.e. not based on the environment, user, or context resulting in low-contextual relevance which might affect the effectiveness of the learning content. While 5 indicates, total relevancy with the context, where the system provides learning content highly correlated to the user context, and environment context.

Multimodality: Scale of 1-5. This scale is based on the capability of the system to provide multimodal input and output. 1 indicates supporting one mode of input/output. (Visual, audio, tactile), whereas 5 indicates supporting multiple modes providing seamless integration of multiple inputs to enhance the learning experience.

5. Iterative Prototyping

In this chapter, we will explore five iterative prototypes that aim to enhance learning and visualization support through MR. Each prototype focuses on a specific aspect of learning, ranging from language learning to real-time content generation, and incorporates different tools such as Augmented Reality (AR), 3D visualization, Speech, and AI-generated content. Each prototype will be introduced with its motivation, followed by a description of the process used to develop it. The outcomes and architecture of each prototype will also be discussed, along with reflections on their effectiveness and potential for future development. Overall, this chapter aims to showcase the potential of technology to enhance the learning experience in various ways, providing learners with personalized, engaging, and effective tools for acquiring knowledge and skills.

5.1 Prototype 1: Language learning with AR through Context

To start with the research, the first prototype was aimed to explore the first and the second sub-question (SQ1, SQ2), "How can a Context-aware educational system acquire context from mixed reality devices to support educational lessons?" and "How can acquired context be used to present an educational concept for the learner?". Exploring these questions through the first prototype provided me with a better understanding of how sensors could be used to acquire context and further present educational information to the learner.

5.1.1 Personal Motivation

When I arrived in Canada in 2021, I used various methods in an attempt to acquire proficiency in the French language. These approaches included utilizing popular language-learning applications like Duolingo, completing targeted exercises to strengthen my knowledge of French grammar, and immersing myself in French cinema. While these techniques undoubtedly facilitated my acquisition of certain grammatical structures and vocabulary, I found that they lack a certain level of practical application when attempting to comprehend real-world language usage. For example, as children, we learn the name of an object such as a bicycle through direct observation and parental instruction which is different from how the above-mentioned approaches work. After a bit of research reading about the theory of immersion, I realized that immersion can significantly accelerate the learning process and I was curious to build an app that could make learning languages fun and highly efficient for me and all the other language learners.

5.1.2 Process

With an inspiration to make language learning fun and efficient for certain concepts, I started researching user personas for language learners (Figure 8).

For this prototype, I wanted to target language learners who have started out as language exchange travelers or students. Building a tool for advanced concepts needed a deeper understanding of different languages and learning instructional design for teaching languages. A few language learners who are beginners travel to other countries to immerse and look for as much immersion as possible to make the most of their limited time travel. While some people are able to travel to the language-speaking country of the language they want to learn, It is not feasible for most people to travel for immersion, and in such cases, technology can help provide virtual immersion.

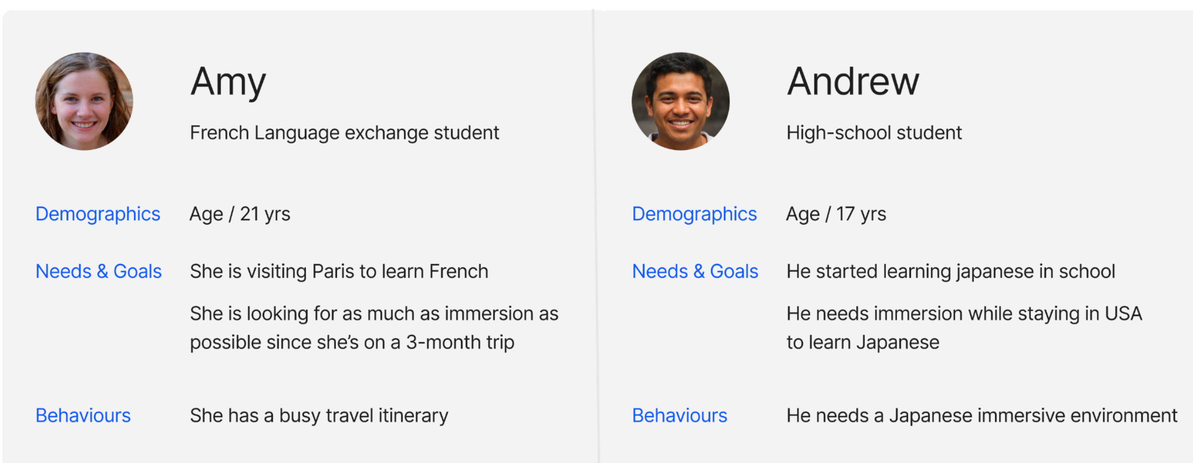

Figure 8: User personas to represent the target audience for Prototype 1

Most language-learning apps introduce new languages by introducing users to simple nouns such as a dog, cup, etc. with images that aren't presented in the context when they experience such objects in the real world. Also, language learning apps such as Duolingo rely support on pedagogical techniques for spaced repetition where a concept is introduced again after some time to result in longer retention for vocabulary, but it doesn't regard the learner's context which is their environment where things and situations do repeat and could be used to provide language lessons to the user. Immersion shouldn't require learners to travel to other countries and should have a natural interaction mechanism where the technology is hidden, supporting learners to grasp vocabulary and grammar effortlessly similar to first-language acquisition.

Krashen (2017) talks about comprehensible input in his paper and the theory of immersion. He talks about providing contextual information with hearing and reading comprehensible input is enough and most effective when acquiring a second language.

The concept behind this prototype was to explore language learning with smart glasses which would have major benefits over Mobile-based AR in terms of immersion, usability, and interactivity through ubiquitous design techniques (Abowd et al., 1999). Using mobile-based AR limits learners to have constant immersion while smart glasses allow being immersed in the learning content while being actively present in the real environment.

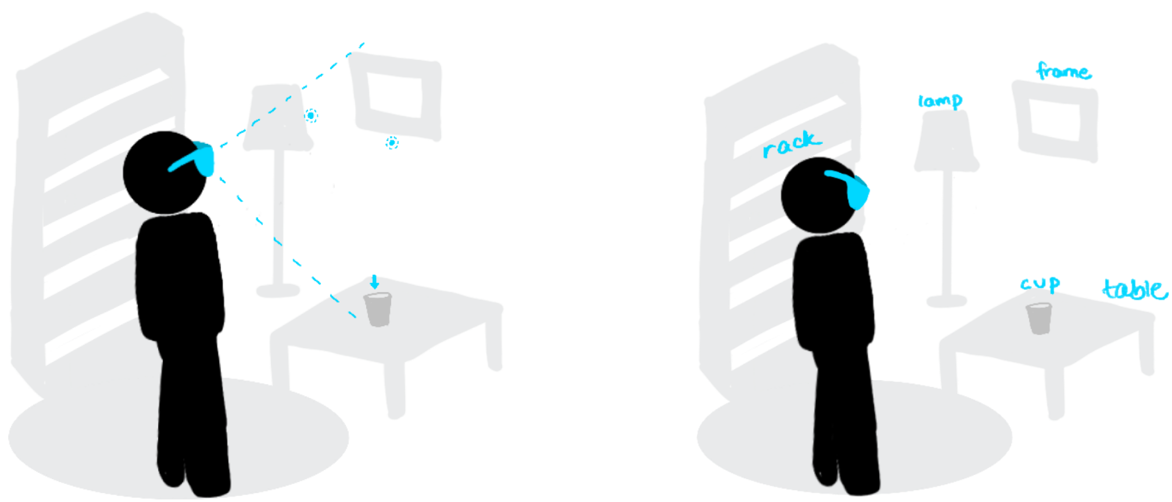

Figure 9: Sketches to visualize multiple object recognition

I brainstormed several scenarios through sketches to visualize the capabilities of the basic context-aware system (Figure 9) and its various use cases. After that, I also designed high-fidelity mockups to understand interactions that could help users learn vocabulary or grammar through smart glasses (Figures 10 & 11).

Figure 10: Sketches to visualize use cases with the user wearing smart glasses

Figure 11: High-fidelity mockups for learning nouns, adjectives, and grammar using environment context

5.1.3 Outcomes

After designing some sketches and high-fidelity mockups, I had to choose an HMD for this project. With the various available HMDs such as Oculus Quest 2, Hololens 2, and other smartglasses, I chose the Vuzix blade smart glasses (Vuzix Corporation, 2018) as they are ideal for daily outdoor use, and have a compact form factor that made them a suitable choice for the target audience and my familiarity with Android OS on which the Vuzix blades run on. Steve Mann[1] was kind enough to lend them to me for this project. Vuzix Smartglass are android smartglasses with a camera, a microphone, a speaker, and a heads-up display.

The app built (Figure 12) for this prototype uses image recognition to recognize objects in the user's surroundings and is followed by translation to translate objects near the user. When users wearing smart glasses look at any object around them, they see what that object is called in the target language on the HUD(heads-up display) screen. It is able to support multiple languages and allows the users to switch between languages by tapping on the capacitive touch panel on the glasses. The smart glasses displayed the translation of the object's name and spoke out the pronunciation as well.

Figure 12: First-person-view screenshots of the language learning application

The glasses also have in-built speakers and could also speak out the pronunciation of the [1] target object for multimodal feedback. The app was selected to be presented at VRTO 2022 Expo[2] (Figure 13). Around 50 people tried it during the 2-day period. People could see it as a potential tool for language learning and helpful when traveling to foreign countries.

[1] Mann, S. (1996). Wearable computing. IEEE Technology and Society Magazine, 15(3), 12-23.
[2] VRTO 2022. (2022). About VRTO. Retrieved from https://conference.virtualreality.to/about/

Figure 13: Prototype 1 demonstration using Vuzix blades (Vuzix Corporation, 2018) at the VRTO 2022 Expo, OCAD University, Toronto.

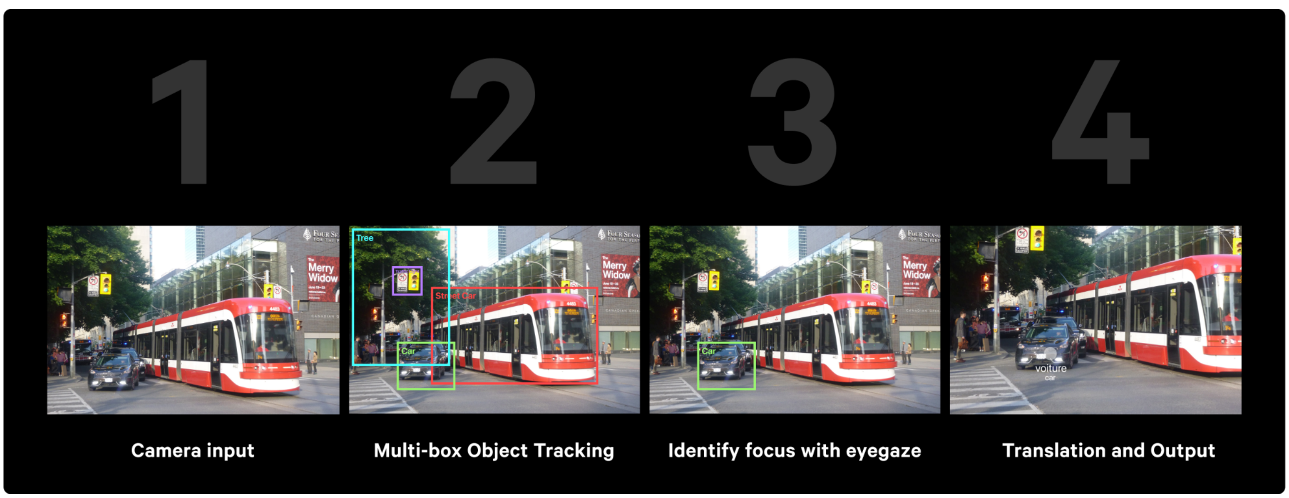

Figure 14: Visual representation of the working of Prototype 1

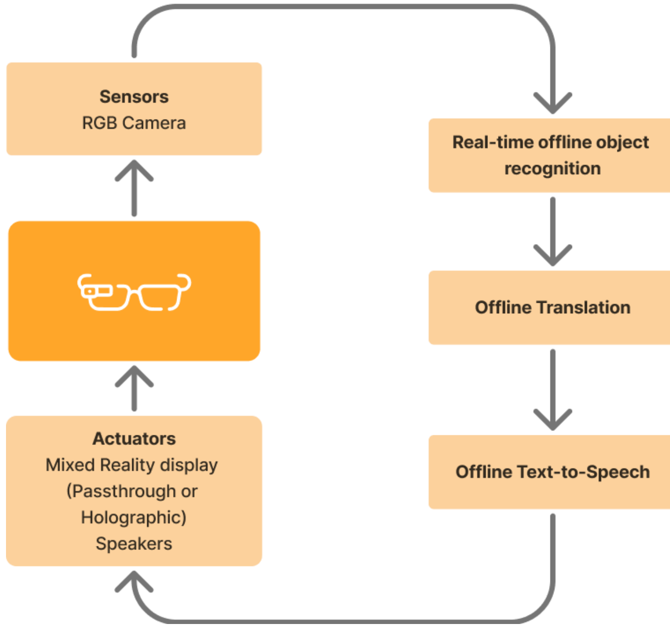

Figure 15: System architecture diagram for Prototype 1 (See Appendix E for details)

5.1.4 Reflection

The prototype was built for daily wearable smartglasses with a small display augmented onto the user's vision. The user interface simply had a text for the translated object and hence had a minimum information overlay on top of the real world. According to the Reality–Virtuality Continuum (Milgram et al., 1994), this work lies closer to the point of AR in the spectrum and has less of the MR component to it.

The system was able to recognize the objects in the user's environment and had a set user preference for the language the users wanted to learn. Although it was limited to identifying 80 objects, they were common enough to help a user learn from their environment. Being able to comprehend other parameters in the surroundings could have helped in providing variety in the educational information. Apart from recognizing the object categories; colors, texture, pattern, and scene descriptions can be a great addition to the contextual awareness system. Providing real-time translations and pronunciation guidance can help language learners feel more confident and motivated to explore and learn new words and phrases. They would also better understand and remember new vocabulary, which can spark their curiosity and encourage them to learn more.

This prototype incorporates parts of different learning theories. The key interaction, which is triggered by the user's gaze, initiates the lesson content using a curiosity-based intent as the user looks at an object and the system displays the translated word. The educational content delivered relates to the user's surroundings, exemplifying the concept of situated learning.

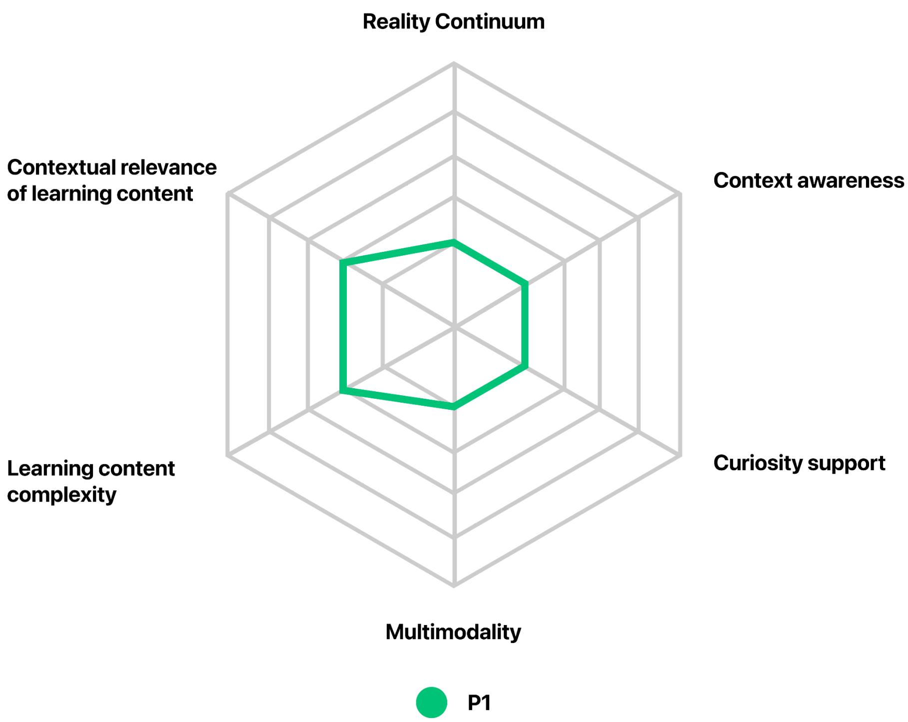

Figure 16: Radar plot visualization for Prototype 1

5.2 Prototype 2: 3D visualization support through paper sketches

Exploring SQ1, and SQ2 through the first prototype provided me a better understanding of how sensors could be used to acquire context and further present educational information to the learner. My experience with the earlier prototype revealed the importance of exploring advanced headsets like 6DoF displays for spatial learning content. Offline object detection and translation are limited, and exploring other models is necessary for versatility. A 2D display may not provide sufficient information for users. Identifying users' desired objects for translation was challenging due to the absence of eye-tracking sensors in Vuzix smart glasses. Reflecting upon the previous prototype, I realized the necessity to continue exploring the research questions with a headset with advanced display capabilities.

This prototype was aimed to explore the third sub-question (SQ3), “How can we build a curiosity-support agent through Mixed Reality?”. The prototype explores a technique that uses Mixed Reality to provide visualization support for learners.

5.2.1 Personal Motivation

When I was a kid, I remember watching a television show, Shaka Laka Boom Boom[3], in which the lead character; Sanju, had a magic pencil that brought his sketches to real life. I always wanted to get that magic pencil however as I grew up I realized that it was nothing more than fiction.

Looking at that television show from a different perspective while working on my thesis, I realized the importance of connecting sketches and art on paper with 3D objects. The concept of turning sketches into objects such as a chair, car, etc. would help learners visualize various visual perspectives and this idea could also be extended to drawing objects which don’t exist like purple apples or magical worlds of mushrooms with abstract gradients as the background. The interaction techniques for this expression in 3D spaces could be through 2D sketches and could provide learners with a natural way to imagine and create 3D spaces/objects. This technique could support curiosity by not limiting the learner’s imagination to the paper.

5.2.2 Process

The aim of this prototype is to understand the benefits and scope of a Mixed Reality visualization agent initiated through sketching on paper. As for the device to be used for this prototype, I decided to use the Hololens 2 because of its advanced display capabilities as opposed to the earlier prototype’s Vuzix blade. Hololens 2 allowed applications to anchor 3D models spatially while the users could move and also use hand interactions to manipulate (scale, rotate, move) 2D/3D holograms in space. The other alternatives were using a Quest 2 / Quest Pro which used a Pass-through AR display (showing a projection of the real world through cameras) whereas the Hololens 2 uses an Optical AR display (direct perception of the real world as we see with our eyes). There was a major drawback in using the Quest platform which led me to choose the Hololens 2. Quest didn’t allow the developers to access the raw

[3] Shaka Laka Boom Boom (2000). Retrieved from https://www.imdb.com/title/tt10510908/

cameras due to its privacy policies, and hence it was not possible to use the Quest cameras to detect characteristics of a sketch from a paper.

