Ethics in Artificial Intelligence for Virtual Reality: Navigating Challenges under the EU AI Act

Reno Yuri Camilleri, Vanessa Camilleri, and Matthew Montebello

University of Malta, Msida MSD 2080, Malta
reno-yuri.camilleri.20@um.edu.mt

Abstract. The integration of Artificial Intelligence (AI) into Virtual Reality (VR) has revolutionised immersive experiences across various domains, including education, entertainment, and immersive training environments. However, this convergence raises critical ethical concerns, such as user manipulation, data privacy, and societal biases. This paper examines these challenges through the lens of the European Union's AI Act, the first comprehensive AI legislation globally, which categorises AI systems by their risk levels and imposes tailored requirements to ensure ethical and transparent use. By exploring how this legislation applies to AI-powered VR, the paper highlights its implications for developers, designers, and users while emphasising the need to balance innovation with ethical responsibility. A proposed modular architecture is introduced to guide the development of AI systems in VR, addressing challenges such as inclusivity, data privacy, transparency, and risk management. This framework ensures compliance with the EU AI Act while fostering innovation. By advocating for responsible design, this paper aims to contribute to the ongoing discourse on creating ethical, safe, and user-centric AI-powered VR systems that respect fundamental rights and societal values.

Keywords: Artificial Intelligence, Ethics, Virtual Reality, EU AI Act, AI Governance

1 Introduction

Artificial intelligence (AI) is rapidly transforming our world, presenting exciting opportunities in various fields, including education. One particularly promising area is the use of AI in Virtual Reality (VR) applications. AI is being integrated with VR to create intelligent virtual environments. This involves using AI techniques to develop autonomous agents that can populate virtual worlds, such as virtual humans or animals [1, 2] that the player can interact with, promoting a more realistic environment. In addition, AI can also improve immersion through the use of more realistic soundscapes and better object interactions [3, 4].

AI-powered VR has shown promise in creating immersive experiences for entertainment, training, and even in realistic simulations such as surgery [1, 2, 5]. As these technologies continue to advance, however, ethical considerations

must be taken into account. The European Union (EU) has recognized this and has taken a proactive step by introducing the "EU AI Act" [6]. This legislation establishes a framework for defining and regulating AI within the European Union.

The EU AI Act aims to mitigate the risks associated with AI systems by categorising them into different risk levels and outlining specific requirements for each category [6, 7]. This categorisation includes unacceptable-risk AI systems (banned), high-risk AI systems (strict regulations) and limited or minimal risk systems (transparency requirements) [6, 7]. This Act impacts AI development and deployment in VR not just for European countries, but it will also affect companies globally who provide products or services to EU citizens [7]. This paper explores the ethical challenges of AI in VR, examining the key provisions of the EU AI Act and their implications for developers, designers, and users of these technologies.

1.1 Motivation

This paper is motivated by the increasing integration of AI into VR experiences. This development necessitates an analysis of the ethical implications, as VR intrinsically blurs the lines between real and virtual, raising significant concerns about the potential influence of AI on various aspects of the user experience [8].

VR, by its very nature, creates highly immersive experiences that have a considerable impact on users, potentially evoking strong emotions and influencing users' perceptions and actions [3, 8]. AI can then significantly improve this immersion by making the world feel more alive, such as through the use of AI NPCs that can listen and talk to a user [2]. This blurring of boundaries between reality and virtuality raises concerns about the potential for AI to manipulate user behaviour, compromise privacy, and negatively affect well-being [3, 8]. Ethical concerns also arise around data privacy, potential manipulation of users, and the responsibility of developers in ensuring user safety [3, 8].

The EU AI Act recognizes these concerns and seeks to ensure the responsible development and deployment of AI systems, fostering trust in the technology while safeguarding fundamental rights [6, 7]. This paper aims to examine the Act's implications for AI within the specific context of VR, raising two crucial questions:

1. What specific ethical challenges arise from the use of AI in VR, and how does the EU AI Act address these challenges through its risk-based approach?
2. How can we ensure the responsible development and use of AI in VR while simultaneously fostering innovation?

By delving into these questions, this paper seeks to contribute to the ongoing discourse surrounding the ethical considerations of developing and using AI in VR, specifically within the framework established by the EU AI Act.

1.2 Document Structure

Chapter 2 will delve into the project's background, discussing the ethical challenges of AI in VR, as well as discuss in depth the categorisation of AI systems within the EU AI Act. Chapter 3 will explore recent papers discussing research in VR as well as recent ethical discussions regarding AI. Chapter 4 will delve into a proposed architecture for tackling ethical issues with a core focus on the EU AI Act. Chapter 5 will discuss the compliance considerations for the Act, with implementation challenges being examined in chapter 6. Chapter 7 will serve as a conclusion and will discuss recommended future work.

2 Background Research

2.1 Ethical Challenges of AI in VR

The convergence of AI and VR presents a complex landscape of ethical considerations that require careful attention. VR, with its inherent capacity to create immersive and highly realistic experiences, amplifies the potential impact of AI on human behaviour, privacy, and well-being. This chapter explores these ethical challenges, examining the intricacies arising from this technology.

AI in VR: Blurring the Lines Between Reality and Illusion
VR's defining characteristic is its ability to transport users into digitally constructed realities, blurring the boundaries between the real and the virtual. This blurring raises concerns about the potential for AI to manipulate user perceptions and influence their actions.

For example, research has shown that VR can cause negative physical effects (Such as dizziness, falling down, or tripping over equipment while immersed in a scenario) [8] as well as psychological effects (Such as triggering post-traumatic stress disorder, desensitization to violence, and decreased empathy) [8].

