Trustworthy AI for Educational Metaverses

Faseela Abdullakutty\textsuperscript{1}, Adnan Qayyum\textsuperscript{2}, and Junaid Qadir\textsuperscript{3}

\textsuperscript{1}Robert Gordon University
\textsuperscript{2}Information Technology University
\textsuperscript{3}Qatar University

March 07, 2024
Trustworthy AI for Educational Metaverses

Faseela Abdullakutty 1, Adnan Qayyum 2, Junaid Qadir 3,*
1 Robert Gordon University, Aberdeen, UK
2 Information Technology University, Lahore, Pakistan
3 Qatar University, Doha, Qatar

Abstract—The metaverse—a 3D virtual universe is expected to significantly impact the education sector by making learning more accessible, personalized, and fun. The advancements in AI, blockchain, extended reality, big data, and cloud computing are the key enablers for the development of educational metaverses. The recent disruptions in AI, particularly, generative AI (GenAI) have transformed educational practices by generating human-like text, automating conversations, providing personalized learning experiences, and supporting students with disabilities. AI advancements together with immersive technologies hold immense potential to transform conventional education and learning by providing an interactive and immersive platform for seamless learning experiences. As GenAI advances, it is expected to generate more accurate and high-quality content, with future applications in the educational metaverse enhancing trustworthiness. This article contributes to background research on AI in education, a detailed study on the educational metaverse, a critical discussion on proactive measures to achieve trustworthy AI, and open research issues in the TAI in the educational metaverse context.

Index Terms—Trustworthy AI, TAI for Educational metaverse, Generative AI for Educational metaverse, Risks and Opportunities

I. INTRODUCTION

The metaverse—a virtual universe facilitating user interaction with digital objects in a 3D environment is expected to make significant strides in various sectors, particularly in education. The educational metaverse enhances immersive and interactive learning experiences, allowing users to engage with content in ways that mirror real-world scenarios [1]. Specifically, the educational metaverses transform the evaluation, mode, and educational environment by making learning fun, engaging, and more accessible. This transformation has gained momentum, especially due to the challenges posed by the COVID-19 pandemic, which has spurred an increased interest in the development of educational metaverses, providing an innovative solution to the limitations of traditional education during global crises. Moreover, the recent advancements in key technologies such as Artificial Intelligence (AI), blockchain, Extended Reality (XR), big data, and cloud computing play pivotal roles in empowering educational metaverses, contributing to their effectiveness and widespread adoption [2].

AI has played a transformative role in reshaping educational practices, leveraging capabilities such as generating human-like text and automating conversations [3]. These advancements enable personalized learning experiences and offer multifaceted support within educational settings. For instance, AI tools can effectively automate administrative tasks, providing educators with more time for instructional activities. Furthermore, AI’s capacity to engage in sophisticated data analysis allows for the identification of patterns and trends, empowering educators to refine instructional strategies and make informed and data-driven decisions. One notable impact of integrating AI technology in education is the provision of virtual tutoring and personalized support, extending educational opportunities beyond traditional classrooms. AI applications in education span a diverse range, encompassing student grading, assessments, retention prediction, classroom monitoring, intelligent tutoring, sentiment analysis, and recommender systems [4]. By automating such tasks, AI not only enhances student engagement and learning outcomes but also streamlines administrative processes, contributing to the overall efficacy of educational institutions.

In the educational metaverse, the importance of Trustworthy AI (TAI) is paramount in terms of providing students with accurate and reliable information [5]. TAI is an ethical approach to the development and use of AI systems, emphasizing transparency, accountability, responsibility, privacy rights protection, and security issues. Importantly, in the realm of AI-empowered education, TAI plays a pivotal role in enhancing learning experiences while promoting fairness and supporting personalized learning. TAI assists educators in identifying areas needing additional support, enabling early intervention and targeted instruction. Specifically, in the educational metaverses, trustworthiness should be characterized by competence, transparency, and fairness to prevent any misplacement of trust in AI applications. TAI aligns with educational standards, creating a secure learning environment and enhancing personalized learning. Additionally, it ensures data privacy, responsible handling of students’ data, and personalized recommendations [6]. Importantly, TAI enables accountability by holding the AI system accountable for its actions.

In the context of educational metaverse applications, the creative and flexible attributes of Generative AI (GenAI) technologies, such as ChatGPT, have attracted attention for their capability to enhance content production and distribution significantly [7]. This is achieved by generating diverse content tailored to address various educational requirements. Specifically, GenAI has the potential to revolutionize the educational metaverse by enabling immersive experiences through realistic game content, such as maps, monsters, and Non-Player Characters (NPCs). This leads to increased player engagement and organic interactions with in-game counterparts. As GenAI technology advances, it is expected to generate...
TABLE I

COMPARISON OF REVIEW PAPERS ON EDUCATION, TRUSTWORTHY AI, AND METAVERSE. LEGEND: GENERATIVE AI (GENAI), TAI (TAI), AND EXPLAINABLE AI (XAI).