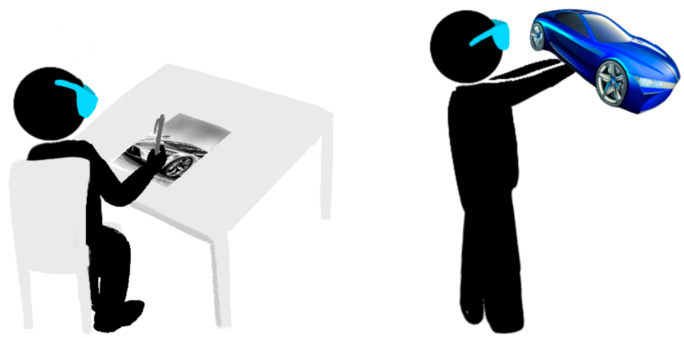

Figure 17: Paper-based sketch visualization support with MR

To begin with the storyboarding, I brainstormed use cases of recognizing user-drawn objects and augmenting them in the MR space and drew some use case sketches to visualize this tool (Figure 17). Users could draw rough sketches on paper but also could be using a whiteboard or a blackboard to draw concepts. They might want to visualize objects, concepts, or spaces (Figure 18). Users could want to visualize what a 3D model of a car concept looks like. They might want to visualize a UI design through a low-fidelity prototype they've drawn. The visualization support could also be beneficial to architects when designing building concepts on paper or through a rough sketch. These and many such use cases, however, have a much broader scope, and before proceeding to build these concepts I had to understand the requirements and technical constraints. For the first step, I looked into the different kinds of information which could be extracted from sketches on paper. I started exploring different datasets which could be used to train ML models for inferring objects or other characteristics from sketches.

Figure 18: Marker-board-based visualization support with MR

One of the interesting datasets which I came across was the largest doodling data set; the Google Quickdraw dataset (Ha & Eck, 2017) has about 350 categories (car, apple, chair, etc.) and 10,000 images for each category. Whereas, the Google Quickdraw dataset allowed for an ability to recognize user-drawn objects on paper. Other datasets which I found relevant were the BendSketch (Li et al., 2017) which allowed the modeling of freeform surfaces through 2D sketching, and OpenSketch (Gryanditskaya et al., 2019); a richly annotated dataset of product design sketches. Most of the other generally available datasets used camera-captured footage of the real world to identify objects and were not suitable for context recognition from paper sketches.

Among the ones mentioned above, the Quickdraw dataset was a better choice to implement and build a prototype considering the time constraints. I decided to work on a prototype that could recognize objects using the Quickdraw dataset and display the 3D model of that object to the user. For eg., the user draws a chair and sees a chair in front of them. Once the context (objects, colors, textures, situation, etc. in the sketch is acquired, it would be used to choose relevant 3D models and display them in the user's surroundings through the MR display.

To enable the generation of 3D models, I looked into quite a few 3D generation techniques; text-to-3D machine learning-trained models such as Magic3D (Lin et al., 2022) and Dreamfusion (Poole et al., 2022) which accept user text queries to generate 3D models based on the query and similar techniques on the rise but they were not feasible to generate models in real-time. There were a few other drawbacks associated with this approach as well. Firstly, the sketch would have had to be transformed into a text query using techniques similar to visual question answering which was a bit cumbersome to work on for an initial prototype. Also, due to the display limitations of the Hololens 2, it is difficult to convert the generated 3D models into a usable and supported format for Hololens 2 on the go.

Another technique that I looked into for generating 3D models was using 3D model databases available for non-commercial uses. I came across the largest 3D model database; SketchFab API (Sketchfab, 2022) which had most of the basic objects available which could be fetched in real-time and displayed on the Hololens 2. I decided to use Sketchfab's API to fetch 3D models for this prototype.

5.2.3 Outcomes

The prototype (Figure 19) was built on Hololens 2 (Microsoft, 2019), with real-time object detection from a sketch to generate 3D models and place them in the user's vision. The model is trained on the Google Quickdraw dataset and works with around 350 common objects (chairs, cars, apples, etc.).

Figure 19: Screenshot from the Prototype 2 experience: Drawing a sketch of an apple loads a 3D model of an apple

The 3D models are fetched from the SketchFab API. This allows for real-time searching for models such as an "apple", filtering and finding a suitable model format and specifications that are supported by Hololens 2, downloading it, and then rendering it in front of the user. The prototype also implements MRTK3 (Microsoft, 2022) to support hand interactions with the 3D models. Implementing this tool allowed the user to use their hands to interact to scale, rotate or move these holographic objects in space.

Figure 20 demonstrates the process of recognizing an object from a sketch to showing it in MR. Figure 21 demonstrates the system architecture diagram for this prototype. It shows how information is transmitted from the headset to the python server and back to the headset using several APIs to support this prototype.

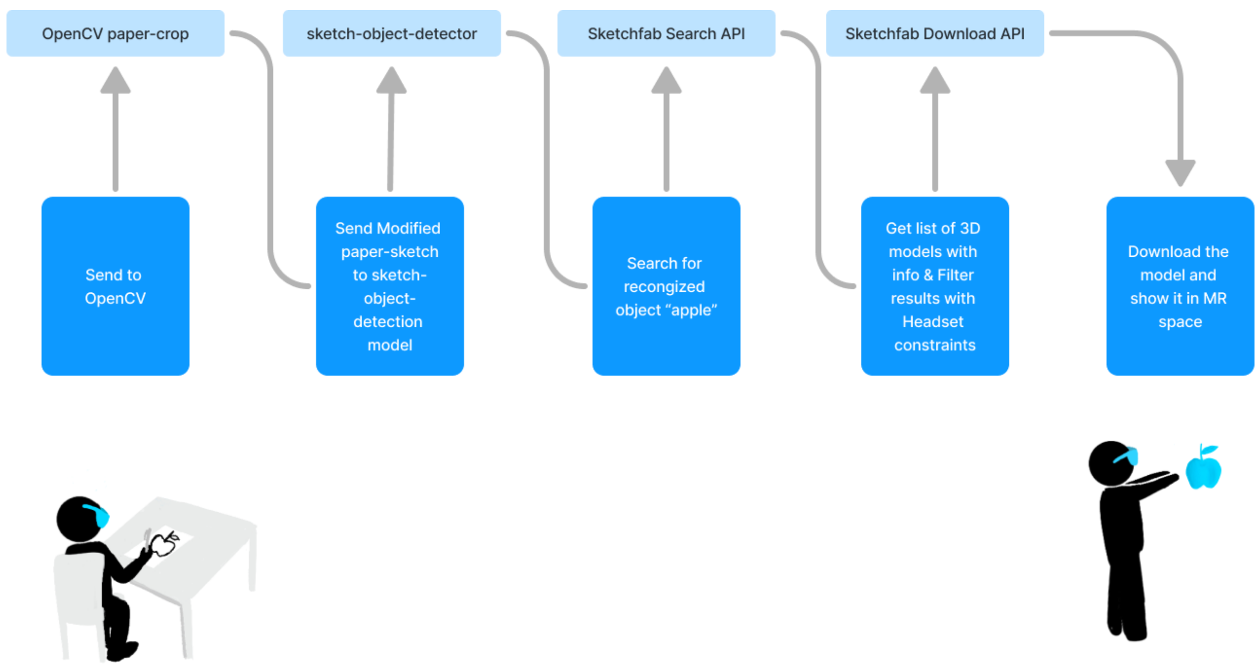

Figure 20: Flow diagram to demonstrate the functionality of Prototype 2, Sketch-to-3D visualization

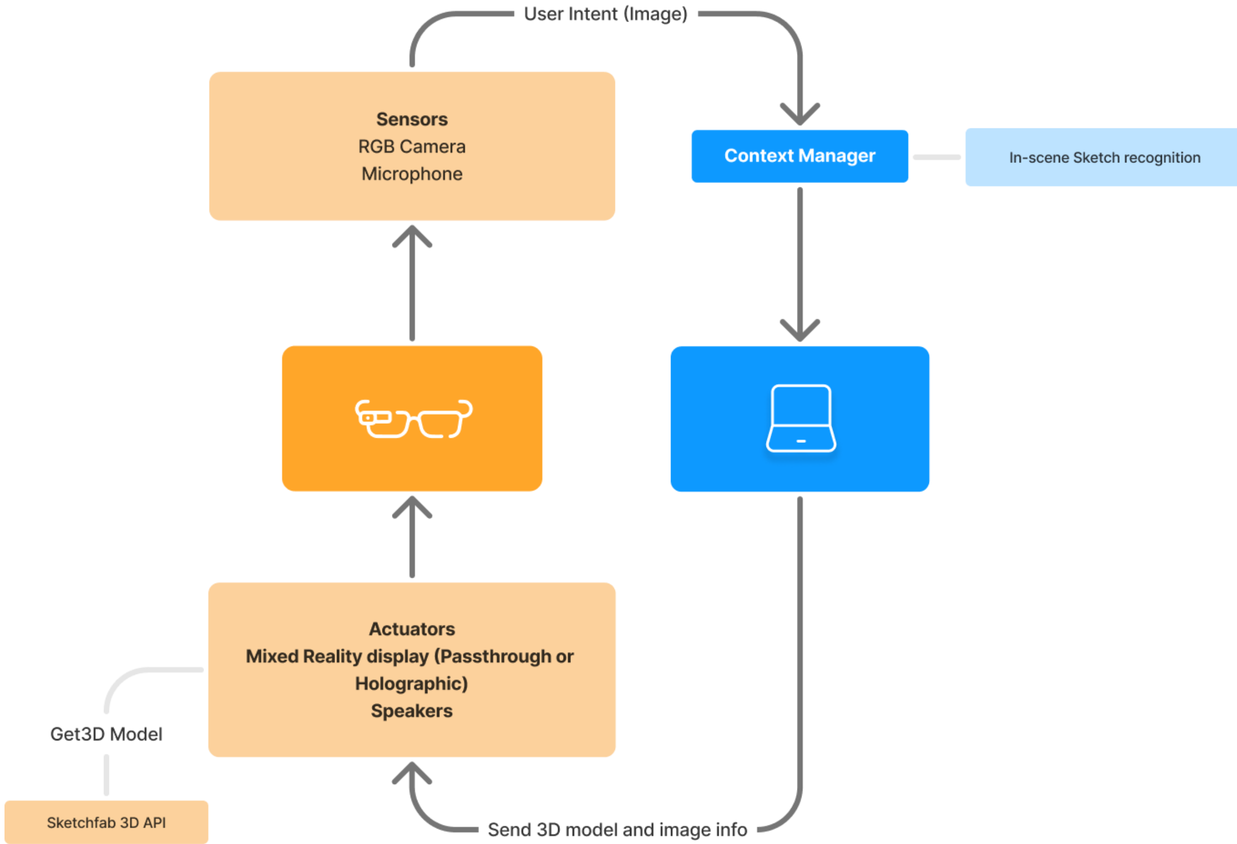

Figure 21: System architecture diagram for Prototype 2 (See Appendix E for details)

5.2.4 Reflection

This prototype would lie close to the central part of the Reality-Virtuality continuum (Milgram et al., 1994). Compared to the previous prototype, it has more virtual components than fixed text such as spatially-anchored and interactive 3D fully-colored models augmented in the user's environment. I think that allowing users to manipulate the virtual objects gives users the ability to be able to fully immerse into the models (Right extremum of the Reality-Virtuality continuum) and also scale it down to have minimal augmentation in the real world. (Left extremum of the Reality-Virtuality continuum).

The context recognition subsystem for this prototype had various limitations. It was able to understand the bounding box of the paper through the camera stream but it wasn't able to recognize if the paper had multiple sketches or a single sketch. Also, if the camera stream had multiple pages in the environment It chose a paper at random. This subsystem helped understand which object has the user drawn assuming one object was drawn by the user on one page with no other visible pages to the camera. Also, the camera on Hololens couldn't capture the entire sketch most of the time, since it had a placement much above the user's eyes and had a narrower field of view. The images had cropped parts of the sketches or sometimes completely missed if the user is quite close to what they are drawing. Lastly, the sketch-object-recognition module recognized incorrect objects occasionally. Different colors of pens, pencils, markers, and lighting situations altered its performance. Being able to retrieve other parameters from the sketch could have helped in generating more relevant 3D models. Apart from recognizing object categories, the ability to derive spatial properties from the sketch, colors, and textures could have helped generate 3D models with specific materials and shaders.

The goal of this prototype was to understand and explore the limitations of a Mixed Reality agent to support users with visualization through sketches. This prototype doesn't directly relate to providing educational content but the generated 3D models can be considered as a visual aid for improving the spatial understanding capabilities of learners. It allows them to visualize objects they're curious about from different perspectives and acts as a visualization support tool for ideation.

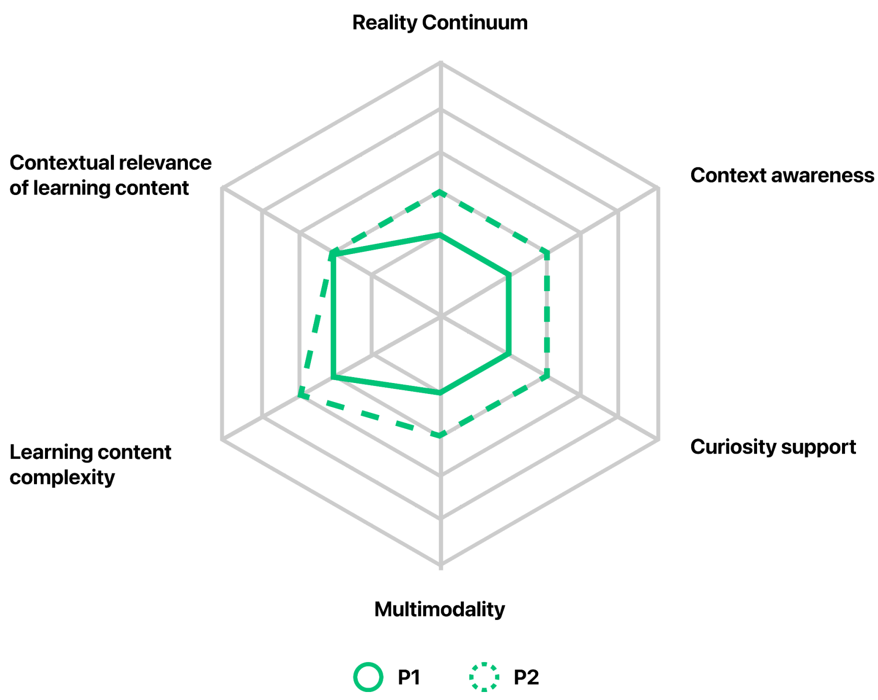

Figure 22: Radar plot visualization for Prototype 2 with the previous prototypes

5.3 Prototype 3: Real-time lesson content generation using Speech

Exploring SQ3 through the previous prototype, an agent to provide 3D visualization support at runtime was built and helpful but I realized the need for an agent which could continuously support learners' curiosity once the 3D models appear. Working on the previous prototype, I could identify technical problems associated with acquiring context from paper sketches and look for an alternative interaction to acquire user intent. I could see a need for an interactive tool that could assist learners to help with imagination and visualize things they are curious about and also answer related queries. For eg. If a learner would like to understand how a car engine works while having a 3D model in their hand, they don't have a tool that would help them visualize and assist with related queries in Mixed Reality.

This prototype was aimed to explore the third and the gift sub-question (SQ3, SQ5), "How can we build a curiosity-support agent through Mixed Reality?" and "How can Mixed Reality use this multimodal content to support interactive learning experiences in context?". This prototype aims to solve such visualization aids through a different approach and continue supporting the user's curiosity.

5.3.1 Personal Motivation

I am a visual learner and wonder about how things work around me. I realized that there is a need for a system that can respond to questions to better understand objects around us. A great resource that helped me visualize this concept was Bret Victor's TED video on "Humane representation of thought". Bret Victor (2014) talks about how we are trapped by current technology to use only our vision and auditory sense while the tactile, kinesthetic, and spatial abilities are disregarded. He talks about a need for a dynamic medium that is responsive, computational, and connected. He mentions infusing computation with physical matter to use most of the human capabilities to interact with computers and each other. However, nanotechnology isn't quite mature yet and this is possible today, but using his ideas as a motivation I realized some of these things could be implemented using Mixed Reality.

5.3.2 Process

The aim of this prototype is to understand how to build an MR application that can have functionalities to not only help users visualize 3D models instantly but also answer relevant questions while they interact with them as if they had a virtual teacher with them. To map out the use case scenarios, I drew some use case sketches (Figure 23) to imagine this tool's possibilities.

Figure 23: Sketches to visualize use cases for Prototype 3. (Top) Visualizing a model and having a query about the model as a whole, (Middle) Visualizing multiple models and having a query about one model, (Bottom) Visualizing a model and having a query about a specific part of the model.

In the first scenario, the user might want to visualize a model and have a query about the model as a whole. This could involve viewing the model's overall structure, understanding its key features and components, and asking questions about how it operates and functions. In the second scenario, the user might want to visualize multiple models and have a query about one specific model. This could involve comparing and contrasting multiple models, understanding their similarities and differences, and asking questions about specific aspects of one model in particular. A third scenario could be in which the user would like to visualize a model and have a query about a specific part of the model. This could involve examining a particular aspect or feature of the model in detail, understanding how that part works and how it fits into the overall model structure, and asking questions about its function and significance.

I decided to focus on the first scenario due to the technical support and timeline limitations. The other two scenarios would require understanding user context from voice, hand recognition, and camera image to deduce the part of the models the user is asking a query about.

To build for the first scenario, the starting step was to consider different methods to initiate a 3D model in space. The user could either select a model from the UI with options or type an object's name to search for one or request a model through voice. I decided to draw storyboarding ideas for a voice user interface for initiating 3D models. This interface could be more efficient and natural while wearing the HMD compared to the other approaches.

After brainstorming and sketching ideas, some of the use case scenarios could be the following. The user could say things like "I wonder what an electric car looks like from the inside", "Can I see a model of a laptop PCB" or "I would like to see a moon? Once a relevant model is displayed, the user could then request follow-up questions.

The VUI (Voice User Interface) could provide real-time feedback through text or speech while the user is interacting with the models around them. To be able to have a natural VUI, the first requirement would be to have an intent recognizer in the context-acquisition subsystem. Intent recognizer would help the system understand the user has the intention to visualize a 3D model of Mars from queries such as "Can I see Mars", "I would like to visualize Mars", "How does Mars look like" etc. Among the various intent recognizer pre-trained models available, I found Wit.ai[4] to be quite reliable and customizable for this prototype's use case. Integrating Wit.ai would also

[4] Facebook. (2018). Wit.ai Retrieved from https://wit.ai/

allow the agent to understand if the user would like to see a model, know something else about it, or ask factual questions in relation to what they're curious to learn about.