The Proteus Effect [9], a phenomenon studied in VR research, further highlights these concerns. Studies have shown that user behaviour in virtual environments can be influenced by their avatars' characteristics [8, 9]. For instance, users assigned taller avatars exhibited greater assertiveness in negotiations compared to those with shorter avatars [9].

This effect, while fascinating, underscores the potential for AI-powered VR experiences to shape user behaviour in ways that may not be immediately apparent or consciously controlled. This raises crucial questions about potential unintended consequences in the real world, particularly when AI is employed to enhance realism and personalise experiences [8, 9].

These ethical challenges necessitate a thorough examination of the potential risks associated with AI in VR and the development of appropriate safeguards to mitigate these risks.

Amplifying Existing Ethical Concerns

The immersive nature of VR has the potential to magnify existing ethical issues associated with AI. Concerns about bias in algorithms, a significant problem in AI development, are made even more critical in VR. If algorithms are biased, they can lead to discriminatory outcomes and perpetuate existing societal inequalities in virtual environments. For example, an AI-powered VR experience intended for job training might inadvertently disadvantage certain user groups if the algorithms used reflect biases found in the training data. Data privacy and security concerns are also heightened in AI-powered VR [8]. The comprehensive data gathered by VR systems, including user movements, biometric information, and interactions within the virtual environment, pose privacy risks if not managed responsibly. The potential for misuse or unauthorised access to such data raises significant ethical concerns. Developers need to consider these issues and strive to create ethical AI systems [7]. The EU AI Act, which came into force in 2024, includes penalties for companies that do not comply with regulations on AI systems, such as those that could perpetuate societal inequalities [7].

Ethical Considerations in VR Experiences Designed for Empathy

VR has demonstrated the potential to cultivate empathy by allowing users to experience situations from diverse viewpoints. However, ethical considerations must be addressed when designing VR experiences for empathy. One risk is the creation of a false sense of agency or the oversimplification of complex situations, which could lead to misinterpretations of the intended message. For instance, a VR experience designed to promote empathy for individuals with autism spectrum disorder could inadvertently reinforce stereotypes or misrepresent the lived experiences of autistic individuals if it is not carefully designed and evaluated.

One study, 'Walking in Small Shoes', used VR to help teachers understand what it is like to be a child with autism [3, 10]. The creators of the study found that although it was successful in raising awareness, the experience could have given the wrong impression that autism is simply about controlling external stimuli. This shows how a short VR experience can lead to an oversimplification of a complex condition [3, 10]. The study also found that when parents of children with autism tried the VR experience, they found it traumatic because it amplified negative aspects of the school environment [3, 10]. This highlights the importance of considering how a VR experience might be interpreted by different audiences. Designers need to be aware of the potential for VR experiences to have unintended consequences and should take steps to mitigate these risks, for example by providing support to users before and after the experience and ensuring that the content is accurate and nuanced [3, 10].

The Importance of Responsible Design and Regulation

As AI and VR technologies continue to evolve, it is vital to prioritise responsible design and development practices. Developers have a responsibility to consider the ethical implications of their creations and take steps to safeguard user privacy and ensure fairness and inclusivity within AI systems [6, 7]. It is also crucial

to mitigate potential risks of manipulation [3]. For example, developers of AI-powered VR experiences for job training need to ensure that their algorithms do not inadvertently disadvantage certain groups of users. Similarly, the collection of sensitive data through VR systems needs to be handled responsibly to protect user privacy. Regulatory frameworks, like the EU AI Act, are essential for setting standards and encouraging ethical development in this rapidly changing field.

2.2 The EU AI Act: A Framework for Regulation

The EU AI Act represents a significant step in regulating the use of artificial intelligence. It lays out a risk-based framework designed to protect EU citizens from potential harm while fostering innovation in the field [6, 7]. This chapter analyses the key provisions of the Act, highlighting its relevance to the ethical challenges of AI in VR.

The EU AI Act is the first-ever established comprehensive legal framework on AI worldwide. This framework is essential for addressing the ethical aspects of AI in VR, ensuring appropriate safeguards are implemented according to the technology's potential impact. The Act employs a risk-based methodology, classifying AI systems by their potential threat to fundamental rights and user safety [6, 7]. This categorization is vital in navigating the ethical complexities of AI within VR. The Act defines various roles within the AI ecosystem: providers, users, importers, distributors, and manufacturers, holding all parties involved accountable. It also has implications for those operating outside the EU, for instance Switzerland, if their AI system's output is intended for use within the EU [7].

This risk-based approach ensures that AI systems are subject to varying levels of scrutiny and regulation, with stricter requirements for those deemed high-risk. For example, high-risk systems must meet specific requirements and undergo a conformity assessment before being introduced to the market [6, 7]. They must also be registered in an EU database [7]. Furthermore, the Act prohibits AI systems that present an unacceptable risk, including social scoring systems and those that exploit vulnerabilities of particular groups. The Act's emphasis on a risk-based approach allows for tailored regulations, encouraging innovation while safeguarding against potential harms.

The penalties for non-compliance are substantial, serving as a deterrent against unethical practices. By establishing clear guidelines and consequences for unethical practices, the EU AI Act aims to promote the responsible development and use of AI, such as within VR applications. It is worth noting that the EU AI Act does not aim to serve as the sole regulation of AI within Europe, but rather as a guideline for all European countries to develop their own laws regarding AI usage.

Unacceptable Risk: Prohibited AI Systems

The Act prohibits AI systems deemed to pose an unacceptable risk to individuals' safety or fundamental rights. These prohibited systems include those deemed a threat to people, such as:

- AI systems that employ subliminal techniques to manipulate individuals, exploiting vulnerabilities of specific groups, such as for instance programs that encourage children to perform dangerous behaviour [6].
- Social scoring systems that evaluate and rank individuals based on their social behaviour or characteristics [6].
- Biometric identification and categorisation of people [6].
- Real-time remote biometric identification, such as facial recognition [6].