<table>
<thead>
<tr>
<th>Authors</th>
<th>Year</th>
<th>Topics of discussions in educational context</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ahmad et al. [4]</td>
<td>2020</td>
<td>General, Metaverse, GenAI, Data Privacy &amp; Security</td>
</tr>
<tr>
<td>Holmes et al. [7]</td>
<td>2021</td>
<td>General Ed., Metaverse, GenAI, TAI, Ethical AI, XAI</td>
</tr>
<tr>
<td>Akgun [8]</td>
<td>2022</td>
<td>K-12, Metaverse, GenAI, TAI, XAI</td>
</tr>
<tr>
<td>Gupta et al. [1]</td>
<td>2023</td>
<td>Higher Ed., Metaverse, GenAI, TAI, XAI</td>
</tr>
<tr>
<td>Qian et al. [13]</td>
<td>2023</td>
<td>General Ed., Metaverse, GenAI, TAI, XAI</td>
</tr>
<tr>
<td>Nguyen et al. [14]</td>
<td>2023</td>
<td>General Ed., Metaverse, GenAI, TAI, XAI</td>
</tr>
</tbody>
</table>

more accurate and high-quality content including textual, speech, video, and avatar generation, etc. Future applications of GenAI include healthcare, education, and finance, as it becomes more intelligent and makes independent decisions. This opens up a horizon of innovative possibilities in various fields including education.

Existing Surveys: Within the research landscape of AI in education, various survey and review papers have focused on areas encompassing trustworthy AI, the educational metaverse, and GenAI. Nevertheless, to the best of our knowledge, there remains an absence of a comprehensive review addressing the realm of TAI in the educational metaverse, encompassing imminent challenges and future trajectories while incorporating GenAI. Consequently, this review extensively intertwines the dimensions of TAI with the context of a GenAI-powered educational metaverse. The article further delves into ethical AI, explainable AI, and foundational models within the purview of the trustworthy educational metaverse, exploring challenges, potential solutions, and prospective directions. To aid in understanding the unique contributions of our work, in Table I we provide a detailed comparison of our paper with existing similar survey and review articles.

Scope of this Paper: Figure 1 illustrates the areas that fall within the purview of the paper. This article examines recent developments in AI applied to education, AI within the metaverse, the influence of GenAI in the educational metaverse, and the concept of trustworthy AI, with a focus on its exploration in the educational metaverse context. Reviewing these topics in depth reveals a number of imminent challenges, particularly related to the use of GenAI within the educational metaverse. Moreover, the article discusses opportunities and potential future paths for establishing trustworthiness in an educational metaverse that is powered by GenAI.

Contributions of this Paper: The main contributions of this article are:

1) We present a comprehensive review of educational metaverses, encompassing an examination of its enabling technologies and a thorough analysis of the impact of TAI in the context of the educational metaverse.

2) We present an overview of various proactive measures to achieve TAI in educational metaverses.

3) We elaborate upon various open research issues in the realm of TAI in educational metaverses.

Review Methodology: This section explains the methodology used for reviewing and selecting articles to include in this review paper. The selection process was conducted following the approach shown in Figure 2. Articles were screened primarily based on their publication date, focusing on the years 2020 to 2023. The identification of relevant articles was facilitated using keywords, including “metaverse,” ”trustworthy AI,” “AI in the educational metaverse,” and ”generative AI for the educational metaverse.” Thereafter, the identified articles underwent a two-step filtering process. Initially, an examination of the abstracts and conclusions was conducted to narrow down the selection. Subsequently, a full-text examination of the remaining papers was carried out, and the final selection was made based on their relevance to the specified keywords. In total, this review considered 82 articles that have been selected from various publishers including IEEE Xplore, ACM, Science Direct, Springer, and MDPI.
II. BACKGROUND

In this section, we present an overview of related terminologies on which this paper is focused including immersive technologies, GenAI, foundation models, and TAI.

A. Extended Reality

Augmented Reality (AR) enhances user perception and interaction by overlaying digital data onto the real world, as shown in Figure 3 (a). It can be experienced using various devices like smart glasses, smartphones, and tablets, which use sensors and cameras to detect the real-world environment. AR can potentially revolutionize education by providing interactive and immersive learning experiences, promoting critical thinking, collaboration, and active learning.

Virtual Reality (VR) provides a simulated and immersive environment that could enhance teaching and learning. It can be realized using VR headsets and it enables users to immersively interact with simulated objects and navigate intricate scenarios. For educational purposes, VR can be leveraged to create interactive simulations of complex concepts, it can also be used for historical recreations, as well as for scientific explorations, providing an unparalleled engagement. This facilitates students’ profound comprehension of complex concepts, enhancing their learning, knowledge retention, and skill acquisition. By encouraging critical thinking, teamwork, and active learning through simulations and interactive experiences, VR can potentially redefine education, shaping a cohort of learners characterized by technological proficiency and knowledge-driven skills. Figure 3 (b) illustrates an example VR environment in education.