For the next part, to generate the 3D models, this prototype used the same technique as the previous prototype (using SketchFab API). Lastly, the system needed to display the results of users' queries. For eg., Once the user sees a Mars model, they might be curious to learn about the planet's properties, or wonder "The possibility of life on mars, or how long would it take to reach mars from a rocket?".

The scope of the user's questions could be boundless and the aim of this system is to be able to answer as many kinds of queries as the user has and support the learner's curiosity. To achieve this capability, I decided to use OpenAI's GPT-3 API which supports a variety of general text-related tasks such as question-answering, summarization, translation, etc. It has immense capabilities in being able to answer users' queries which could range from philosophy to core sciences. GPT-3 had a few known drawbacks and provided misinformation for logical and computational-based results. For eg. "What is the size of a plant cell?" provided answers such as "<1cm" or "very small of the size of a grain of salt" which isn't accurate as learning content. To provide users with a response to such queries, I decided to use Wolfram Alpha's API which is known to be one of the best computational engines on the web.

5.3.3 Outcomes

The prototype was built on Hololens 2 and allowed users to talk and request 3D models by speaking out prompts such as "I would like to see mars", or "Show me earth" (Figure 24). The system used SketchFab API to get 3D models which have a collection of 500,000+ models. The users could interact with the models and ask questions like "How far is earth from mars?" or "How long is 1 Earthian day on mars?" using different APIs to request answers. The user was able to see and read the relevant answer in the text box and the system also speaks out the answer content.

Figure 24: Interacting with planets and learning about them through the Voice User interface

This prototype uses a similar flow as the previous prototype to gather user context and show a 3D model. Figure 25 depicts how models are rendered in real-time using the user's voice context.

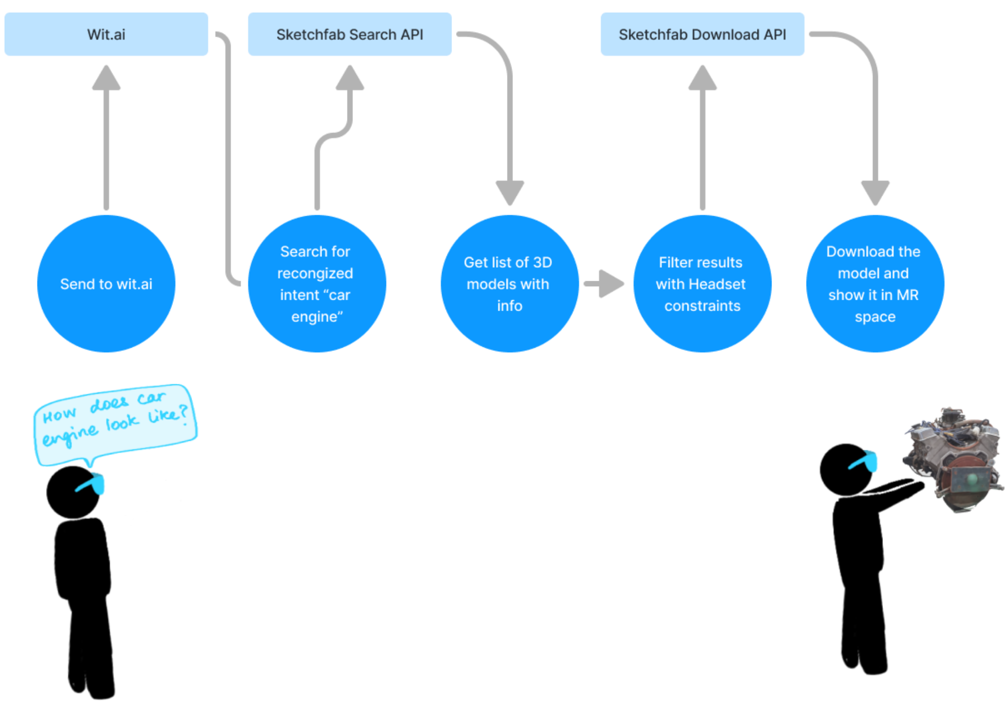

Figure 25: Process explaining real-time 3D model rendering through a Voice-user interface

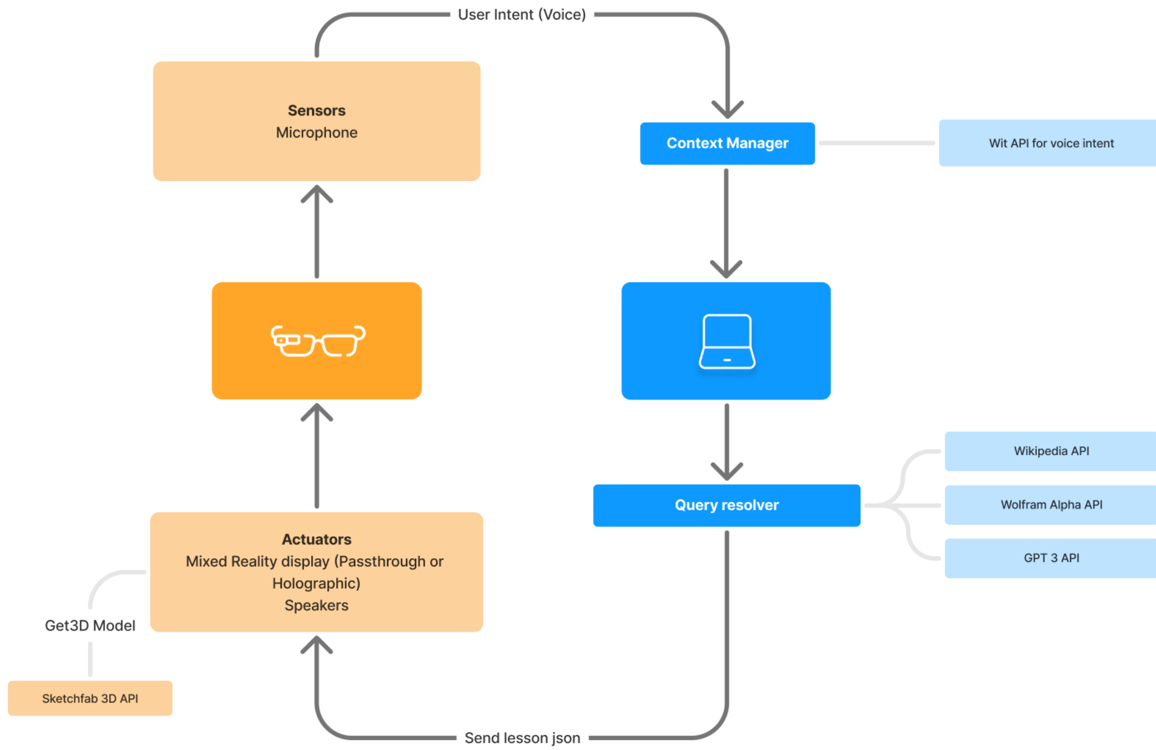

Figure 26: System Architecture diagram for Prototype 3 (See Appendix E for more details)

Figure 26 demonstrates the system architecture diagram for this prototype. It shows how information is transmitted from the headset to the python server and back to the headset using several APIs to support this prototype.

5.3.4 Reflection

This prototype would lie close to the central part of the Reality-Virtuality continuum (Milgram et al., 1994). Similar to the previous prototype, it has spatially anchored, interactive, fully-colored 3D models augmented in the user's environment. Compared to the previous prototype, the context recognition subsystem was better at identifying users' intents. Voice as a query interface was capable of more complex requests allowing users to mention colors, textures, and various other object-specific characteristics of what they'd like to visualize. One of the technical challenges of this prototype was related to the voice command initiations. There was no wake word associated with the VUI and the microphone was always active to find possible intents from the recognized speech. Due to this, it was very challenging to disregard speech detected from people nearby and converse with the people around while engaging with this prototype. The other technical challenge was to understand the context of the object which the user wanted to visualize. For e.g., If a user were to ask about the "earth" required for a plant to grow while looking at a plant cell, the system would show a model of the planet earth instead of recognizing that the user is referring to the soil.

Similar to prototype 1, this prototype incorporates parts of different learning theories. The ability to visualize and interact with objects through a VUI ties back to the principles of curiosity-based learning. The same theory also is exemplified in the learning content generation where the user could ask what they're curious to learn about and get answers. This prototype also benefits from the constructivist learning theory. During the experience, the learners actively engage and construct information through their own understanding similar to how Bada et al. (2015) describe it.

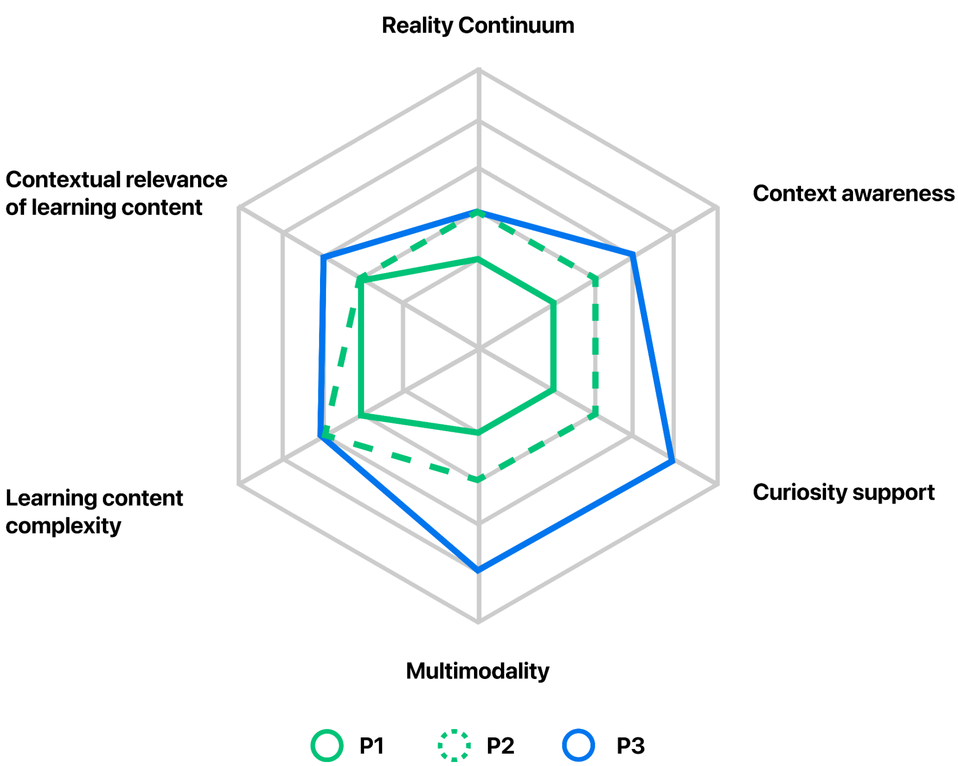

Figure 27: Radar plot visualization for Prototype 3 with the previous prototypes

5.4 Prototype 4: Serendipitous AI-generated learning content based on environment context & learner context

Exploring SQ1, SQ2, SQ3 & SQ5 through the previous prototypes helped me understand the techniques to acquire context, build curiosity-support agents, and use multimodal content to support interactive learning experiences. A part of SQ5 to support these interactive learning experiences in context was something that was yet to be explored.

This prototype was aimed to discover and analyze the missing pieces of SQ5; possibilities of interactive learning experiences incorporating the environmental context and the learner context, and, parts of SQ4; exploring how these adaptive-context aware systems create environment-based multimodal mixed-reality lessons to teach concepts to the learner.

5.4.1 Personal Motivation

In our daily lives, we often see parents teaching their kids about the things around them in their environment. For instance, a parent might show their child how to tie their shoes or explain the different types of trees in the park. This type of learning is highly effective, as it allows children to learn in a natural and engaging way, using the objects and experiences around them. However, this kind of learning is not always possible in traditional educational settings.

By integrating a contextual learning system framework into educational content, we can create a learning environment that is tailored to the user's individual needs and interests. Users could interact with real-world objects like chairs, lamps, and other household items while simultaneously learning about their history, functioning, and inventors. Additionally, such a system could make learning more accessible to a wider range of users, including those with different learning preferences and abilities.

5.4.2 Process

I made a few sketches to visualize the use cases. After brainstorming and storyboarding the possible interactions, I realized the independence of users' interests to the surrounding objects for generating educational content. For eg. If someone is curious to learn physics, they can learn physics concepts through a plant, through a table, or through a basketball. To learn physics concepts through a plant, the mixed reality system could simulate various physics experiments, such as measuring the plant's growth rate, the force required to move its leaves, or the amount of energy required for

photosynthesis (Figure 28). Mixed reality can allow users to examine the properties of the table, such as its weight, height, and center of mass. This can help users understand concepts like gravity, mass, and energy. For example, they can see how the table's center of mass affects its stability and how its weight affects the amount of force needed to move it (Figure 29). Users can also bounce a basketball and visualize the trajectory of a ball, which helps them understand the concepts of motion, velocity, acceleration, and how the force of gravity affects motion (Figure 30).

Such learning techniques could change users' perceptions of the objects and add a learning perspective to the things they experience, making them curious to learn more about the subject they're interested in.

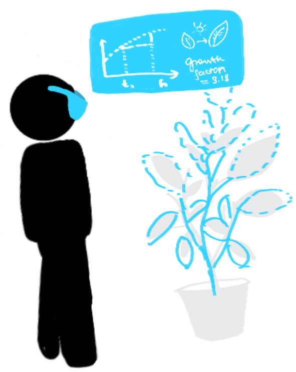

Figure 28: Learning about thermodynamics and growth rate from a plant

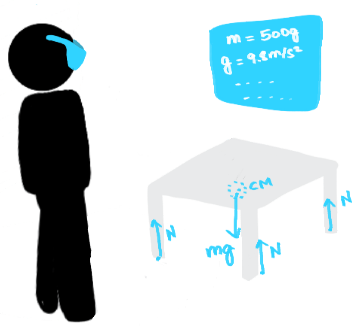

Figure 29: Learning about force and center-of-mass concepts from a table

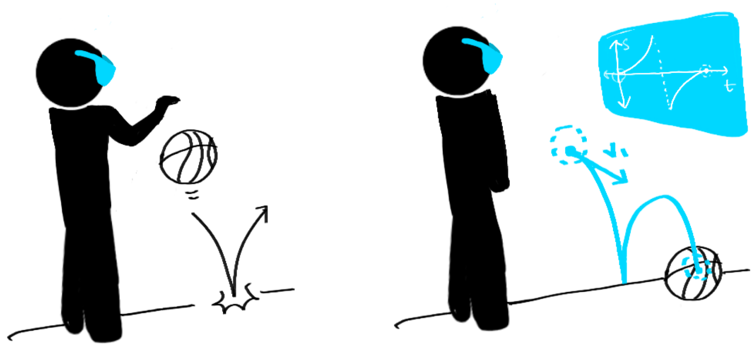

Figure 30: Learning velocity-time graphs while playing with a basketball

Also, the lesson information need not be a text similar to a single translated word in the first prototype but could also be a long text and contain images/videos/3D models.

The first part to build this prototype was to explore the interaction techniques to generate the learning content. One of the options was gaze intent; which would allow the user to gaze at objects and initiate a curiosity intent further loading the lesson content. This technique however may occlude the environment with information and constantly update as the user moves their eyes and head around. A rather modified interaction called “Gaze Pinch” solves this concern. It is available in MRTK3 and allows for intent activation based on both a user’s gaze and a pinch gesture. Another interaction that can define a user’s explicit interest would be to have a curiosity-initiation prompt anchored to objects in the user’s environment.

The next part was to understand how to spatially anchor content to real-world objects. To achieve this, the system needed objects’ spatial 3D coordinates in the user’s environment. I designed a few mockups to visualize the spatial anchor for the lesson content (Figure 31).

To accomplish this, I tried using the depth sensor on Hololens 2 but the development documentation for it is complex and requires a deeper understanding of its hardware interface libraries. As an alternative, I decided to use an external camera and a depth sensor instead. Among the various external depth sensors, Zed Mini (Stereolabs, 2018) with its Python SDK provides access to the RGB camera stream and Depth sensor stream. Zed mini served the requirements of this prototype quite well as it could be used for real-time object detection and to get the 3D coordinates of each object in the environment. Figure 32 depicts a sketch to visualize Zed mini, Hololens 2, and a Python socket server in the user space.

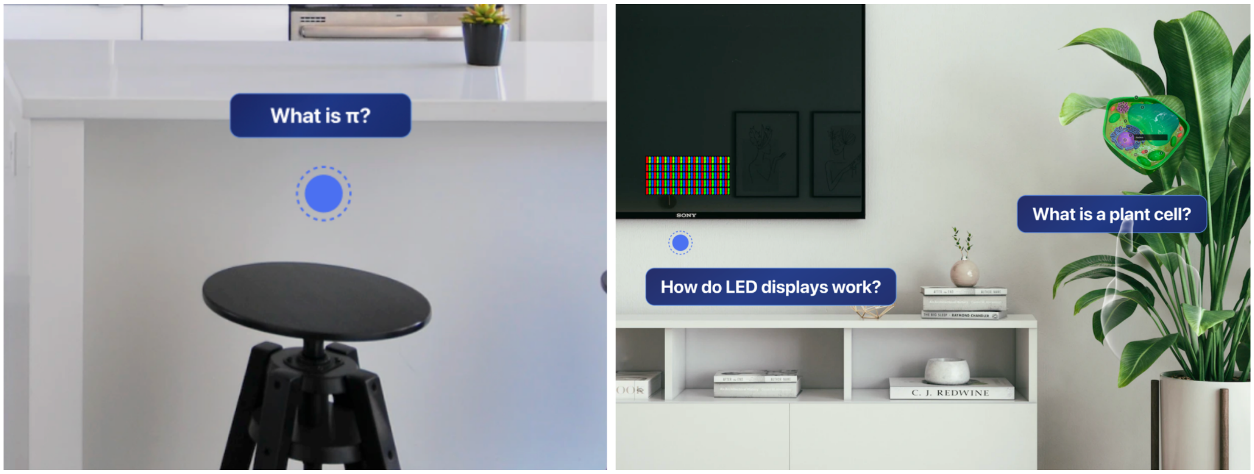

Figure 31: High-fidelity mockups using MRTK3 (Microsoft, 2022) design system to depict Object-anchored curiosity-initiation prompts

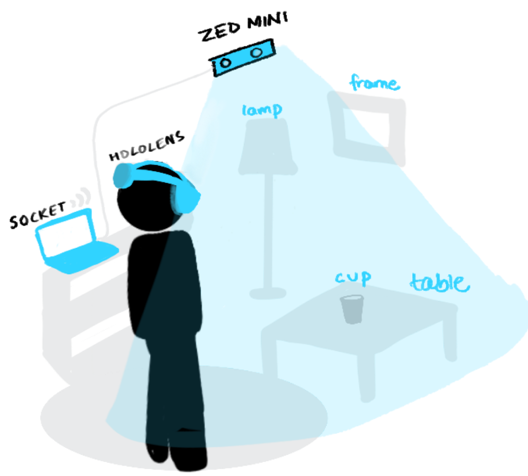

Figure 32: Low-fidelity sketch to visualize real-time object detection in 3D space using a Zed Mini, Hololens 2, and Python socket server

To support lessons such as the ones mentioned above, a context acquisition agent was required which could gather information about the objects such as their category, and other properties like color, material, and texture. And for specific cases, it should be able to recognize specifications as well. For eg. for a plant, find out the type of the plant, its common name, and its scientific name. For a building, its name if it's a monument. Such tasks are categorized as instance recognition tasks and there aren't machine learning models that currently solve this task in general. It was also challenging to acquire complex properties and behaviors of surrounding objects. For eg. tracking real-time 3D coordinates of the basketball, or mesh approximation of a plant, or table’s dimensions and weights of its different parts. Although such a complex context is possible to derive through sensors and could be used for generating highly informative and contextual learning content, but, due to the time constraints and to limit this prototype to a proof-of-concept validity requirement I decided to use an object detection model to only recognize the object category. The desired agent should then be able to identify the object and tag it as “plant”, “apple”, or “cup” etc. which could be used to generate relevant learning content.