Certain exceptions may be permitted such as for law enforcement purposes [6]. Remote biometric identification systems operating in real-time may be authorised in a restricted range of serious cases [6]. In contrast, systems that perform biometric identification, following a substantial delay, may be utilised to prosecute serious offences but only after obtaining court approval [6]. These prohibitions reflect the EU's commitment to protecting fundamental rights and freedoms, ensuring that AI is not used in ways that could undermine human autonomy and dignity.

High-Risk AI Systems: Stringent Requirements

The Act defines high-risk AI systems as those that could potentially have a negative impact on safety or fundamental rights. These systems face stricter regulations to guarantee responsible development and deployment. Some of the key requirements for high-risk systems include:

- Conformity assessments: Before a high-risk AI system can be launched in the market, it must undergo a rigorous conformity assessment to ensure it aligns with the Act's safety and ethical standards. This assessment involves evaluating the system's compliance with the Act's requirements [7].
- Risk management systems: Developers of high-risk AI systems must implement comprehensive risk management systems. This includes identifying potential risks, implementing mitigation strategies, and establishing procedures for ongoing monitoring and evaluation [7].
- Data governance: Recognising the crucial role of data quality in AI, the Act stresses the importance of using high-quality data for training, testing, and validating AI systems. This entails ensuring data sets are relevant, representative, and free from biases, which could lead to unfair or discriminatory outcomes [6].
- Human oversight: To prevent potential harm, high-risk AI systems necessitate human oversight. This means that there needs to be human accountability for the system's decisions and the ability for humans to intervene if the AI system is making decisions that could lead to negative consequences [7].

In addition to the aforementioned, all AI systems are split into two categories [6], those that are used in products falling under the EU's product safety legislation (Such as toys, aviation, cars, medical devices and lifts), and those that fall into specific areas such as

- Management and operation of critical infrastructure [6]: Such as power grids, transportation systems, and water supply systems.
- Education and vocational training [6]: AI systems used in educational and vocational training can have a significant impact on individuals' opportunities and future prospects.
- Employment, worker management, and access to self-employment [6]: Such as AI systems used in hiring, performance evaluations, and other aspects of employment.
- Access to and enjoyment of essential private services and public services and benefits [6].
- Law enforcement [6]: The use of AI in law enforcement raises critical concerns regarding fundamental rights, including the right to a fair trial and privacy.
- Migration, asylum, and border control management [6]: AI systems are being employed in sensitive areas like migration and border control, which directly impact individuals' freedom of movement and right to seek asylum.
- Assistance in legal interpretation and application of the law [6]: While AI can potentially support legal professionals, the Act classifies AI systems assisting in legal interpretation and application of the law as high-risk.

AI that follow these specific requirements are required to be registered in an EU database [6]. These examples illustrate the Act's focus on regulating AI applications that have the potential for significant societal impact, ensuring these systems are developed and deployed responsibly.

Transparency Requirements for Minimal Risk AI Systems

The EU AI Act recognises that not all AI systems pose the same level of risk. While some applications require stringent oversight due to their potential impact on fundamental rights and safety, others carry minimal risk. AI systems deemed to pose limited or minimal risk are subject to transparency requirements, where users must be informed when interacting with an AI system, such as in the case of Generative AI [6] and Deepfakes [6].

Generative AI refers to AI systems that can create new content, such as text, images, audio, and video. Examples include:

- ChatGPT [11], a large language model that can generate human-quality text in response to prompts. It can be used for various tasks, including writing different types of creative content, translating languages, and answering questions in an informative way.
- DALL-E 2 [12], another model from OpenAI, can create realistic images and art from natural language descriptions. For example by typing "a cat riding a unicorn on the moon", the AI generate a detailed image depicting exactly that.
- Jukebox [13], also from OpenAI, can generate music in different styles, including singing in the style of various artists.

Deepfakes, on the other hand, specifically involve AI-generated or manipulated images, audio, or video files. These manipulations can create incredibly

realistic, but fabricated, depictions of individuals saying or doing things they never actually did. The primary concern with Deepfakes is their potential to spread misinformation, damage reputations, and erode trust in digital content [14, 15].

The EU AI Act addresses concerns surrounding generative AI and Deepfakes by mandating transparency requirements such as:

- Disclosing that the content was generated by AI [6].
- Designing models to prevent the generation of illegal content [6].
- Publishing summaries of copyrighted data used for training [6].

These transparency requirements aim to foster trust in AI systems by ensuring users are informed about the nature of their interactions with AI-powered technologies.

Supporting Innovation While Ensuring Ethical Development
The EU AI Act aims to support innovation while ensuring ethical development by offering opportunities for startups and small and medium-sized enterprises to develop and train AI models before their release [6]. The Act requires national authorities to provide testing environments that simulate real-world conditions, enabling companies to assess the performance and safety of their AI systems before deployment [6].

3 Literature Review

3.1 Recent Research in VR and AI

Recent research highlights the ethical challenges surrounding the integration of artificial intelligence (AI) and virtual reality (VR). A significant concern is the potential for VR experiences to induce psychological harm, such as Depersonalization / De-realization Disorder [16]. This disorder can blur the lines between reality and virtuality, potentially impacting users' well-being.

Several studies have explored the ethical implications of using VR for specific purposes. For instance, research on immersive VR experiences for children has raised concerns about the acquisition of false memories [4, 10]. Another study examined the impact of VR on young adults, noting that it can increase physiological arousal and aggressive thoughts, particularly in interactive VR scenarios compared to observational ones [4]. This raises questions about the responsibility of VR designers in mitigating potential negative psychological impacts.