One of the main goals for this prototype was to identify a tool capable of producing educational content based on users' interests and objects. While I considered using knowledge graphs such as Wikipedia, which associate objects with facts or concepts, I found them to be unsuitable for generating brief lessons. Instead, I found that GPT-3, which was used in a previous prototype, was a viable option. For eg., if the user is interested to learn physics, it could provide answers to queries such as “teach me a concept of *physics* using a *mug*“ or if the user is interested to learn history, it could also answer prompts such as “tell me a *historical* fact about *chair*”. And then the generated text could be directly used to display a lesson for the user. The other tool which seemed a viable solution was Wolfram Alpha’s summary box API which could provide summary cards for objects such as monuments, physical elements like metals, animals, or plants, etc. These summary cards included an image and a basic set of information but generating lesson content through these cards disregards users’ interests and rather provides general information.

5.4.3 Outcomes

With this Hololens 2 prototype, the user was able to set preferences of the topic of interest such as Physics, Maths, History, Science, etc. by simply speaking out a phrase like “I would like to learn physics”. Example use cases are shown in Figure 33. Wit.ai was used to recognize users’ intent to change their subject of interest. Once the topic of interest is set, the user can gaze at objects around them and see a text prompt of a fact or concept anchored to the object. Figure 34 depicts a flow diagram for this system architecture.

Figure 33: Screenshots from Prototype 4, the user is able to see object-anchored AI-generated learning content about different objects

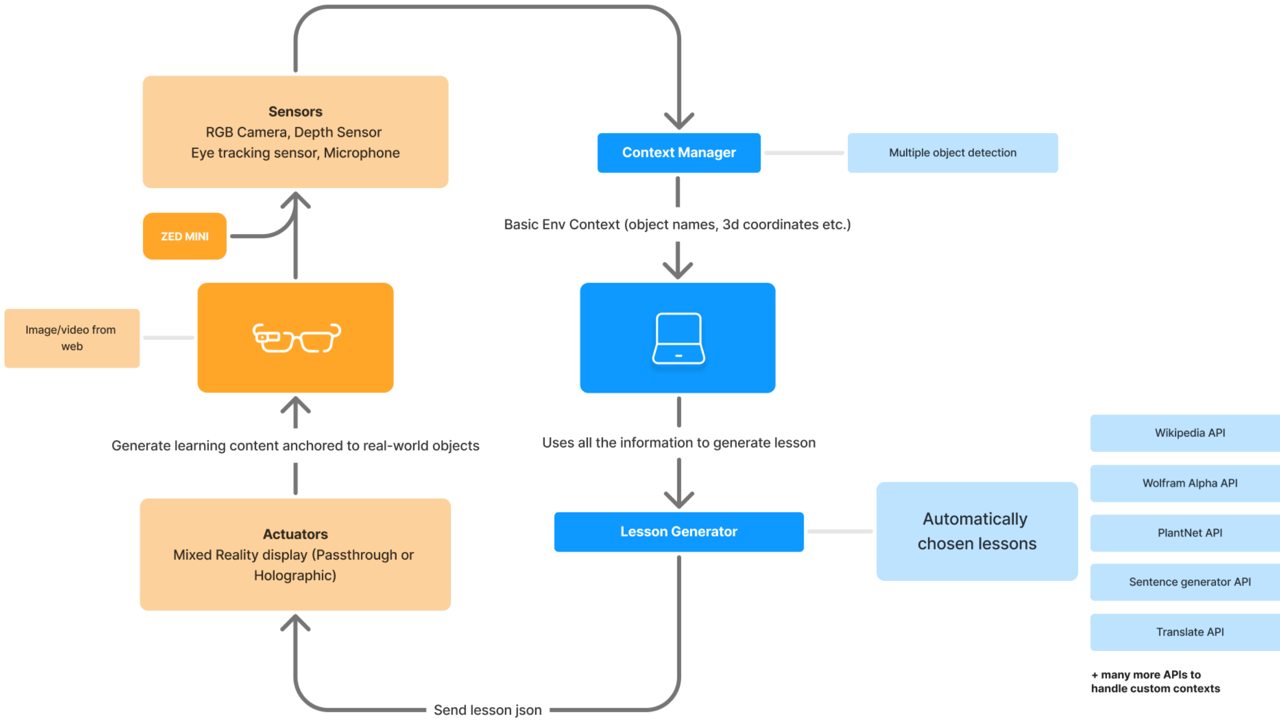

Figure 34: System architecture diagram for Prototype 4. See Appendix E for more details.

5.4.4 Reflection

This prototype would lie close to the Augmented reality part of the Reality-Virtuality continuum (Milgram et al., 1994). Similar to prototype 1, it has augmented text in the user's environment but unlike a fixed display the learning content is spatially anchored.

The context recognition subsystem was a bit different as compared to the previous prototypes. The agent used multiple sensors to be able to derive objects' spatial coordinates in real time and another agent monitored the user interest preferences.

This prototype focused on the principles of situated learning where the surrounding objects generated the learning content for the user. It also had components of curiosity-based learning which allowed the user to set their topic of preference using the VUI.

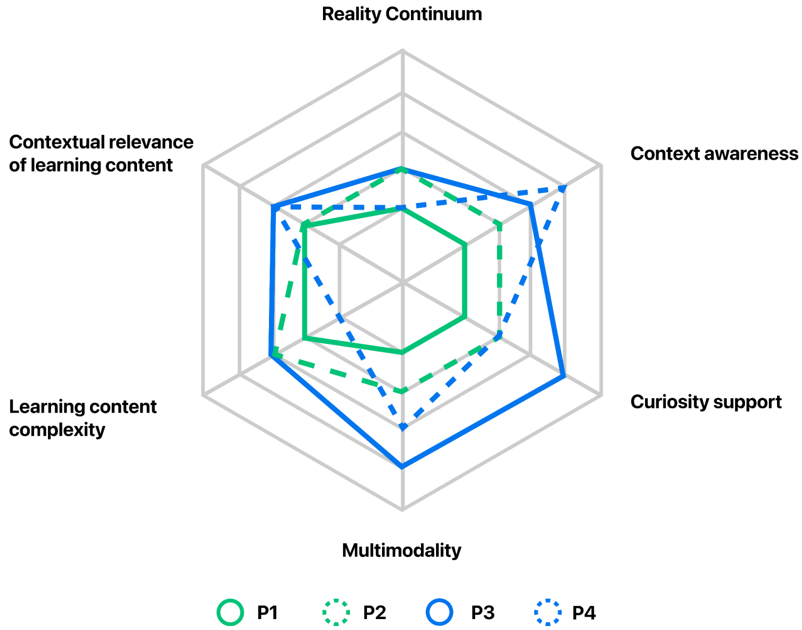

Figure 35: Radar plot visualization for comparing Prototype 4 with the previous prototypes

There were however a few challenges in the learning content generation. Firstly, Even though the answering APIs like GPT-3 allowed for open-ended queries, not all the topics were correlated to all the objects around the user. Eg. “Biology fact from a chair”, or “Geography lesson from a monitor” resulted in weird facts that didn’t make sense. The Former resulted in the output “A chair is a piece of furniture that people sit in to support their weight.” and the latter resulted in “The largest monitor in the world is the Ultra HD monitor from Samsung.“ After engaging with basic queried AI-generated learning materials, it became apparent to me that the availability of a supplementary source of carefully curated learning content is crucial in enhancing the quality of educational resources.

5.5 Prototype 5: Framework to support the creation of contextual MR mini-lessons

After exploring AI-generated learning content for MR learning experiences, I decided to work on a prototype to explore the ways to provide educators with tools for creating learning content for contextual MR lessons. This prototype focuses on exploring the remaining part of the SQ5 which was left unexplored in the previous prototype; “How can educators create environment-based multimodal mixed-reality lessons to teach concepts to the learner?”

5.5.1 Personal Motivation

In recent years, I have reflected on the impact of online video-sharing platforms such as YouTube on learning. I recognized the potential of video as a medium for sharing experiences and knowledge, which has significantly transformed traditional learning methods based on text and image-based content. However, I also realized that video content is limited to a 2D display, which restricts interactivity with the material. I acknowledged the need to offer greater flexibility to users in directing their learning content in a way that aligns with their interests. I recognized that traditional video-based learning can promote passive learning, while interactive MR environments can provide a more active, immersive, and engaging learning experience.

5.5.2 Process

To build a prototype that could support multiple domains of knowledge, I started sketching various formats of MR lesson content that could benefit learners. These lessons could vary in multiple dimensions, the dimensions would be the amount of self-directedness, the media types of learning content (text,2D, 3D), affordances, and interactability.

Building a platform that could support the creation of such dynamic aspects of these lessons was challenging and required more time and technical support. To overcome these challenges, I decided to proceed with a template-based design solution wherein the educators could choose a template that best supports the lesson requirements and they could then fill in the required details to create a lesson (Figure 37). Since this had relatively limited information, I describe it as mini-lessons. The templates would be a predetermined format of the learning content with a set of features incorporated. This would also make it easy for educators to create MR mini-lessons. The templates could incorporate a variety of elements. For e.g. One could learn so many things about plants (how much water does it need, why are leaves green, how much can it grow when full size, what is a plant cell, etc.). All these mini-lessons would need support elements like text, video, image, or 3D models which the learner could interact with while engaged in the mini-lesson.

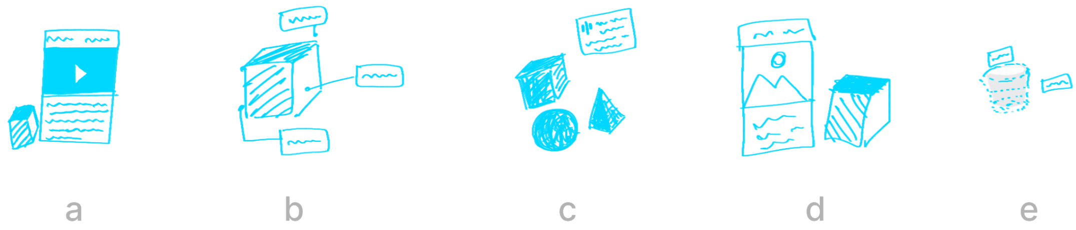

Figure 36: Various formats of MR lesson content a) Video, text, and 3D models, b) Interactive annotated 3D models, c) VUI and 3D model-based learning, d) Image, text, and 3D models, e) Real-world object augmented with interactive 3D models and annotated labels.

The educators would also need a tool to be able to create such mini-lessons easily. I designed mockups to visualize the UI for educators. They would have to select an object with which the mini-lesson is associated, apply relevant object filters (eg. Cactus if they select a plant), and select a template for the mini-lesson which would have components like title, description, models, images, etc. This would help them focus on designing the learning content and not be limited by the technical knowledge to build Mixed Reality applications.

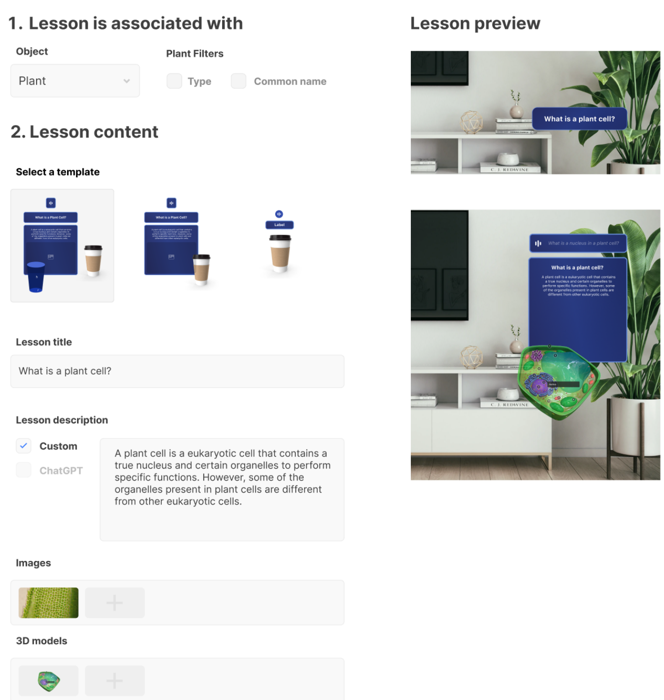

Figure 37: Mini-lesson creation web UI mockup for Educators

5.5.3 Outcomes

The Educator interface was built on React Native[5] and allows educators to create a mini-lesson by opening a form on a web/iOS/Android (Figure 39). The developed version is a simpler form of the mockup and only supports a subset of the mini-lesson configuration. In the app, an educator can select an object to associate a mini-lesson with, and enter a title, description, and model name for the mini-lesson; they can also add tags to associate this mini-lesson with different topics. This mini-lesson information is then added to a database and will be displayed to users based on the semantic match of the user's topic of interest with the mini-lesson tags for a corresponding environment object.

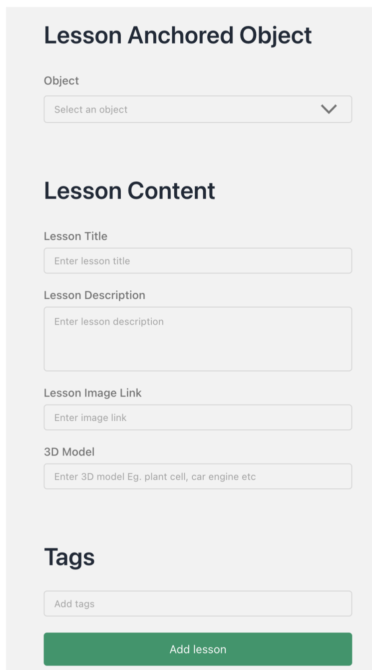

Figure 38: Educator interface to create an MR mini-lesson

As users explore the environment using this prototype, they will be prompted with curiosity prompts that are relevant to the objects in the environment. For example, a prompt such as "What is a plant cell?" would appear near a plant, while a prompt such as "How to measure the volume of this cup?" would appear when a user sees a cup (Figure 38). These works were accepted at IEEEVR 2023 in the Technical Video Track (Vaze et al., 2023).

[5] Facebook. (2015). React Native. Retrieved from https://reactnative.dev/

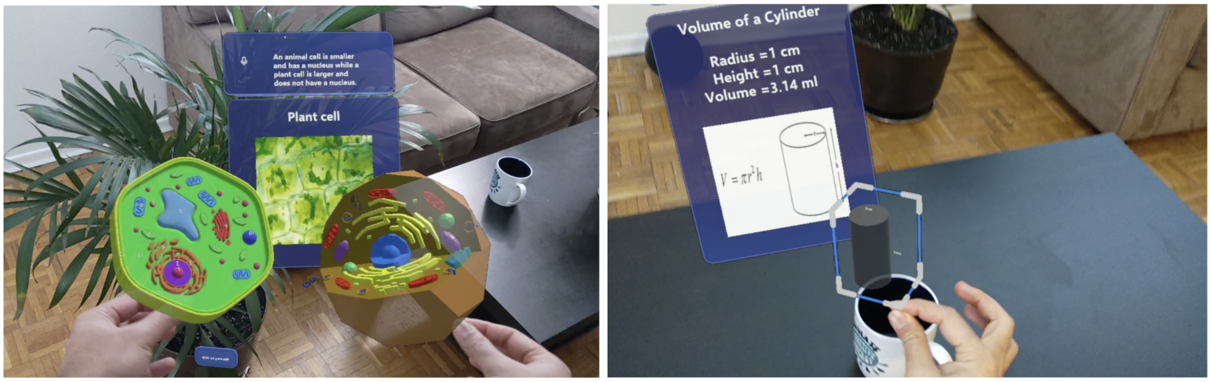

Figure 39: MR learning experiences using MRTK3 for Prototype 4 (left) Comparing an Animal cell with a Plant cell near a real plant, (right) Scaling Cylinder to learn about the volume of a cylinder near a real cup.

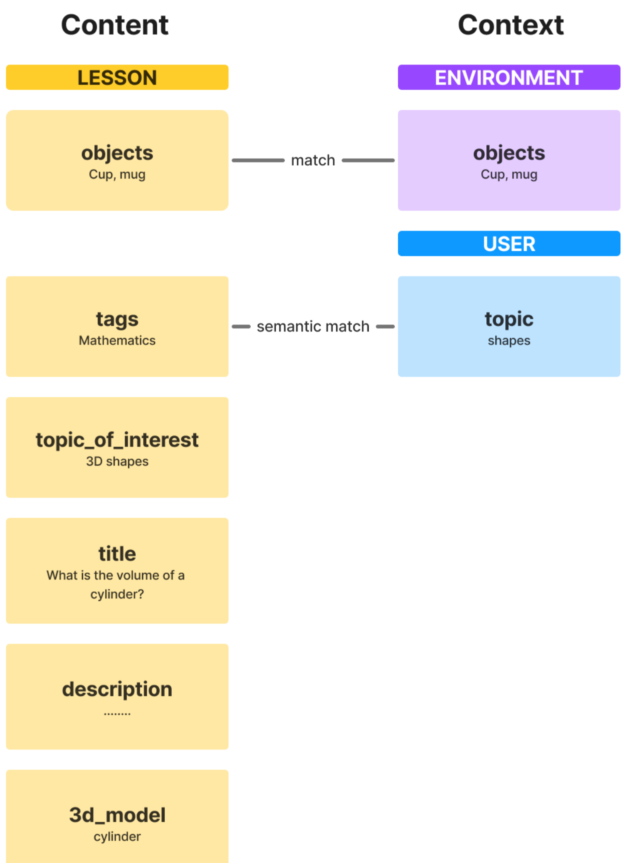

Figure 40: Content-context mapping, (Left) Represents data of a lesson, (Right) represents the Context (User, Environment). The diagram depicts a case when this lesson is chosen. This happens when “objects” in Lesson data match the “objects” in the environment and “tags” in the lesson data semantically match the “topic” of the user’s interest.

A new additional component added to this prototype (Figure 40) helps to choose a mini-lesson based on the environmental context and user context. The learner is able to set up their interests while using the interface, either through VUI or updating it in the settings. The user can select a topic of interest: “topic”. Eg. “subject: mathematics, philosophy, technology, reality, etc.” etc. Figure 41 depicts the system architecture for Prototype 5.

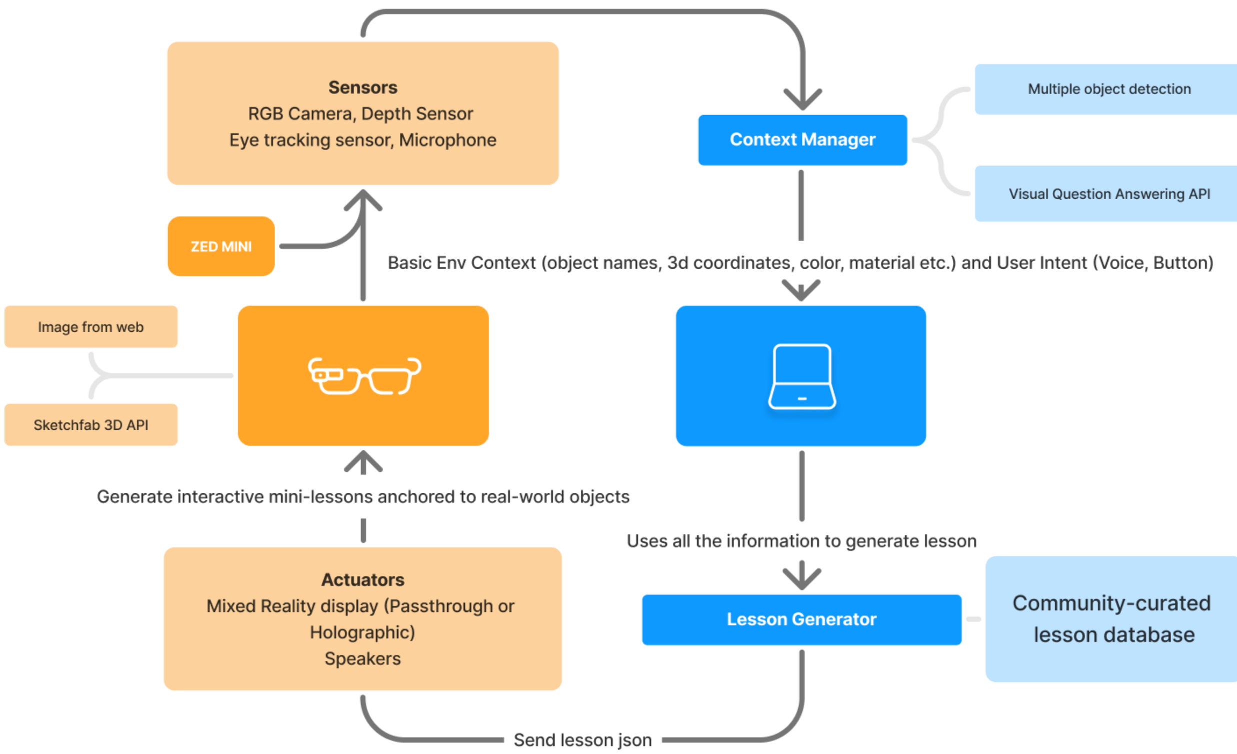

Figure 41: System architecture for Prototype 5. See Appendix E for details.

5.5.4 Reflection

This prototype would lie close to the central part of the Reality-Virtuality continuum (Milgram et al., 1994). Similar to prototype 3, It has spatially anchored, interactive, fully-colored 3D models and text augmented in the user’s environment. The context recognition agent system was also similar to the previous prototype, determining objects’ spatial coordinates and user preferences. This prototype similar to the previous prototype focused on the principles of situated learning where the surrounding objects generated the learning content for the user. It also had components of curiosity-based learning which allowed the user to set their topic of preference using the VUI.

The lesson content presented to the user had multiple components such as text, images, and 3D models which would ideally be curated by an educator. This allows for better learning content complexity. The content is also determined based on the learner's interests.

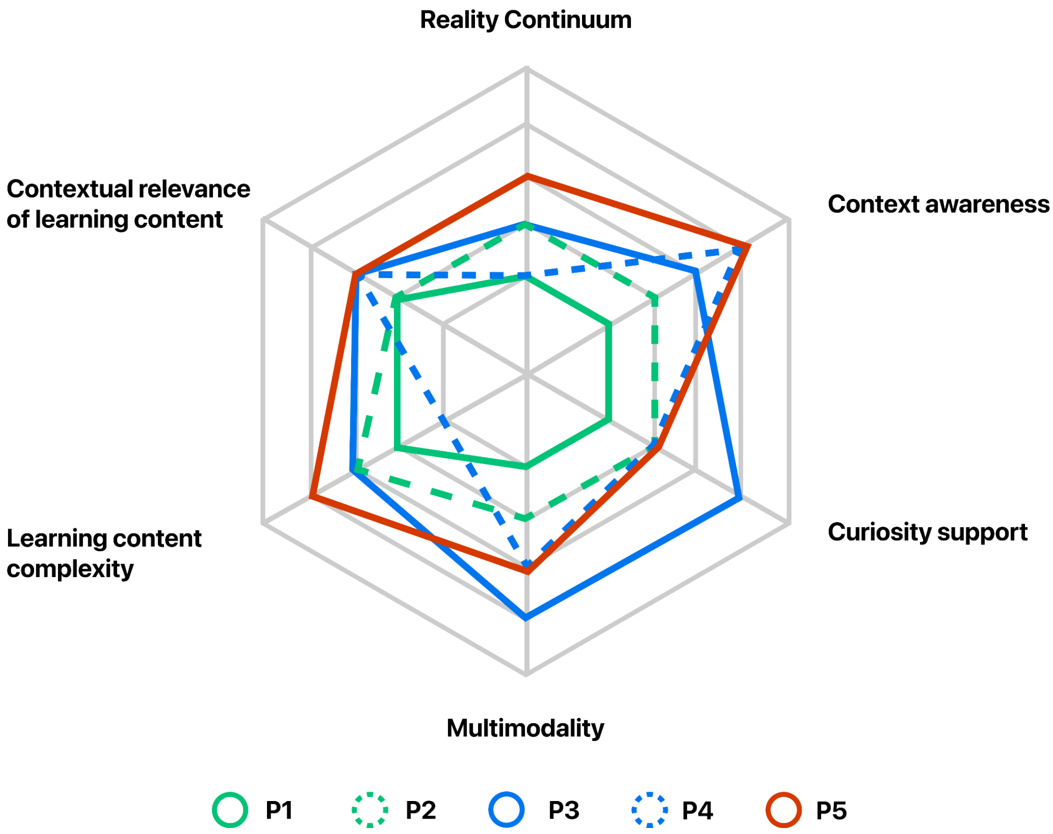

Figure 42: Radar plot comparison of all prototypes

There were a few challenges associated with this prototype. Firstly, there wasn't any interface available for learners to support their queries after they interacted with the template content. Another challenge was the repetition of the same lesson experiences due to a limited amount of educator-curated lessons which wouldn't be able to make the best use of educational content on the web. Hence, I considered combining parts of all the prototypes that provided value to the learning experiences and building a final prototype, "Curiosity XR" which I discuss in the next chapter.

6. Final Prototype: Curiosity XR

After having explored various sub-questions through the past prototypes, I decided to combine the modules built into an MR educational application called “Curiosity XR”. Curiosity XR aimed to benefit through the learnings by building the earlier prototypes.

I started building Curiosity XR around Jan 2023 on the Hololens 2 but during building it, Meta released the Quest Pro headset with color passthrough capabilities. Quest Pro provided much more advanced hardware, a better field of view, better rendering capabilities, and ease of sharing this application through the Quest store. Due to these benefits compared to the Hololens platform, I decided to switch this application to the Quest platform during the Winter of 2023.

This prototype also explored the last sub-question SQ6; “ How can we evaluate if such lessons support the learner’s curiosity and retention of knowledge?”. User participants were recruited to experience a lesson through Curiosity XR which helped analyze its potential to support learners’ curiosity and retention of knowledge as compared to the traditional learning methods.

6.1 Process

To start laying down the requirements for this application, I listed down the modules which needed to be imported from the previous prototypes and noted the learnings and analyzed how they could be incorporated into this application.

P1; the language learning prototype was an efficient example of translating objects around the user to teach languages through a minimal UI. The language translation module would be required to support language learning for Curiosity XR. That prototype was quick and provided translations in real-time due to the offline computation for the learning content. To carry forward these learnings into Curiosity XR, I decided to have translation and text-to-speech (TTS) modules running offline to provide minimal latency.

The next prototype, P2, explored supporting curiosity through sketches. Although the prototype worked well for a certain set of use cases, it had a lot of limitations due to technical constraints.

Later, an alternative interface, VUI in the next prototype was analyzed as a better interface to support learners' curiosity. Hence, I decided not to explicitly use any of the modules of P2, rather using many functionalities of P3 (some of which were imported from P2) would be quite relevant for this application.

Reflecting upon the learnings from P3, I decided to import the intent recognition agent which allowed users to use a natural language interface to select their topics of interest, request 3D models, ask questions, and get results as text. I would also be importing the Text-to-speech module to display the response in the form of a speech.

The final two prototypes, P4 and P5 featured spatial object detection with 3D coordinates using the zed mini which allowed for learning content anchored to the objects in the real environment. This environment context recognition system would be imported along with both the AI-generated and user-curated lesson support. A new agent would be necessary to choose whether an AI-generated lesson or an educator-curated lesson is more relevant to present to the user.

After laying out the learnings from the earlier prototypes, I proceeded with storyboarding and sketching concepts for Curiosity XR (Figure 43). This helped me visualize the application as a whole and how a mixed reality environment would look like for the learners when provided with different functionalities. Storyboarding also helped me visualize two ways to initiate curiosity-support for the users, either agent-based or user-based.

To choose an appropriate educational concept for the user at a given time when using Curiosity XR, various context parameters would play an important role. Eg. the User's past encountered lessons, what they're curious to learn, the environment objects and the physical properties of these objects, the user's location, the time, etc. Once a relevant concept has been chosen, using the Prototype 4 and Prototype 5 modules, an AI-assisted lesson and a user-curated lesson are generated. The system can be trained to choose which lesson to choose for the learner through A/B testing. (Figure 44)

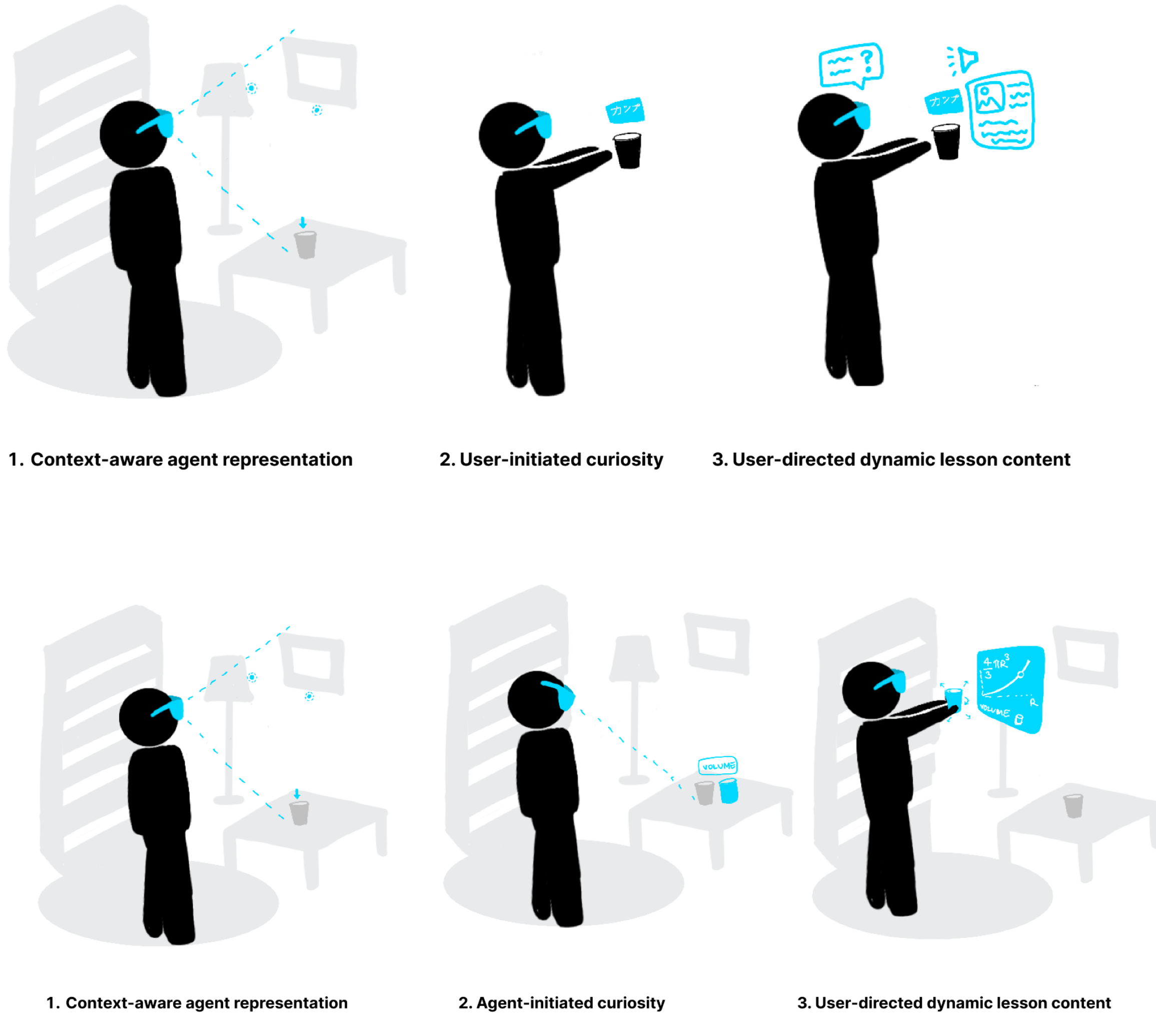

Figure 43: Storyboard concept for user-initiated curiosity(Top), Storyboard for agent-initiated curiosity (bottom)

The next step would be to provide capabilities for the possible templates which the lessons could use. The modules imported from the past prototype allowed for lessons to display a title, description, images, and 3D models. It is possible that video content could be a great addition to some lessons for eg. If a user needs to be shown a lesson about how paper is recycled, a video would be quite helpful, while an interactive 3D model wouldn't help much in that case. Hence I decided to incorporate the capability of adding a video to the lesson as well.

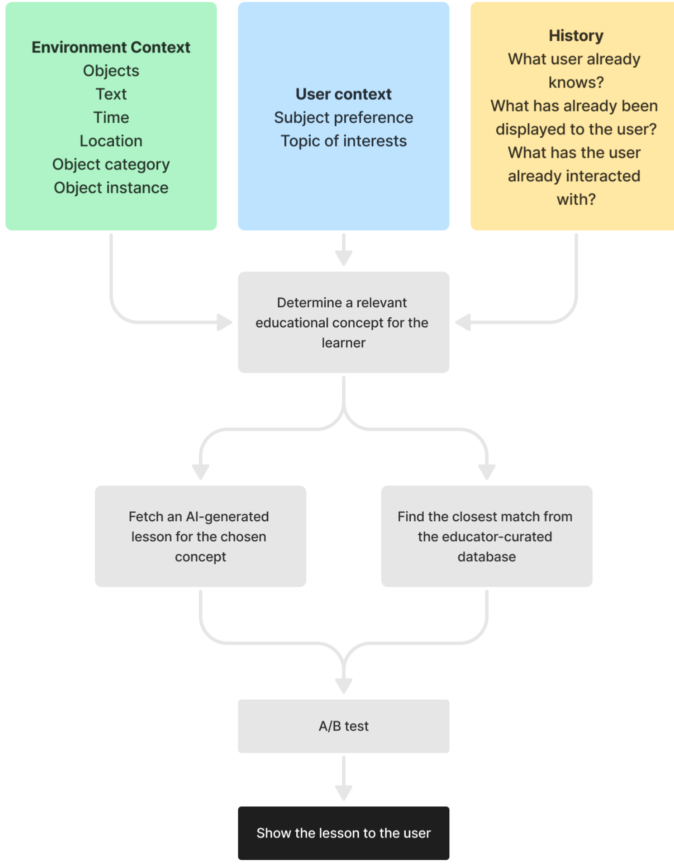

Figure 44: Determining a relevant concept and a lesson for the user

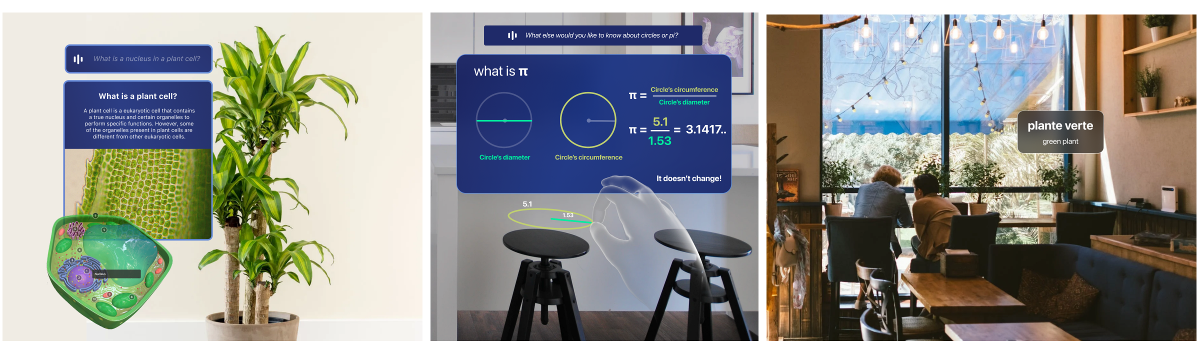

Figure 45: High-fidelity prototypes for Curiosity XR

As VUI is a major component for interactions and was required to perform various tasks merging from the different prototypes, it was necessary to rethink its flow. Reflecting upon the learnings from Prototype 3, where the VUI was always active causing false interpretations, I had to reimagine the voice activation interaction. One of the methods would be to use a wake word similar to “Hey Siri”, or “Alexa” before speaking out a command. However, I wanted to provide an alternative way to initiate voice commands, which were better suited for an educational environment and intuitive as well. After some ideation, I decided to proceed with a gesture-based initiation. This approach had various benefits, including, helping provide active multimodal participation of learners, the hand provided an anchor for placing 3D objects as if they were in the user's hand, and seemed a natural interaction to initiate a query.