Ethical considerations also extend to the design process of VR applications. Designers are urged to integrate ethical analysis throughout the development stages [4]. This involves anticipating potential harms, such as misuse or unintended consequences, and implementing safeguards to protect users [4]. For instance, a study investigating the use of VR in education emphasized the need for continual research and assessment to define and apply ethical guidelines for VR use in educational settings [8].

3.2 Transparency in Design

Another crucial aspect is the transparency of AI systems in VR environments. Users should be clearly informed about the artificial nature of content, particularly in cases where AI is used to generate or manipulate images, audio, or video [4, 6, 17]. This is especially relevant in the context of "deepfakes," where AI-generated content can convincingly mimic real individuals or events. Ensuring transparency helps maintain trust and prevents potential deception.

Researchers also advocate for a "student focus" approach when designing VR for educational contexts [8]. This emphasizes the importance of considering students' developmental needs and potential vulnerabilities, including their susceptibility to emotional distress or privacy violations [8]. This approach promotes a user-centric design philosophy that prioritizes the well-being and safety of students in VR learning environments.

AI systems, particularly those using machine learning, are often described as "black boxes" due to the difficulty in understanding their decision-making processes [18]. This lack of transparency makes it challenging to understand how an AI arrives at a particular output, which creates issues for human oversight and guidance [18]. The opacity of AI systems, especially complex models like deep neural networks, makes it difficult to interpret their rationale, creating problems with accountability and trust. This is a significant ethical concern as it hinders the ability to identify and correct errors, biases, or discriminatory outcomes, and reduces the ability of humans to monitor the AI [18, 19]. Transparency issues also extend beyond the algorithms themselves to the development and deployment processes of the AI systems.

3.3 AI Respecting User Privacy

While some AI systems operate by collecting and processing data on a centralised server, posing potential privacy risks, other AI systems like federated learning prioritise data protection, while still maintaining high-performance. Federated learning enables AI models to train on data distributed across multiple devices without directly sharing the raw data. This approach ensures user privacy, a particularly important consideration in sensitive contexts like education.

A recent study titled "Beyond the Maze: How AI Personalizes Learning and Drives Engagement in Educational Games" [20] utilised this system, whereby by incorporating Federated Learning not only did the game's performance improve as not everything was running on a centralised server, but only necessary data within the VR environment was returned to the centralised system, providing an additional barrier of safety for user data.

4 Proposed Architecture

This chapter proposes an architecture for developing ethical AI systems in VR environments, specifically addressing the challenges outlined by the EU AI Act.

The architecture aims to serve as a guiding framework for developers, enabling the creation of VR experiences that are not only immersive but also ethically responsible. By focusing on ethical considerations, this framework seeks to ensure compliance with evolving regulations.

4.1 How The Proposed Architecture Was Designed

The proposed architecture was designed to address the ethical challenges of AI-powered Virtual Reality (VR) while ensuring compliance with the EU AI Act. It was built on principles of transparency, inclusivity, and user-centric design, informed by research into the psychological and ethical risks of VR, such as data privacy concerns, biases in AI, and the potential for psychological harm from immersive environments. The EU AI Act's risk-based regulatory framework further shaped the architecture's structure and functionality.

A modular design approach was chosen to ensure flexibility and adaptability. The architecture is divided into four key modules—VR Environment, AI Agent, User Interface and Interaction, and Ethics Monitoring and Evaluation—allowing each component to be developed and assessed independently. This approach also facilitates updates as technology and ethical standards evolve, ensuring long-term compliance and innovation.

The architecture integrates safeguards such as transparency features, risk management protocols, and robust data privacy measures like federated learning. Accessibility and inclusivity were prioritised, with features such as adjustable immersion levels to mitigate psychological distress and mechanisms to moderate harmful content. Additionally, real-time ethical monitoring and user feedback channels were embedded to proactively detect and address ethical concerns during deployment.

This framework balances the need for immersive and innovative VR experiences with the ethical obligations outlined in the EU AI Act. By aligning with regulatory requirements and addressing potential risks, the architecture serves as a robust foundation for creating responsible AI-driven VR systems.

4.2 VR Environment Module

The VR Environment Module encompasses the virtual world in which users interact, including its simulated environments, objects, and characters. Ethical considerations at this stage are crucial, as the design choices significantly influence user experiences and outcomes.

One key aspect is balancing realism and immersion. While creating realistic VR experiences can enhance engagement, excessive realism may lead to unintended negative consequences, such as depersonalisation or psychological distress. Developers must carefully evaluate the level of immersion to mitigate these potential adverse effects [4].

Another important consideration is accessibility and inclusivity. VR environments should be designed to accommodate a diverse range of users, accounting

for varying physical and cognitive abilities as well as cultural sensitivities. By prioritising inclusivity, developers can ensure that VR experiences are equitable and welcoming to all users.

Lastly, content moderation is an essential component of this module. Mechanisms should be implemented to detect and address harmful content, such as harassment, hate speech, or other inappropriate behaviour within the VR environment. By proactively managing content, developers can create safer and more respectful virtual spaces [6].

4.3 AI Agent Module

The AI Agent Module focuses on the design and functionality of AI-powered entities within the VR environment. These agents play a critical role in shaping user interactions and overall experiences. Ethical considerations for this module are multifaceted and must align with the principles outlined in the EU AI Act.

Transparency and explainability are essential for building user trust. AI agents should behave in ways that are understandable to users, with clear explanations of their decision-making processes and potential biases [18]. This transparency is particularly important for high-risk systems, as emphasised by the EU AI Act [17].

Fairness and non-discrimination are equally critical. AI agents must be designed to treat all users equitably, avoiding any discriminatory outcomes based on sensitive attributes such as race, gender, or disability. By prioritising fairness, developers can prevent bias and promote ethical interactions within the virtual environment [6, 17].

Additionally, user autonomy and control should be upheld. Users should have the ability to manage their interactions with AI agents, including making informed decisions about data sharing and personalisation. Empowering users with control over their experience fosters a sense of agency and respect for individual preferences [18].

4.4 User Interface and Interaction Module

The User Interface and Interaction Module centres on how users engage with the VR environment and its AI components. Ethical design in this module is crucial for ensuring a safe and user-friendly experience.

One primary consideration is informed consent and data privacy. Developers must provide users with clear and concise information about data collection practices, ensuring they can make informed decisions about their participation [4]. This aligns with the EU AI Act's focus on transparency and data governance [19].

Safety and well-being are also important. Interaction designs should minimise risks to users, including physical discomfort or psychological distress. Developers should consider potential triggers, such as anxiety-inducing scenarios or motion sickness, and implement safeguards to mitigate these effects [4].

Finally, responsible use guidance should be integrated into the interface. Users should receive clear instructions on ethical behaviour within the VR environment, encouraging respectful interactions with both other users and AI agents. This guidance helps maintain a positive and constructive virtual community.

4.5 Ethics Monitoring and Evaluation Module

The Ethics Monitoring and Evaluation Module provides a continuous oversight mechanism to ensure the ethical performance of the VR system. This module plays a pivotal role in identifying, addressing, and preventing ethical concerns throughout the lifecycle of the VR experience.

One critical aspect of this module is real-time ethical issue detection. Systems should be equipped to identify potential ethical problems as they occur during user interactions [18], allowing for timely interventions and resolutions.

Data logging and analysis are also essential for identifying patterns and trends related to ethical issues. By collecting and analysing user data, developers can gain insights into recurring challenges and implement targeted improvements to address them effectively.

User feedback mechanisms should be established to empower users to report ethical concerns and provide suggestions for system enhancements. This two-way communication fosters transparency and collaboration between developers and users, contributing to the system's ethical evolution.

Regular ethical audits should also be conducted to evaluate the system's performance against established guidelines and the requirements of the EU AI Act, in addition to local laws. Periodic assessments ensure that the VR system remains aligned with ethical principles and regulatory standards over time.

4.6 System Diagram

:

4.7 Practical Applications of the Architecture

To illustrate the practical implementation of the proposed architecture, this section presents two concrete examples of AI-powered VR systems developed within this framework. These examples demonstrate how the architecture's modules can work together to create ethical and compliant VR experiences that align with the EU AI Act's requirements.

Medical Training Simulation

The first example involves a VR-based medical training system for surgical procedures that implements the four core modules of the architecture. The VR Environment Module creates an immersive operating theatre with anatomically

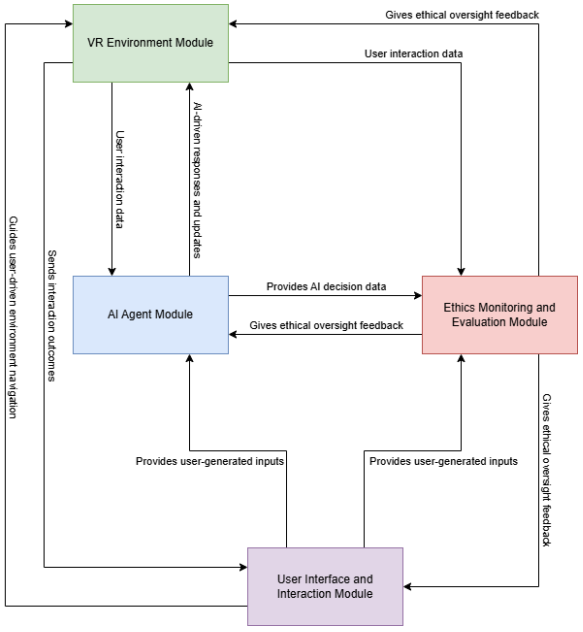

Fig. 1: System Diagram

accurate representations and precise physics-based interactions. This environment incorporates advanced tissue deformation models and realistic environmental conditions while ensuring accessibility for users with varying physical capabilities. Within this environment, the AI Agent Module implements an intelligent virtual instructor system that provides real-time guidance and performance analysis. The AI agents utilise specialised foundation models fine-tuned on validated medical datasets to ensure accuracy and safety in training scenarios. These agents employ federated learning techniques similar to those described by Silva et al. (2020) [21] to maintain patient data privacy while enabling continuous model improvements through distributed learning.

The User Interface and Interaction Module integrates sophisticated consent mechanisms and multi-modal interaction systems that prioritise user comfort and safety. This implementation pays particular attention to providing clear information about data collection practices and maintaining transparent communication about AI system involvement in the training process. The system incorporates haptic feedback mechanisms that enhance the realism of surgical tool manipulation while maintaining strict safety protocols. Meanwhile, the Ethics Monitoring and Evaluation Module continuously assesses both user performance and system behaviour. This module implements comprehensive logging systems that document all training sessions while maintaining user privacy. The system performs regular automated audits of AI decision-making processes to detect and mitigate potential biases in performance evaluation. This implementation

demonstrates full compliance with the high-risk system requirements under the EU AI Act, particularly concerning Articles that refer to data quality and human oversight.