6.2 Outcomes

Curiosity XR was initially built for the Hololens 2. The application supports AI-based and user-curated Mixed Reality mini-lesson with curiosity-driven and personalized lesson content generation to support learning experiences across various domains of Math, Science, History, Geography, Language, etc. Later in Feb 2023, Curiosity XR was transitioned to the Quest platform (Figures 46 & 47), providing a better overall Mixed Reality experience on Quest Pro due to the color passthrough but also supported Quest 1 / 2 with the greyscale passthrough.

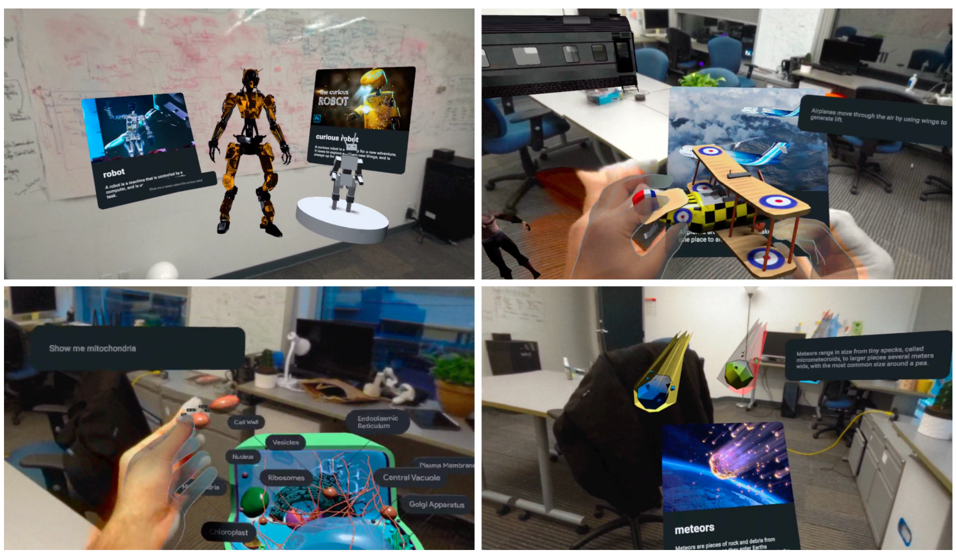

Figure 46: Screenshots from Curiosity XR, (Top-right) Learning about robots using robot models[6], (Top-left) Learning about airplane aerodynamics using airplane model[7], (Bottom-left) Requesting to show a specific part of a plant cell while looking at a plant cell model[8], (Bottom-right) Learning about meteors with meteor models[9].

---

[6] Biped robot (https://skfb.ly/KBnH) by Willy Decarpentrie is licensed under Creative Commons Attribution (http://creativecommons.org/licenses/by/4.0/).

[7] "The Flying Circus: Stylized WW1 Airplane" (https://skfb.ly/6ZFPI) by Alyssa Valcorza is licensed under Creative Commons Attribution (http://creativecommons.org/licenses/by/4.0/).

[8] Plant Cell (https://skfb.ly/oC8YM) by brianj.seely is licensed under Creative Commons Attribution (http://creativecommons.org/licenses/by/4.0/).

[9] Meteors (https://skfb.ly/6SAQp) by octopushh is licensed under Creative Commons Attribution (http://creativecommons.org/licenses/by/4.0/).

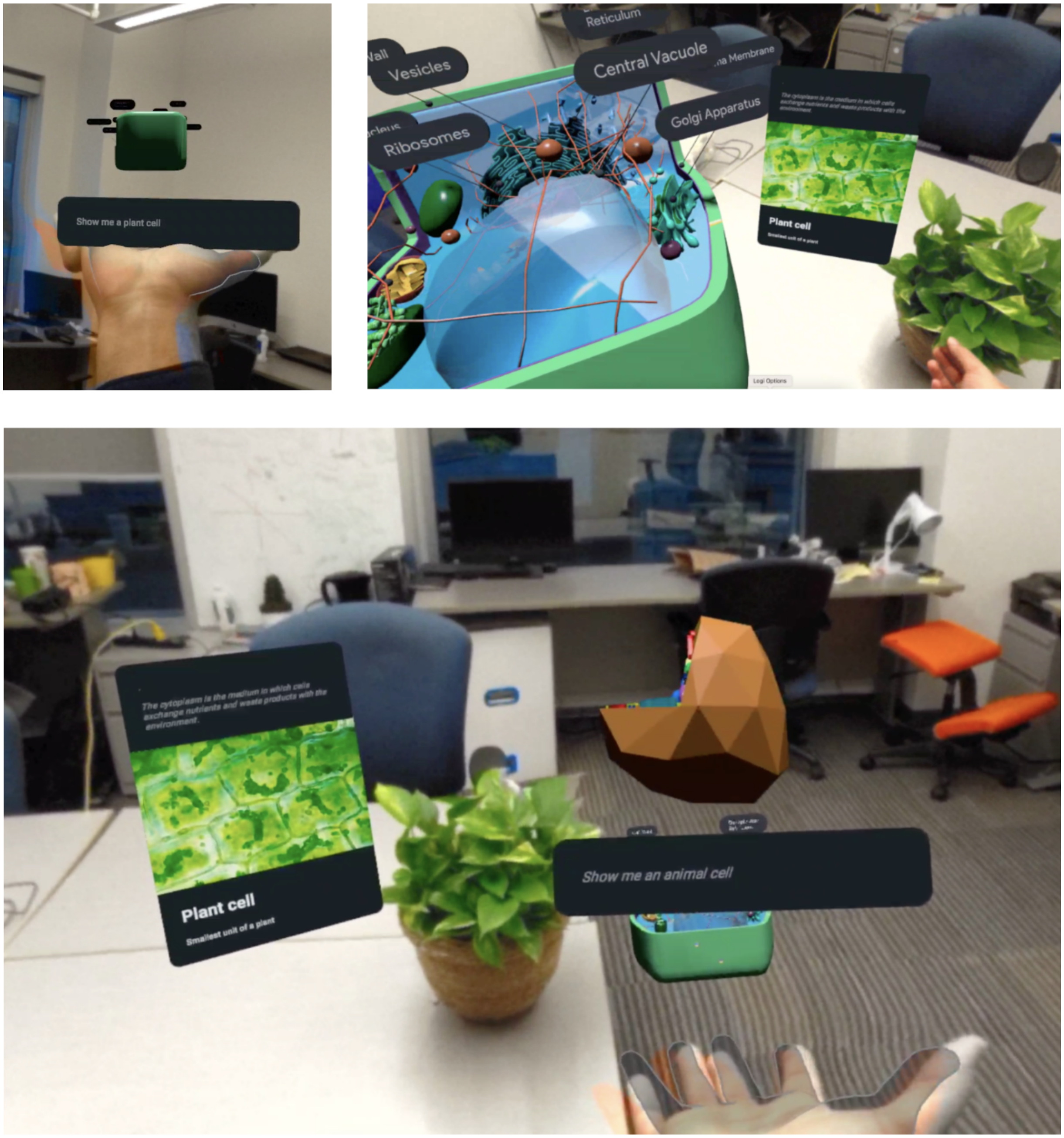

Figure 47: Screenshots from Curiosity XR, (Top-left) Open-palm gesture to activate voice commands, (Top-right) Plant-cell lesson content (title, description, image, plant cell[10] model) initiated through a real plant in the surrounding, (Bottom) Open-palm gesture to request and load Animal cell model[11] other 3D models.

---

[10] Plant Cell (https://skfb.ly/oC8YM) by brianj.seely is licensed under Creative Commons Attribution (http://creativecommons.org/licenses/by/4.0/).
[11]Animal Cell (https://skfb.ly/6nZZP) by Forged1212 is licensed under Creative Commons Attribution (http://creativecommons.org/licenses/by/4.0/).

6.3 Architecture

The architecture system (Figure 48) for Curiosity XR is designed to support multiple components working together. First, the Context-awareness subsystem (Figure 49) gets the user and environment context. Secondly, the curiosity-initiation subsystem uses the acquired context to determine potential lessons and present curiosity prompts to the user as they explore their environment, and lastly, the multimodal learning-support subsystem supports users' curiosity through interactive educational information in MR.

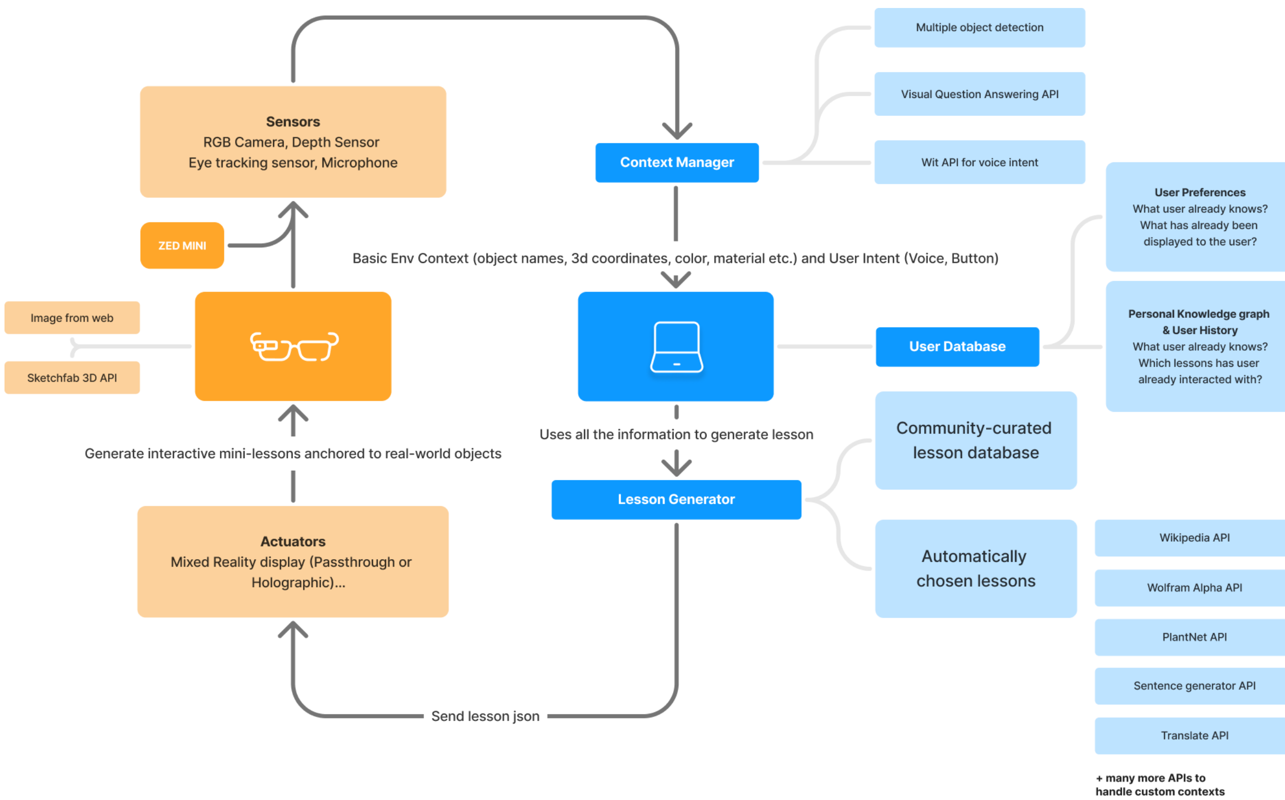

Figure 48: System architecture diagram for Curiosity XR. See Appendix D for more details on the architecture.

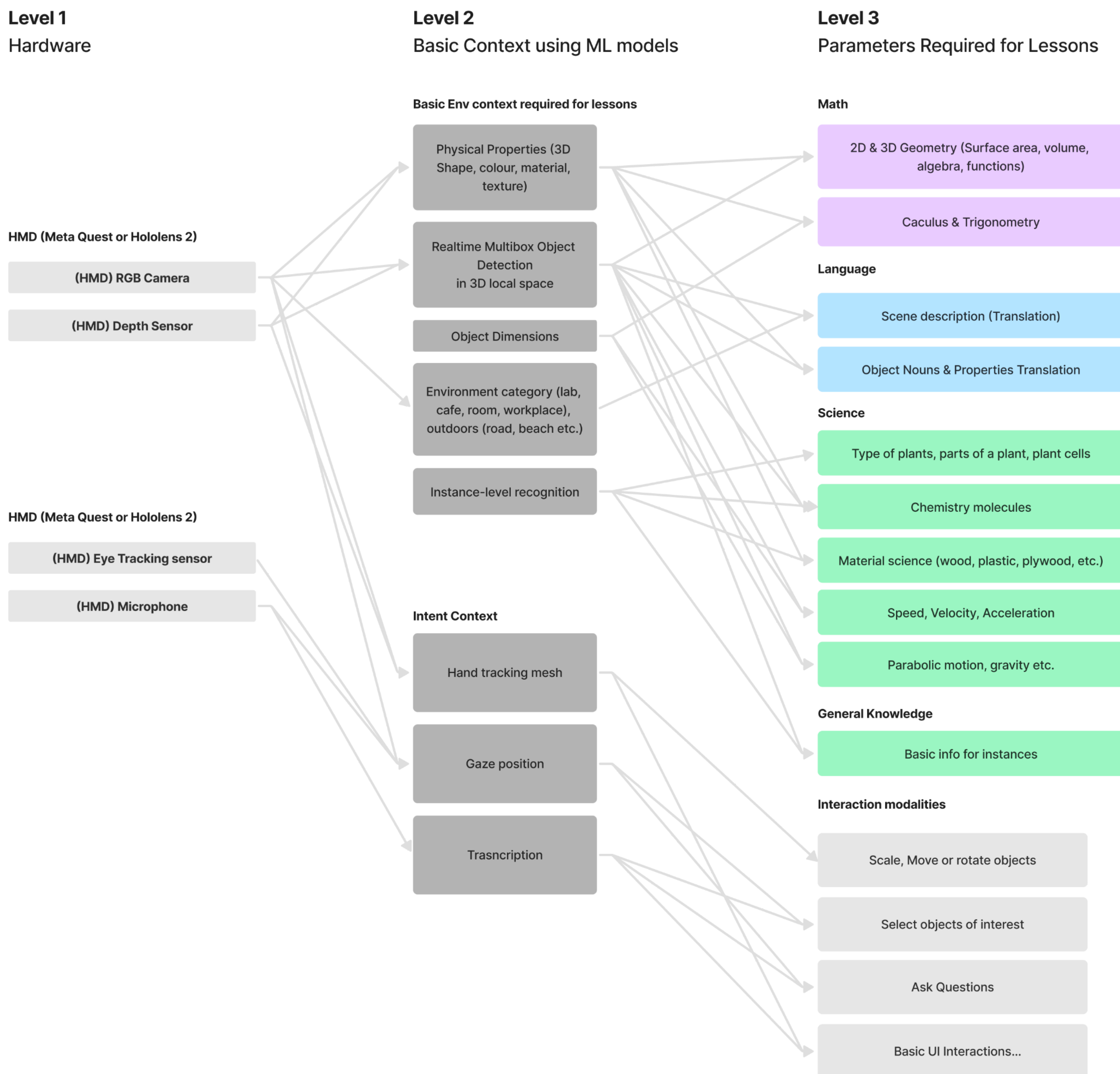

Figure 49: Context abstraction layers and their correlation to the learning content requirements.

6.4 User Study Evaluation

To evaluate Curiosity XR once it was complete and functional, I conducted an REB-approved user study with 6 participants (Figure 50) wherein the session provided users with the opportunity to experience learning concepts in Mixed Reality (MR) using an MR headset (Hololens 2 / Quest Pro) and later the same concept through a former method using a laptop. The study focused on understanding and measuring how well this application enhances learners' curiosity and knowledge recall as opposed to traditional methods of learning.

Figure 50: User participants using CuriosityXR to experience a lesson about a Plant cell

The research study evaluated participants' experiences using mixed reality as a learning tool for understanding plant cells and further exploring learning content using CuriosityXR. 2 of the participants used Hololens 2 and 4 of them used a Quest Pro for this study. Participants generally found the tool engaging and immersive, allowing them to interact with 3D models in real-world settings and request new information. However, some participants found parts of the tool challenging to use, particularly with voice recognition and gestures. The tool was hard to use due to the issues with the accuracy of voice recognition where the system was unable to understand various speech patterns and dialect accents. They felt that the application didn't understand their questions and answers were sometimes difficult to understand. They also reported interface issues not detecting hand gestures when modifying models.

Participants wanted more 3D visualizations and self-directed learning prompts and wished the tool had a browser panel to access additional information on the web. They also wished for more control over queries and the ability to have different models. The tool's effectiveness was rated about 3.5 out of 5, with most participants rating it moderately helpful.

Due to the limited number of participants for this part of the study, nothing concrete can be said about enhancing curiosity or knowledge recall using these research prototypes. Both the participants had an engaging experience but some of the major concerns were related to the usability of CuriosityXR. Figure 51 shows a bar chart visualization of the evaluation of six user participants in terms of effectiveness in aiding learning (rated by participants), engagement level, and ease of use (indicative from questionnaire transcription notes).

Overall, the study demonstrated the potential of mixed reality as a learning tool and highlighted the importance of improving voice recognition and more visualization support.

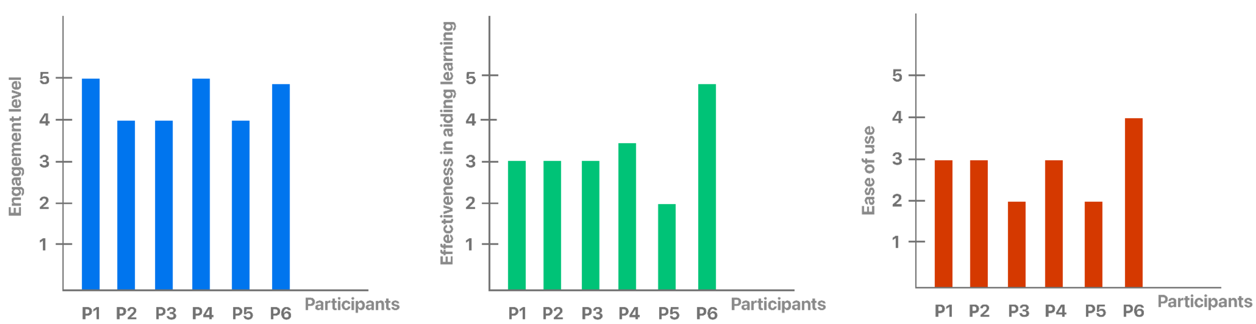

Figure 51: Ratings for CuriosityXR by 6 user participants. Left, Engagement level (indicative from questionnaire transcription notes). Middle, Effectiveness in aiding learning (Rated by participants). Right, Ease of use (indicative from questionnaire transcription notes).

6.5 Comparative subjective analysis

6.5.1 Comparison with previous prototypes

As compared to the previous prototypes, CuriosityXR covers a combination of abilities to help users provide curiosity support and a more immersive learning environment. (Figure 52)

With respect to the Reality-Virtuality continuum (Milgram et al., 1994), CuriosityXR had a similar setup to P5, allowing different types of learning media as part of the MR experience. It was able to support 3D models, text, images, and videos which had much more virtual content capabilities than any of the earlier prototypes, P1 - P4. The context recognition agent shares similarities with P4 & P5 in terms of identifying the location and category of objects in 3D space and understanding user learning preferences. It also adheres to the principles of situated learning, where the surrounding objects serve as the basis for learning material similar to all the previous prototypes. Furthermore, it incorporates elements of curiosity-based learning and constructivist learning. It enables users to specify their preferred topics using the VUI, ask questions to drive the learning content dynamically, and request models to visualize what they are curious to know about. CuriosityXR also supports educator-curated and AI-generated learning content which enables it to provide more contextually-relevant content.

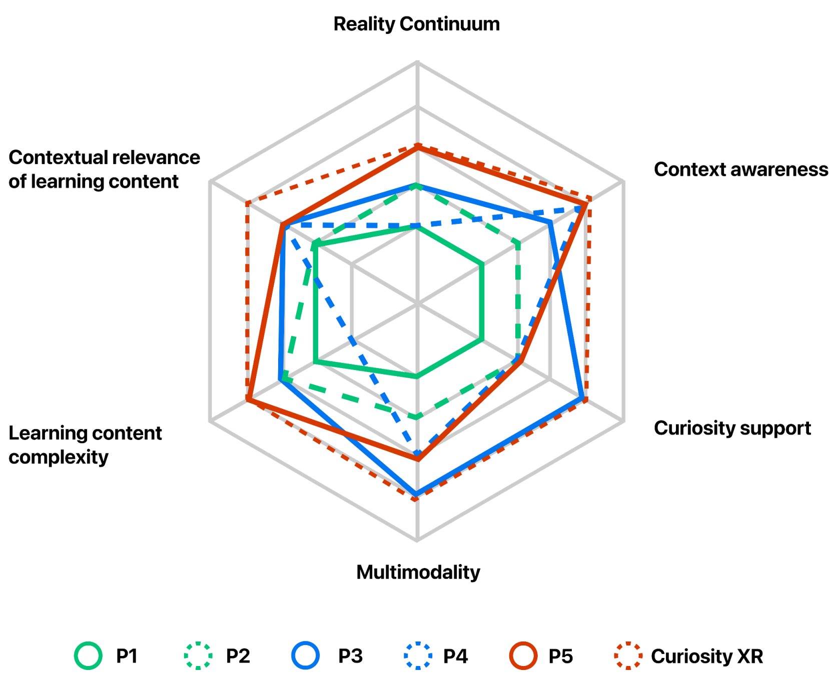

Figure 52: Comparison of the final prototype; CuriosityXR with all the prototypes

6.5.2 Comparison with related works

Compared to the related works, CuriosityXR offers a distinct learning experience through the use of Mixed Reality. Figure 53 showcases a Radar plot that subjectively compares CuriosityXR with each related work.

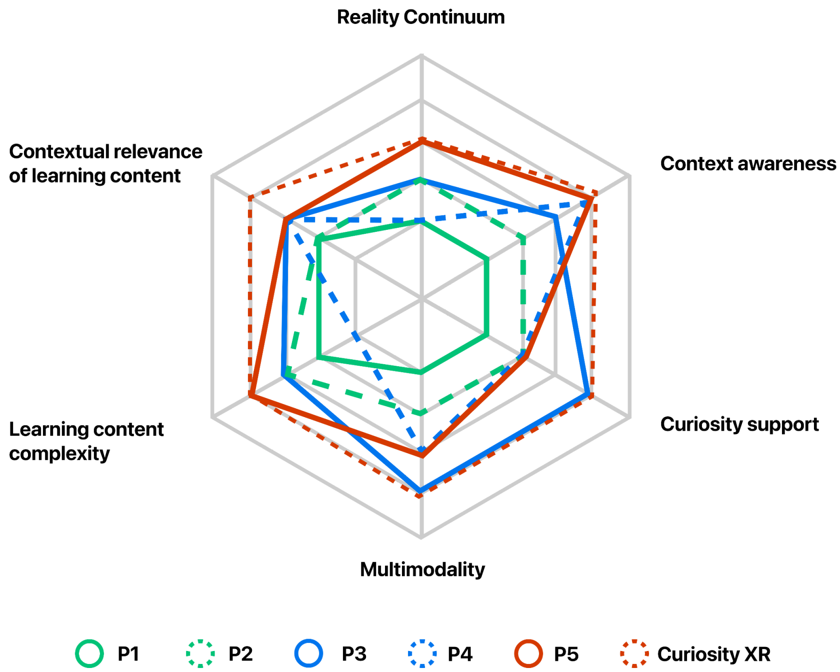

Figure 53: Comparison of the final prototype; CuriosityXR with the related works.

While each of the related works utilized a specific learning pedagogy to facilitate their applications, CuriosityXR leveraged a combination of learning pedagogy frameworks to create an immersive learning environment.

None of the related works provided hand interaction capabilities for immersive learning, unlike CuriosityXR. With this feature, learners can directly explore models and learning content with their hands, providing a more tangible and natural interaction within a relevant contextual setting. For example, manipulating a plant cell model with one hand while touching a leaf with the other. The related work applications used various techniques to provide contextually relevant learning content, whereas CuriosityXR enables much deeper, connected environment-context-based learning content.

CuriosityXR aims to deliver learners an immersive and dynamic experience, but it does have technical limitations and design issues. It solely uses object category context to generate learning content and doesn't capture accents, dialects, or support multimodal context, which could allow for a better understanding of audio and visual information.

7. Reflection & Future work

7.1 Reflection

This research aimed to explore how Context-aware agents could support curiosity and knowledge recall through Mixed Reality. This question was explored through different sub-questions that investigated topics such as acquiring context from mixed reality devices, creating curiosity-support agents, and evaluating the effectiveness of mixed reality in supporting interactive learning experiences.

During this thesis, a series of iterative prototypes were designed, built, and analyzed to address various sub-questions that ultimately contributed to answering the main research question. Through this process, a range of contexts was explored, which informed the selection and integration of relevant educational content. Additionally, agents were designed and analyzed to deliver diverse forms of learning content that supported learners' curiosity.

An important aspect of the project was the analysis of the prototypes based on parameters that were deemed relevant to mixed reality educational environments. Through this evaluation, I was able to refine and improve the design of the system, taking into account considerations such as design solutions and interactions.

To measure the effectiveness of CuriosityXR in terms of knowledge recall and curiosity support, a user study was conducted. Although the results were limited due to time constraints and a small sample size, the study provided valuable insights into the system's capabilities, challenges, and benefits and helped analyze the design solutions. It is important to note that the study did not provide a concrete answer to the main research question directly; however, it offered valuable guidance for future work in this area, particularly in terms of design approaches, architectural frameworks, and curiosity-support agents that could be used to achieve these goals.

Overall, the iterative prototype process and user study helped to inform the development of CuriosityXR, resulting in a system that provides engaging and immersive learning experiences that support learners' curiosity. While further research is necessary to fully evaluate the system's effectiveness, this work represents a significant step towards creating educational experiences that harness the potential of mixed reality technologies.

7.2 Future Work

Curiosity XR has the potential in several key areas for further research and development.

(1) A near-future work would be to implement a feedback system, where users can provide feedback on the lessons they experience to evaluate their effectiveness and improve future iterations.

(2) A multimodal assistant to better support learning in MR, similar to how chatbots are trained and are trained on text -> text models, this multimodal assistant but with the ability to process a wide range of multimedia inputs such as text, image, and video, and 3D content, and hand interaction reference. The assistant would capture this multimedia context and provide an output that would include a combination of text, images, video, and 3D content, or even present an interactive subpart of the lesson to best support the user’s query considering the multimedia context.

(3) The inclusion of animation support and dynamic control for 3D models. Providing educators the ability to provide labels to subparts of 3D models, making it easier for learners to play with components and explore each component separately would greatly enhance the learning experience. Also, having the ability to provide animated 3D models could help better design instruction for various use cases. Eg. working of a bicycle, breakdown of a laptop, or growth of a plant, etc.

(4) A (Mixed Reality) interface application could be designed and built which would allow educators to create and design learning experiences for MR in MR. This would provide an immersive and interactive platform for educators to create the course material and better visualize the learning content presented.

(5) An addition of an auditory element to the environmental context would provide opportunities to create unique learning experiences, which would trigger relevant lesson content based on specific sounds or cues in the environment, such as the sound of a

bird triggering a lesson about that bird. These developments have the potential to greatly enhance the effectiveness and engagement of digital learning experiences and provide a more holistic and interactive learning environment for students.

(6) One potential avenue for future research is exploring how AI can be integrated into our surroundings as an object itself, equipped with contextual information about the object. The learner could then have a conversation with the object, asking questions to learn more about it and the AI would respond through an embodied object perspective.

(7) Support for local multi-user learning experiences would provide the following benefits. Local multiplayer MR experiences in educational applications can facilitate increased collaboration and socialization among students. By working together in the same physical space, students can view models, discuss concepts, solve problems, and help each other learn, which can lead to a more engaging and interactive learning experience. It would also benefit the teachers to be able to produce interactive learning content on-demand and facilitate an immersive collaborative learning experience with the students.

8. Conclusion

8.1 Revisiting Objectives

Much of the literature has explored the use of emerging technologies for learning through pre-determined instruction that may not be applicable or significant to each student. The literature also has gaps in exploring curiosity-based learning using MR.

The goal of this research is to create and evaluate an application, Curiosity XR to facilitate self-guided learning and foster personally relevant and engaging learning experiences driven by curiosity. Curiosity XR hypothetically helps better support learners with contextual educational content and self-directed learning tools as opposed to other traditional learning techniques such as reading books or watching videos on 2D displays.

8.2 Revisiting Contributions

In this thesis paper, a mixed reality learning system, Curiosity XR is presented to meet the research goals. Curiosity XR allows users to learn languages, science, history, general knowledge, and mathematics concepts through the objects around them and the environment. It also provides tools to self-direct the learning content through immersive interactions in MR.

(1) A thorough literature review at the intersection of three fields: Mixed Reality (MR), Context, and educational theories which are necessary for the creation and evaluation of a research project similar to Curiosity XR.

(2) An approach to mapping multiple context sources for multi-modal learning through agents using APIs (deep learning models) to support curiosity and knowledge recall among learners.

(3) An architectural framework to support a new form of learning and teaching supported by MR multimodal mini-lessons. For educators, this architecture provides support by allowing them to create new mini-lessons based on 3D user interface templates for various educational topics. The architecture also supports self-driven learning in context through the use of 3D visualizations, access to the educational content, and conversational agents which allow them to explore further into the topic by querying the system. This is achieved by combining head-mounted displays, external sensors, conversational agents, 3D user interfaces, and machine learning models. This architectural framework has been instantiated for the proof-of-concept use cases below involving mathematics and biology.

(4) CuriosityXR: Overall, the iterative prototype process and user study helped to inform the development of CuriosityXR, resulting in a system that provides engaging and immersive learning experiences that support learners' curiosity. While further research is necessary to fully evaluate the system's effectiveness, this work represents a significant step towards creating educational experiences that harness the potential of mixed reality technologies.

(5) Validation of the benefits of MR and its potential to support curiosity for education and learning purposes.

8.3 Revisiting Limitations

This work does not focus on building a structured pedagogy or helping curriculum design for education. This work is not targeted to a specific user audience and aims to explore a general platform that could support learning content across different fields.

Although this work was aimed to explore metrics for knowledge retention, due to a limited number of participants and time constraints it does not have concrete conclusions about whether the techniques produced are better in the knowledge retention metrics as opposed to the traditional learning methods.

There are a few remarks about this tool as a curiosity-support agent that depicts positive experiences from the participant study but does not have quantitative comparisons of measuring curiosity and comparing it with other techniques. The curiosity here is used more in the sense of a general connotation of “wanting to know/learn more” aspect and does not talk about or refer to the psychological or philosophical nature of this term and is not the scope of this project either.

8.4 Final Remarks

In summary, this thesis paper explores the potential of mixed reality (MR) technology to support self-guided and curiosity-based learning experiences. The paper presents Curiosity XR, a mixed reality learning system, and an architectural framework that supports self-driven learning in context through the use of 3D visualizations, access to educational content, and conversational agents. While the paper does not focus on building a structured pedagogy, it validates the benefits of MR technology and its potential to support curiosity for educational purposes. However, the paper acknowledges its limitations in terms of limited data on knowledge retention and a lack of quantitative comparisons to other learning techniques in terms of measuring curiosity. As we move forward, it is crucial for researchers, educators, and developers to continue exploring and refining the use of MR technology in education to create more engaging and effective learning experiences. By embracing curiosity as a driving force for learning, we can create a future where XR systems change the landscape of education and provide students with the tools they need to explore and discover their interests.

9. Bibliography

Abowd, G. D., Dey, A. K., Brown, P. J., Davies, N., Smith, M., & Steggles, P. (1999). Towards a better understanding of context and context-awareness. In *Handheld and Ubiquitous Computing: First International Symposium, HUC'99 Karlsruhe, Germany, September 27–29, 1999 Proceedings 1* (pp. 304-307). Springer Berlin Heidelberg.

Agrawala, A. (2020). Ubiquitous Computing [Lecture notes]. Department of Computer Science, University of Maryland.

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. Educational research review, 20, 1-11.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational researcher*, *25*(4), 5-11.

Bacca Acosta, J. L., Baldiris Navarro, S. M., Fabregat Gesa, R., & Graf, S. (2014). Augmented reality trends in education: a systematic review of research and applications. *Journal of Educational Technology and Society, 2014, vol. 17, núm. 4, p. 133-149*.

Bada, S. O., & Olusegun, S. (2015). *Constructivism learning theory: A paradigm for teaching and learning.* Journal of Research & Method in Education, 5(6), 66-70.

Borg, M. (2021). GLTFast [Computer software]. *https://github.com/madebyhiro/gltfast*

Brown, J. S., Collins, A., & Duguid, P. (1989). *Situated cognition and the culture of learning.* 1989, 18(1), 32-42.

Bruner, J. S. (1966). *Toward a theory of instruction* (Vol. 59). Harvard University Press.

Colman, A., Hussein, M., Han, J., & Kapuruge, M. (2014). Context aware and adaptive systems. *Context in Computing: A Cross-Disciplinary Approach for Modeling the Real World*, 63-82.

Dasgupta, A., Manuel, M., Mansur, R. S., Nowak, N., & Gracanin, D. (2020). Towards Real Time Object Recognition For Context Awareness in Mixed Reality: A Machine Learning Approach. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 262–268. https://doi.org/10.1109/VRW50115.2020.00054

Dey, A. K. (2000). *Providing architectural support for building context-aware applications*. Georgia Institute of Technology.

Dey, A. K. (2001). Understanding and using context. *Personal and ubiquitous computing*, *5*, 4-7.

Dey, A. K., Abowd, G. D., & Salber, D. (2001). A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human–Computer Interaction*, *16*(2-4), 97-166.

Egger, J., & Masood, T. (2020). Augmented reality in support of intelligent manufacturing – A systematic literature review. *Computers & Industrial Engineering*, *140*, 106195. https://doi.org/10.1016/j.cie.2019.106195

Essa, I. A. (2000). Ubiquitous sensing for smart and aware environments. *IEEE personal communications*, *7*(5), 47-49.

Facebook Research. (2018). Detectron2: Detectron2 is FAIR's next-generation platform for object detection and segmentation [Computer software]. Retrieved from https://github.com/facebookresearch/detectron2

Fernando, S. Y., & Marikar, F. M. (2017). Constructivist Teaching/Learning Theory and Participatory Teaching Methods. *Journal of Curriculum and Teaching*, *6*(1), 110. https://doi.org/10.5430/jct.v6n1p110

Fischer, G. (2012). Context-aware systems: The “right” information, at the “right” time, in the “right” place, in the “right” way, to the “right” person. *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI ’12*, 287. https://doi.org/10.1145/2254556.2254611

Fitzgerald, F. T. (1999). Curiosity. *Annals of internal medicine*, *130*(1), 70-72.

Fyfe, E. R., McNeil, N. M., Son, J. Y., & Goldstone, R. L. (2014). Concreteness Fading in Mathematics and Science Instruction: A Systematic Review. *Educational Psychology Review*, *26*(1), 9–25. https://doi.org/10.1007/s10648-014-9249-3

Fyfe, E. R., & Nathan, M. J. (2019). Making “concreteness fading” more concrete as a theory of instruction for promoting transfer. *Educational Review*, *71*(4), 403–422. https://doi.org/10.1080/00131911.2018.1424116

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, *84*(2), 486-496.

Gryaditskaya, Y., Sypesteyn, M., Hoftijzer, J. W., Pont, S. C., Durand, F., & Bousseau, A. (2019). OpenSketch: a richly-annotated dataset of product design sketches. *ACM Trans. Graph.*, 38(6), 232-1.

Ha, D., & Eck, D. (2017). A neural representation of sketch drawings. *arXiv preprint arXiv:1704.03477.*

Hincapie, M., Diaz, C., Valencia, A., Contero, M., & Güemes-Castorena, D. (2021). Educational applications of augmented reality: A bibliometric study. *Computers & Electrical Engineering*, *93*, 107289. https://doi.org/10.1016/j.compeleceng.2021.107289

Hong, J. I., & Landay, J. A. (2001). An infrastructure approach to context-aware computing. *Human–computer interaction*, *16*(2-4), 287-303.

Hong, J., Suh, E., & Kim, S.-J. (2009). Context-aware systems: A literature review and classification. *Expert Systems with Applications*, *36*(4), 8509–8522. https://doi.org/10.1016/j.eswa.2008.10.071

Huuskonen, J., & Oksanen, T. (2018). Soil sampling with drones and augmented reality in precision agriculture. *Computers and Electronics in Agriculture*, *154*, 25–35. https://doi.org/10.1016/j.compag.2018.08.039

Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental review*, *32*(2), 125-160.

Kaufmann, H., & Papp, M. (2006). Learning objects for education with augmented reality. *Proceedings of EDEN*, 160-165.