Educational Virtual Training Simulator

The second example presents an AI-enhanced teacher training simulator, categorised as a limited-risk system under the EU AI Act. The VR Environment Module creates a highly detailed virtual classroom that simulates diverse and challenging educational scenarios. The environment replicates authentic classroom dynamics, including students with varying behavioural patterns, learning difficulties, and emotional states, enabling novice teachers to gain experience in managing complex classroom situations. The AI Agent Module implements sophisticated student avatars that demonstrate realistic classroom behaviours and interactions. These AI agents utilise advanced behavioural modelling systems based on documented classroom scenarios, operating within carefully defined ethical boundaries to avoid stereotyping or bias. Each virtual student exhibits unique personality traits, learning styles, and behavioural patterns, creating dynamic classroom situations that respond naturally to the teacher's pedagogical approaches and interventions. The system's sophisticated dialogue management ensures that virtual students' responses remain contextually appropriate while presenting authentic classroom challenges, such as disruptive behaviour, learning difficulties, or emotional distress.

The User Interface and Interaction Module facilitates natural teaching interactions, allowing trainee teachers to practise classroom management techniques, pedagogical strategies, and crisis intervention methods. The system provides immediate feedback on teaching approaches while maintaining transparent communication about the AI-driven nature of student responses. This implementation particularly emphasises the development of empathy and understanding for diverse student needs, aligning with the EU AI Act's requirements for transparency while fostering inclusive educational practices. Meanwhile, the Ethics Monitoring and Evaluation Module employs continuous assessment protocols that evaluate both teaching strategies and system behaviour. The system utilises a federated learning approach similar to that described in Checker et al. (2024) [22], enabling personalised feedback while protecting sensitive interaction data. This module records detailed analyses of teacher-student interactions, paying particular attention to the ethical handling of simulated behavioural challenges and emotional situations. This example demonstrates how the proposed architecture can be adapted to create meaningful training experiences while prioritising ethical considerations and user privacy.

5 EU AI Act Compliance Considerations

This chapter provides a detailed overview of the key considerations for ensuring compliance with the EU AI Act when developing AI-powered VR experiences. As previously mentioned, the Act adopts a risk-based approach to categorising

AI systems, setting requirements that vary according to the assessed level of risk. Understanding and adhering to these requirements is essential for creating VR experiences that are both innovative and compliant.

One of the foundational steps is identifying the risk category of the AI system within the VR application. The EU AI Act classifies AI systems into four categories: unacceptable risk, high risk, limited risk, and minimal risk. Each category carries its own set of implications and regulatory expectations.

5.1 Dealing with Higher-Risk AI

To ensure compliance and avoid creating AI systems that are banned under the EU AI Act, developers must take proactive measures to mitigate risks. This includes refraining from implementing subliminal techniques designed to manipulate user behaviour or deploying systems that exploit the vulnerabilities of specific groups. Additionally, systems performing social scoring are prohibited under the Act.

Transparency, fairness, and respect for fundamental rights are paramount for avoiding prohibitions. Developers should strive to create AI systems that empower users while safeguarding their rights and autonomy. For instance, biometric identification systems, which typically fall under higher-risk categories, may only be utilised under strict conditions. The EU AI Act allows such systems in exceptional cases, such as delayed identification for prosecuting serious crimes, and only with court approval [6].

High-risk AI systems require adherence to a conformity assessment process mandated by the EU AI Act. This process includes rigorous data quality control, risk management protocols, comprehensive technical documentation, and mechanisms for human oversight [19] so as to allow for human intervention or review of AI-driven decisions to prevent and mitigate potential harm. In VR applications, this might involve implementing processes that allow human operators to intervene when necessary, ensuring user safety and ethical alignment. High-risk AI falling under certain categories are also required to be registered in an EU database [6].

Developers must implement structured processes to meet these requirements and ensure compliance. Furthermore, the Act grants individuals the right to file complaints with designated national authorities, underscoring the importance of accountability. In addition to this, all High-risk systems must also uphold the systems mentioned in the next section.

5.2 Dealing with Lower-Risk AI

For AI systems classified as limited or minimal risk, the EU AI Act emphasises the need for transparency and ethical responsibility. Transparency requirements are particularly critical, ensuring users are fully informed and able to engage with AI systems responsibly.

Clear communication is a fundamental aspect of transparency. Users should be made aware that they are interacting with an AI system [7], whether through

explicit disclosures or user-friendly interface design. Explainability is also essential, as it enables users to understand the AI system's decision-making processes and the rationale behind its outputs [17].

Data labelling is another important consideration. AI-generated content, such as Deepfakes, should be clearly labelled [6] to prevent misinformation and promote trust. Robust data governance frameworks are essential to ensure that training data is relevant, representative, and free from bias [19]. Meeting these requirements not only enhances compliance but also fosters fairness and accuracy in AI systems.

Post-market monitoring is another vital requirement under the EU AI Act. Developers must establish mechanisms for ongoing evaluation of AI systems after deployment. This includes collecting user feedback, monitoring for unexpected behaviours, and implementing updates to address any issues that arise. Such monitoring ensures the continued compliance and ethical operation of the AI system over its lifecycle.

Navigating these compliance considerations requires a multidisciplinary approach. Collaboration between developers, legal experts, ethicists, and user experience designers is crucial for embedding ethical and regulatory requirements into the development process. By prioritising these considerations, developers can create AI-powered VR experiences that are not only cutting-edge but also responsible, trustworthy, and fully compliant with the EU AI Act.

6 Implementation Challenges

The development and deployment of ethical AI-powered VR systems within the framework of the EU AI Act present several significant implementation challenges. These challenges span technical, organisational, and regulatory domains, underscoring the need for a multidisciplinary approach to overcome them effectively.

6.1 Technical Complexity and Integration Issues

Integrating AI technologies into VR environments involves addressing complex technical requirements. VR systems demand high computational power, precise synchronisation of hardware and software, and advanced AI algorithms for realistic interactions [1, 4]. Achieving compliance with the EU AI Act adds another layer of complexity, requiring systems to incorporate transparency, fairness, and human oversight mechanisms. Developers often face difficulties in balancing these requirements while maintaining system performance and user experience. Additionally, ensuring that AI algorithms are free from biases, representative of diverse user groups, and aligned with ethical principles requires sophisticated data governance frameworks and robust validation processes [19].

6.2 Cost and Resource Constraints

Ethical and regulatory compliance can increase the cost of developing AI-powered VR systems, since experts would need to be consulted, over developer intuitions which may not be as reliable [4, 8]. The conformity assessment process mandated for high-risk systems, including documentation, data quality controls, and risk management protocols, involves additional resource investment. Small and medium-sized enterprises (SMEs), in particular, may struggle to allocate the financial and human resources needed to meet these stringent requirements. Furthermore, the need for ongoing post-market monitoring and regular ethical audits adds to the operational costs, posing challenges for organisations with limited budgets.

6.3 Regulatory Ambiguities and Evolving Standards

While the EU AI Act provides a comprehensive framework, some aspects of its implementation remain open to interpretation. Ambiguities in defining risk categories or specifying technical requirements for conformity assessments can create uncertainty for developers. Moreover, as AI and VR technologies evolve rapidly, regulatory standards may also change, requiring developers to continuously adapt their systems to remain compliant. Keeping up with these evolving standards can be a daunting task, particularly for organisations without dedicated legal or compliance teams.

6.4 User Diversity and Ethical Variability

VR environments cater to a highly diverse user base, encompassing individuals with different cultural, cognitive, and physical characteristics. Designing systems that are inclusive and accessible to all users while adhering to ethical principles is a challenging endeavour [8]. Developers must navigate varying ethical norms, expectations and rights across different regions and demographics [18], balancing global compliance requirements with local sensitivities. For instance, ensuring that AI-powered VR systems are culturally sensitive and do not inadvertently marginalise or disadvantage specific user groups requires careful consideration and extensive testing [18].

6.5 Data Privacy and Security Risks

VR systems inherently collect extensive data, including biometric information, behavioural patterns, and interaction logs [16]. Safeguarding this data against breaches or unauthorised access is paramount to maintaining user trust and compliance with the EU AI Act. Implementing advanced encryption methods, anonymisation techniques, and secure data storage systems is essential but can be technically and financially demanding. Additionally, developers must ensure that data collection practices align with the principles of informed consent and transparency, providing users with clear information about how their data will be used.

6.6 Collaboration and Multidisciplinary Coordination

Effective implementation of ethical AI-powered VR systems requires collaboration among diverse stakeholders, including developers, ethicists, legal experts, and user experience designers. Coordinating these efforts to ensure alignment with ethical and regulatory standards can be challenging [4], especially when stakeholders have differing priorities or levels of expertise. Establishing clear communication channels and fostering a shared understanding of compliance goals are crucial for overcoming these coordination challenges.

7 Conclusion

The combination of artificial intelligence and virtual reality offers exciting opportunities to transform how people interact with technology. From immersive training to new ways of learning, AI-powered VR has great potential. However, this progress comes with ethical and regulatory challenges, especially under the EU AI Act, which requires developers to carefully design systems that are fair, transparent, and respectful of users' rights.

This document has highlighted the importance of following the EU AI Act's risk-based framework, which categorises AI systems into different levels of risk. By addressing ethical considerations like transparency, data privacy, and fairness through a modular system design, developers can ensure compliance while fostering innovation. These modules help to tackle challenges in a systematic way, enabling the creation of safe and responsible VR environments.

Moving forward, the future of AI-powered VR depends on balancing innovation with responsibility. Developers must continue to focus on user well-being, inclusivity, and ethical integrity. With the EU AI Act providing guidance, there is a clear path to building systems that earn user trust, enhance creativity, and bring meaningful benefits to society.

7.1 Future Work

The intersection of Artificial Intelligence (AI) and Virtual Reality (VR) demands ongoing exploration to address evolving ethical, technical, and regulatory challenges. Personalisation in VR, driven by AI, offers immense potential to enhance user experiences but requires robust frameworks to prevent bias and manipulation. These frameworks should prioritise fairness and transparency to ensure equitable treatment while fostering trust.

Long-term studies are essential to evaluate the psychological and behavioural effects of prolonged VR use across diverse demographic groups. Such research can help identify risks like dependency, depersonalisation, or behavioural changes and inform strategies to create safer, more user-centric VR environments. These insights will be critical for developers aiming to balance immersion with user well-being.

Inclusivity must remain a key focus in future developments, with an emphasis on adaptive technologies that accommodate diverse physical, cognitive,

and cultural characteristics. Ensuring accessibility and cultural sensitivity will make VR systems equitable and welcoming for all users, expanding their societal impact.

Real-time ethical monitoring systems present an opportunity to enhance user safety by dynamically detecting and addressing issues like inappropriate content or unsafe interactions. By integrating such systems, developers can improve trust and accountability in VR environments. Additionally, harmonising the EU AI Act with global regulations would reduce compliance burdens and establish consistent ethical standards worldwide.

Innovations such as federated learning models can address data privacy concerns by processing information locally on user devices while maintaining system performance. Ethical handling of biometric data, including secure storage and informed consent, is equally critical. By addressing these areas, future work can ensure that the integration of AI and VR remains both responsible and impactful.