Kokkonen, T., & Schalk, L. (2021). One Instructional Sequence Fits all? A Conceptual Analysis of the Applicability of Concreteness Fading in Mathematics, Physics, Chemistry, and Biology Education. *Educational Psychology Review*, *33*(3), 797–821. https://doi.org/10.1007/s10648-020-09581-7

Klopfer, E., & Squire, K. (2008). Environmental Detectives—the development of an augmented reality platform for environmental simulations. *Educational technology research and development*, *56*, 203-228.

Krashen, S. (2017). The case for comprehensible input. Language Magazine, 7(1), 1-6.

Kurubacak, G., & Altinpulluk, H. (Eds.). (2017). *Mobile technologies and augmented reality in open education*. Igi Global.

Li, C., Pan, H., Liu, Y., Tong, X., Sheffer, A., & Wang, W. (2017). BendSketch: Modeling freeform surfaces through 2D sketching. *ACM Transactions on Graphics*, *36*(4), 1–14. https://doi.org/10.1145/3072959.3073632

Litman, J. A., & Spielberger, C. D. (2003). Measuring epistemic curiosity and its diversive and specific components. *Journal of personality assessment*, *80*(1), 75-86.

Markey, A., & Loewenstein, G. (2014). Curiosity. *International handbook of emotions in education*, 238-255.

Maw, W. H., & Maw, E. W. (1961). Establishing criterion groups for evaluating measures of curiosity. *The Journal of Experimental Education*, *29*(3), 299-305.

McCarthy, J. (2007). What is artificial intelligence.

Meta (2019). Meta Quest 2 [Hardware].

Meta (2022). Meta Quest Pro [Hardware].

Microsoft. (2019). Hololens 2 [Hardware].

Microsoft. (2022). Mixed Reality Toolkit [Computer software]. *https://microsoft.github.io/MixedRealityToolkit-Unity/README.html*

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12), 1321-1329.

Niemann, K., & Wolpers, M. (2010). Enabling the Use of Real World Objects to Improve Learning. *2010 10th IEEE International Conference on Advanced Learning Technologies*, 259–263. https://doi.org/10.1109/ICALT.2010.77

OpenAI. (2022). GPT-3 OpenAI API. *https://openai.com/api/*

Perera, C., Zaslavsky, A., Christen, P., & Georgakopoulos, D. (2014). Context Aware Computing for The Internet of Things: A Survey. *IEEE Communications Surveys & Tutorials*, *16*(1), 414–454. https://doi.org/10.1109/SURV.2013.042313.00197

Poole, B., Jain, A., Barron, J. T., & Mildenhall, B. (2022). Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*.

Reiners, D., Davahli, M. R., Karwowski, W., & Cruz-Neira, C. (2021). The Combination of Artificial Intelligence and Extended Reality: A Systematic Review. *Frontiers in Virtual Reality*, *2*, 721933.

Rodden, T., Cheverst, K., Davies, K., & Dix, A. (1998, May). Exploiting context in HCI design for mobile systems. In *Workshop on human computer interaction with mobile devices* (Vol. 12).

Rosello, O., Exposito, M., & Maes, P. (2016). NeverMind: Using Augmented Reality for Memorization. *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, 215–216. https://doi.org/10.1145/2984751.2984776

Schilit, B. N., & Theimer, M. M. (1994). Disseminating active map information to mobile hosts. *IEEE network*, *8*(5), 22-32.

Schutera, S., Schnierle, M., Wu, M., Pertzel, T., Seybold, J., Bauer, P., Teutscher, D., Raedle, M., Heß-Mohr, N., Röck, S., & Krause, M. J. (2021). On the Potential of Augmented Reality for Mathematics Teaching with the Application cleARmaths. *Education Sciences*, *11*(8), Article 8. https://doi.org/10.3390/educsci11080368

Shafer, S. A., Brumitt, B., & Cadiz, J. J. (2001). Interaction issues in context-aware intelligent environments. *Human–Computer Interaction*, *16*(2-4), 363-378.

Sharp, H., Rogers, Y., & Preece, J. (2019). Interaction Design: Beyond Human-Computer Interaction (5th ed.). *John Wiley & Sons*.

Siew, C. Y., Ong, S. K., & Nee, A. Y. C. (2019). A practical augmented reality-assisted maintenance system framework for adaptive user support. *Robotics and Computer-Integrated Manufacturing*, *59*, 115–129. https://doi.org/10.1016/j.rcim.2019.03.010

Sketchfab. (2022). 3D Models [Online platform]. *https://sketchfab.com/*

Speicher, M., Hall, B. D., & Nebeling, M. (2019). What is Mixed Reality? *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3290605.3300767

Stereolabs. (2018). ZED Mini [Hardware]. *https://www.stereolabs.com/zed-mini/*

Suh, S., Lee, M., & Law, E. (2020). *How Do We Design for Concreteness Fading?Survey, General Framework, and Design Dimensions*. 8.

Sumadio, D. D., & Rambli, D. R. A. (2010, March). Preliminary evaluation on user acceptance of the augmented reality use for education. In *2010 second international conference on computer engineering and applications* (Vol. 2, pp. 461-465). IEEE.

Vaze, A., Clarke, I., Morris, A. (2023). Towards a Mixed Reality Agent to Support Multi-Modal Interactive Mini-Lessons That Help Users Learn Educational Concepts in Context.

Vazquez, C. D., Nyati, A. A., Luh, A., Fu, M., Aikawa, T., & Maes, P. (2017). Serendipitous Language Learning in Mixed Reality. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 2172–2179. https://doi.org/10.1145/3027063.3053098

Victor, B. (2014). Humane Representation of Thought [TED Talk]. *http://worrydream.com/HumaneRepresentationOfThought/note.html*

Von Glasersfeld, E. (1989). Cognition, Construction of Knowledge, and Teaching. *Synthese*, *80*(1), 121–140.

Vuzix Corporation. (2018). Vuzix Blade [Smart glasses]. *https://www.vuzix.com/products/blade-smart-glasses*

Wang, C., Tang, Y., Ma, X., Wu, A., Popuri, S., Okhonko, D., & Pino, J. (2020). fairseq s2t: Fast speech-to-text modeling with fairseq. *arXiv preprint arXiv:2010.05171*.

Yaghmaie, M., & Bahreininejad, A. (2011). A context-aware adaptive learning system using agents. *Expert Systems with Applications*, *38*(4), 3280-3286.

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007, April). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 493-502).

Zimmerman, J., Stolterman, E., & Forlizzi, J. (2010). *An Analysis and Critique of Research through Design: Towards a formalization of a research approach*.

10. Appendix

Appendix A: REB study overview

This research aims to develop a platform that supports educators and education content creators to create mixed-reality learning lessons that can be accessed by students through Mixed reality headsets. The lessons are aimed to teach users about the environment. For eg. Looking at a coffee cup could teach a learner about coffee plantations or interacting with a plant could show a lesson about photosynthesis. We would like to analyze this learning technique instead of the conventional methods of using books, images, or videos.

What: A context-aware Mixed Reality system to support environment-based & learner-directed multimodal educational lessons that help improve curiosity and knowledge retrieval in learners.

Who: 15 participants will be selected from an applicant pool based on subject knowledge, and interests. Participants should be able to visit the OCAD U campus to engage in MR lesson experience (to be conducted in compliance with OCADU's social distancing guidelines and equipment use restrictions) and to access requisite MR equipment. There are no required demographics for the participants.

Why: We would like to get usability feedback and understand if an interactive self-directed learning method in context can help recall information and result in a deeper understanding of concepts. It would be impossible to carry out this study without human participants. Building a platform to support such learning techniques and learning their benefits carries out a significant potential both for learners and educators and the technical support systems for the current education system.

Where: Hosted on-site at 230 Richmond St W, OCAD U.

Structure:

1. Completing an in-person pre-session survey on educational interests. (5 mins)
2. Setup and training session for the Mixed Reality headset. (5 min)
3. Learning session and UX interview. (30 mins)
4. 7 days later: Recall session and exit interview(15 min) (10 min)

Time Requirement (in hours): 1 hour

Detailed activities can be found in Appendix G: Research Protocol

Appendix B: REB study session details

The research will be scheduled to run 2 sessions and ensure that this MR experience is organized in a socially-distanced context and is compliant with OCADU Safety Protocols. All equipment will be thoroughly cleaned and disinfected between sessions.

This session will be hosted in 2 parts, the first session with the major experience and a follow-up second session for an interview to measure recall with a total time up to a maximum of 1 hour per participant.

Details of the in-person learning session experience:

Total time: 55 mins

1. **First session (In-person Mixed Reality experience and interview): 40 mins**

a. **In-person pre-session survey:** Choose a lesson that the user would interact with based on their educational interests in science, math, or languages. **(5 min)**

b. **Setup and training session** for the Mixed Reality headset. **(5 min)**

c. **Learning session and UX interview**: Wearing a mixed reality headset to experience a lesson to learn either a science concept, a math concept, or a language concept. During this mini-lesson users would be interacting through voice and touch with real objects (coffee cup, plant, table, chair, etc.) and virtual learning content (audio, 2D media, and 3D models). Learning a different concept of the same topic with a textbook and/or video content on a laptop. Completing a Post-Session in-person Interview for usability feedback and user experience or enthusiasm to learn more to compare traditional techniques with our technique. **(30 mins)**

2. **Second session (Recall session and exit interview): 15 mins**

After the preliminary screening survey, the selected participants would have to complete a pre-session survey on prior educational knowledge. The aim of this survey is to select a lesson suitable for the participant to minimize bias in measuring the learning experience outcomes.

The 10-minute online survey will be hosted on Microsoft Forms and be completed individually.

1. What is your prior knowledge in learning Maths on a scale of 1-5? (1-Beginner, 5-Expert)
2. Could you highlight your prior experience in learning Maths?
3. What is your prior knowledge in learning Science on a scale of 1-5? (1-Beginner, 5-Expert)
4. Could you highlight your prior experience in learning Science?
5. What is your prior knowledge in learning a second language on a scale of 1-5? (1-Beginner, 5-Expert)
6. Could you highlight your prior experience in learning this language?

POST-SESSION IN-PERSON INTERVIEW QUESTIONS

A 15-minute debrief interview takes place after the learning experience and seeks to capture reflections from the experience:

1. Tell us about your experience.
2. What did you find yourself enjoying most during the learning process? Why?
3. How did you find the experience of learning through mixed reality as opposed to watching a video or reading the concept from a book?
4. Do you think you are more curious to learn about the concept you learned now?
5. At what points was the tool hard to use?
6. How did this tool feel to use?
7. What is one thing you wish was different in this tool and which would have made your learning experience better?
8. What, if anything, will you take from this experience?
9. On a scale of 1-5, how well do you feel it helped you learn (1 - not at all, 5 - extremely helpful)
10. For the subject that you were assigned, how well do you think this system helped you learn?
11. What specific information are you able to recall from the lesson that you learned?

(RECALL AND EXIT INTERVIEW) POST- 1 WEEK IN-PERSON INTERVIEW QUESTIONS

A 15-minute in-person interview will take place after 7 days from the learning experience:
Recall session interview questions:

1. For the subject that you were assigned, how well do you think this system helped you learn?
2. For the subject that you were assigned, are you interested or curious to learn more about it?
3. What specific information are you able to recall from the lesson that you learned?

Appendix C: Unity C# code and project website

https://github.com/thisisvaze/the-learning-sense

https://www.thisisvaze.com/works/things/

Appendix D: CuriosityXR Architecture

The context-awareness subsystem is equipped with multiple processing layers and uses Zed mini's depth sensor and camera streams for real-time 3D object awareness, instance-level recognition techniques, and basic shape, color, material, and object name properties to support context for various lessons. Real-time scene understanding is set up by offloading heavy GPU tasks for inference and processing onto a dedicated local server. This subsystem is equipped with APIs for multiple object detection, text-to-speech, speech-to-text, speech-intent-recognition (Wit.ai), visual-question-answering, language translation, etc. This subsystem also features multithreading capabilities to run constant inference for multi-box object tracking using the point-cloud, and video stream data stream from Zed Mini/Headset, keeping it readily available for context analysis. Real-time information updates such as user interactions, user intents, user speech, status updates, etc. are sent via WebSocket one-to-many communication between the server and the headset.

The front end of the system is powered by Unity using Oculus Integration SDK, which offers support for Quest 1/2, Rift, and Quest Pro. The system features several interaction design elements, including hand interactions, natural language interface, and contextually-aware interactions. The curiosity-initiation system uses interaction design to create a concept-context map and understand how it relates to curiosity. The system can generate both user-generated and system-generated curiosity components. The multimodal learning-support subsystem allows for various modalities during the learning experience such as the gaze, touch (scale, rotate, and move), speech, audio output, hand-interactions, and tactility through the relevant real-world objects while learning. The system also allows users to move around the space for enhanced perceptions and interactions with the real world and virtual objects. Overall, the architecture system for Curiosity XR offers a robust and comprehensive platform for immersive and engaging learning experiences.

Appendix E: Prototype Architectures

Prototype 1

The application uses simpler technologies and doesn't need heavy computationally intensive processing and can work without the need for the internet or external computation. Figure 14 provides a simpler visualization of how the agent system works.

Figure 15 depicts the architecture diagram for this prototype. A live stream from the RGB camera in the smartglasses is used to get offline real-time multi-box object detection and based on the head position the closest object to the user's gaze is chosen and the translated object along with the object's name in the native language (chaise - chair) is

presented on the HUD screen. The application uses TensorFlow lite for offline object detection and Offline Google Translate API which supports 50+ languages.

Prototype 2

There are majorly four technical parts to this prototype, first, getting the paper crop from the camera stream, then, detecting the user-drawn object on the paper, further using SketchFab to display the model, and finally adding the hand interaction abilities (rotate, scale and move) for the 3D model. Figure 20 demonstrates the process of recognizing an object from a sketch to showing it in MR.

For the first part, the prototype uses an OpenCV custom module, which is similar to the techniques used to crop pages to create PDFs from images. The parameters were fine-tuned to meet the requirements of this use case. For the second part, a custom-trained model using the Google Quickdraw dataset was used. The Google Quickdraw dataset was collected from mouse or touch gesture-drawn sketches on a website and hence is vector-based and needed to be transformed into raster-based drawings to make it usable for this use case. To use the dataset for recognizing objects from the Hololens' camera stream, a custom-trained model was trained through transfer learning using a Vision Transformer pre-trained model.

For the third part, the SketchFab search API was used to request a model, for eg. "apple", which responds with search results in JSON format. The developed subsystem filters and finds the most suitable 3D model with a mesh having less than 50,000 triangles (a technical limitation of the Hololens 2) and fetches the model ID. This model ID is used to download the 3D model in GLB/GLTF format and display it to the user using an open-source library; GLTFast (Borg, 2021) to load GLTF models in real-time using Unity.

For the final part, the system uses MRTK3 (Microsoft, 2022) which provides hand-tracking and object manipulation capabilities allowing users to perform hand gestures such as pinch, poke or tap to select objects and other intuitive gestures to rotate, scale, or move them around.

Figure 21 demonstrates the system architecture diagram for this prototype. It shows how information is transmitted from the headset to the python server and back to the headset using several APIs to support this prototype.

Prototype 3

This prototype uses a similar flow as the previous prototype to gather user context and show a 3D model. Figure 25 depicts how models are rendered in real-time using the user's voice context.

For the first part of this prototype, the system needed to understand the user's intent from the recognized speech. I used Wit.ai to support speech intent recognition. With the custom-trainable wit.ai tool, It was easy to train the system to recognize intents for speech such as "I would like to see Mars", "Show me a model of a ship" or "Can I see a 3D printer?" to capture two parameters. The first is that the user would like to see the 3D model, and the second is the object itself. The intent recognition system is understood for various such differences spoken by the user to understand parameters for Intent: “Load_3D_Model” and Traits: “Mars” etc. Once the system recognizes the object, a 3D model generation module (using SketchFab API) is used from the previous prototype to display that object to the user. With the model loaded in the space available to the user to interact with, for the final part, the last module enables the user to ask questions and get responses through displayed text or speech.

For this module, I used different tools available namely, the Wolfram API and GPT-3. Wolfram Alpha has a mission to collect and curate all objective data; implement every known model, method, and algorithm; and make it possible to compute whatever can be computed about anything. With their API it was possible to get either spoken or short textual answers to questions about history, science, math, social science, general knowledge, etc. This prototype uses their Spoken Results API and Short answers API. Although it covers a variety of fields and provides accurate results, it does have limitations with the amount of knowledge stored. For eg. it doesn’t provide answers to questions such as “What is the difference between a plant cell and an animal cell?”.

In such cases, the system falls back to the OpenAI API to find answers to users’ queries. OpenAI API uses GPT-3 (OpenAI, 2022) which is a general-purpose neural network that performs a variety of natural language tasks such as question-answering, translation, summarization, etc. This is used to get short responses to users’ queries. Once the query response is available, the system needs to display it to the user through text or speak it out. To display the results to the user through text, the prototype uses MRTK3 to display a 3D UI canvas with text. To speak out the results, the system uses a Text-to-Speech library to generate realistic human-like speech and then sends it to the hololens to play it through its speakers (Wang, 2020).

Figure 26 demonstrates the system architecture diagram for this prototype. It shows how information is transmitted from the headset to the python server and back to the headset using several APIs to support this prototype.

Prototype 4

The system has multiple components working concurrently. Figure 34 depicts a flow diagram for this system architecture. The main connection point is the python server, which is communicating with both the zed mini and the Hololens 2. A separate thread is running on this server which receives the RGB camera stream and point-cloud information of the environment on the server. The other main component is the WebSocket server which is communicating with the Hololens 2. All other modules to communicate with multiple APIs and offline inferences are called upon request.

The frame from the RGB camera stream is sent to the Multiple object detection module every second which uses the detr-resnet-50 (Facebook Research, 2018) trained model for local inference. The recognized objects with their center coordinates are further mapped to Zed mini's point cloud coordinates to acquire the 3D location of each object in space. These 3D coordinates are relative to the Zed mini camera and are further translated to the Hololens 2's position as the origin.

Prototype 5

There are two major components of this prototype, the educator side, and the learner side are discussed below.

For the educator side, the frontend framework is set up using React Native. Using the web, Android, or iOS, a user could select the mini-lesson elements. The user is able to fill up a form to create a mini-lesson. Once the form is submitted, all the mini-lesson component properties are saved in a JSON format which is added to the mini-lesson database. Each mini-lesson comprises an ID, which uniquely identifies the mini-lesson in the main database.

On the other side, there are a few components similar to the previous Prototype 4. The python server communicates with both the zed mini and the Hololens 2, spatially anchoring the learning content to objects in the user environment and uses gazepinch & curiosity-prompt buttons to initiate the mini-lesson. Figure 40 shows the system architecture diagram for Prototype 5.