References

1. M. Luck and R. Aylett, 'Applying artificial intelligence to virtual reality: Intelligent virtual environments,' Applied Artificial Intelligence, vol. 14, no. 1, pp. 3–32, Jan. 2000, doi: https://doi.org/10.1080/088395100117142
2. D. Sciberras, 'VR comm: communication in VR using LLM', bachelor Thesis, University of Malta, 2024. Available: https://www.um.edu.mt/library/oar/handle/123456789/127918
3. V. Camilleri, 'Chapter 11 - Perspectives on the ethics of a VR-based empathy experience for educators', in Ethics in Online AI-based Systems, S. Caballé, J. Casas-Roma, and J. Conesa, Eds., in Intelligent Data-Centric Systems. , Academic Press, 2024, pp. 211–228. doi: 10.1016/B978-0-443-18851-0.00020-2.
4. B. Kenwright, 'Virtual Reality: Ethical Challenges and Dangers [Opinion]', IEEE Technology and Society Magazine, vol. 37, no. 4, pp. 20–25, Dec. 2018, doi: 10.1109/MTS.2018.2876104.
5. E. Yiannakopoulou, N. Nikiteas, D. Perrea, and C. Tsigris, 'Virtual reality simulators and training in laparoscopic surgery', International Journal of Surgery, vol. 13, pp. 60–64, Jan. 2015, doi: 10.1016/j.ijsu.2014.11.014.
6. 'EU AI Act: first regulation on artificial intelligence', Topics — European Parliament. Accessed: Dec. 15, 2024. [Online]. Available: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.
7. K. Meier and R. Spichiger 'The EU AI Act: What it means for your business'. Accessed: Dec. 15, 2024. [Online]. Available: https://www.ey.com/en_ch/insights/forensic-integrity-services/the-eu-ai-act-what-it-means-for-your-business.
8. P. Steele, C. Burleigh, M. Kroposki, M. Magabo, and L. Bailey, 'Ethical Considerations in Designing Virtual and Augmented Reality Products—Virtual and Augmented Reality Design With Students in Mind: Designers' Perceptions', Journal of Educational Technology Systems, vol. 49, no. 2, pp. 219–238, Dec. 2020, doi: 10.1177/0047239520933858.

9. N. Yee, J. Bailensonm and N. Ducheneaut 'The Proteus Effect Implications of Transformed Digital Self-Representation on Online and Offline Behavior', ResearchGate. Accessed: Jan. 08, 2025. [Online]. Available: https://www.researchgate.net/publication/228445790_The_Proteus_Effect_Implications_of_Transformed_Digital_Self-Representation_on_Online_and_Offline_Behavior.
10. V. Camilleri, M. Montebello, A. Dingli, and V. Briffa, 'Walking in small shoes: Investigating the power of VR on empathising with children's difficulties', in 2017 23rd International Conference on Virtual System & Multimedia (VSMM), Dublin: IEEE, Oct. 2017, pp. 1–6. doi: 10.1109/VSMM.2017.8346253.
11. OpenAI, "ChatGPT," ChatGPT, Nov. 30, 2022. Available: https://chatgpt.com/.
12. OpenAI, "DALL·E 2," Openai.com, Apr. 06, 2022. Available: https://openai.com/index/dall-e-2/.
13. OpenAI, "Jukebox," Openai.com, Apr. 30, 2020. Available: https://openai.com/index/jukebox/.
14. J. Kietzmann, L. W. Lee, I. P. McCarthy, and T. C. Kietzmann, 'Deepfakes: Trick or treat?', Business Horizons, vol. 63, no. 2, pp. 135–146, Mar. 2020, doi: 10.1016/j.bushor.2019.11.006.
15. I. Sample, 'What are deepfakes – and how can you spot them?', The Guardian, Jan. 13, 2020. Accessed: Jan. 12, 2025. [Online]. Available: https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them
16. J. S. Spiegel, 'The Ethics of Virtual Reality Technology: Social Hazards and Public Policy Recommendations', Science and Engineering Ethics, vol. 24, no. 5, pp. 1537–1550, Oct. 2018, doi: 10.1007/s11948-017-9979-y.
17. M. Veale and F. Z. Borgesius, 'Demystifying the Draft EU Artificial Intelligence Act', Computer Law Review International, vol. 22, no. 4, pp. 97–112, Aug. 2021, doi: 10.9785/cri-2021-220402.
18. C. Huang, Z. Zhang, B. Mao, and X. Yao, 'An Overview of Artificial Intelligence Ethics', IEEE Transactions on Artificial Intelligence, vol. 4, no. 4, pp. 799–819, Aug. 2023, doi: 10.1109/TAI.2022.3194503.
19. L. Edwards, "The EU AI Act: a summary of its significance and scope Expert explainer," 2022. Available: https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf
20. R. Y. Camilleri and V. Camilleri, 'Beyond the Maze: How AI Personalizes Learning and Drives Engagement in Educational Games', in Proceedings of the Future Technologies Conference (FTC) 2024, Volume 2, K. Arai, Ed., Cham: Springer Nature Switzerland, 2024, pp. 301–320. doi: 10.1007/978-3-031-73122-8_19.
21. S. Silva, A. Altmann, B. Gutman, and M. Lorenzi, 'Fed-BioMed: A General Open-Source Frontend Framework for Federated Learning in Healthcare', in Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning: Second MICCAI Workshop, DART 2020, and First MICCAI Workshop, DCL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4–8, 2020, Proceedings, Berlin, Heidelberg: Springer-Verlag, Oct. 2020, pp. 201–210. doi: 10.1007/978-3-030-60548-3_20.
22. S. Checker, N. Churamani, and H. Gunes, 'Federated Learning of Socially Appropriate Agent Behaviours in Simulated Home Environments', Mar. 12, 2024, arXiv: arXiv:2403.07586. doi: 10.48550/arXiv.2403.07586.