Mixed Reality (MR) seamlessly integrates both virtual and Augmented Reality (AR), establishing a harmonious fusion between the physical and virtual dimensions. It enables users to engage with virtual objects integrated into their physical surroundings, enhancing the overall experience and creating new possibilities for learning and engagement. MR can be employed in various fields, including education, gaming, healthcare, and design, offering unique opportunities for interactive and experiential learning. In healthcare, MR can be used in surgical training and patient education, allowing surgeons to practice procedures in a realistic virtual environment and patients to visualize medical information and treatments. Remote collaboration can be facilitated by allowing users in different locations to interact and collaborate in a shared virtual space. MR has numerous other applications. In marketing, MR creates engaging consumer experiences for brand interaction and product visualization. In gaming, it integrates virtual elements with the real world, enhancing interactivity. For design and architecture, MR enables real-time visualization and manipulation of models, improving decision-making and collaboration. In entertainment and media, MR can create immersive experiences like virtual concerts, interactive storytelling, and tours as in Figure 3 (c). Industrial training can replicate real-world scenarios, offering hands-on experience in performing complex tasks. This not only improves safety measures but also enhances overall efficiency and effectiveness.

B. AI in Education

The advancements in technology, including AI, have significantly altered many educational practices. AI serves as
AI literacy is increasingly pivotal in vocational education, mirroring a wider trend where AI and metaverse technologies are being integrated across various educational domains, including engineering, arts, and sciences, as indicated in studies by Ho et al. [17], Epstein et al. [18], and Chen et al. [19]. A key example of this trend is the transformation in healthcare education through AI, particularly in diagnostic radiology, as explored by Salastekar et al. [20]. This integration not only facilitates learning and enhances educational content quality but also standardizes experimental methods, supporting an open-science approach. Salastekar et al. underscore the positive reception of AI education among radiologists, stressing the importance of developing accurate AI algorithms and fostering collaboration between radiologists and AI developers. Additionally, AI models are enhancing clinical practice by assisting in tasks like drafting appeal letters, simplifying medical records, and aiding in complex diagnostic cases, as noted by Kung et al. [21], thereby improving the delivery of personalized and scalable healthcare.

In the context of the educational metaverse, TAI encompasses two dimensions: human-like trust and functionality trust. Human-like trust involves aligning AI algorithms with educational values, ethics, and cultural norms. This dimension ensures that AI technologies in educational settings reflect and respect diverse perspectives and ethical considerations. On the other hand, functionality trust focuses on the reliability, competency, expertise, and safety of AI technology in supporting educational activities. TAI can be examined at multiple levels, including individual trust, trust in educational institutions implementing AI, and trust based on personal experiences with AI in educational settings. The literature suggests that these dimensions are crucial for ensuring the ethical use of AI technology and fostering trust among educators, students, and other stakeholders in the educational ecosystem [23].

TAI is a concept that emphasizes transparency, interpretability, accountability, fairness, robustness, user-centricity, and ethical considerations. A brief description of these components is provided next.
• **Transparency**: Transparency in AI refers to the ability of AI systems to disclose information about their decision-making processes, promoting trust and accountability. This enables users to comprehend how AI models arrive at decisions, understand the factors influencing them, and assess the system’s fairness and reliability. Transparency also helps identify biases and discriminatory practices, ensuring they don’t disproportionately impact certain groups. It also helps detect potential errors or vulnerabilities, leading to improved system performance. Transparency also facilitates regulatory compliance and ethical considerations.

• **Fairness**: Fairness in AI refers to the absence of bias and discrimination in decisions and outcomes produced by AI systems, ensuring equal treatment for individuals or groups based on protected attributes. The purpose of fairness is to mitigate negative social ramifications from systematic unfairness, such as underprivileged groups experiencing disadvantage in hiring or criminal risk profiling. Fairness promotes social equity, prevents discrimination, and fosters trust in AI systems among stakeholders. TAI systems aim to provide equal opportunities and treatment to all individuals, regardless of their protected attributes, contributing to AI technology development and application.

• **Accountability**: Accountability in TAI involves stakeholders ensuring ethical, transparent, and human-centered development and use of AI systems. This involves considering potential impacts, disclosing design and operation information, and assessing compliance with ethical and legal standards. Accountability promotes fair, unbiased, and respectful AI systems, fostering trust and confidence in AI technologies. It also allows stakeholders to be held responsible for potential negative consequences or biases, leading to improvements in system design, operation, and governance.

• **Robustness**: Robustness in TAI refers to the system’s ability to perform reliably and accurately in various conditions, including adversarial attacks, noisy data, and changes in the input environment. It ensures AI systems can handle unexpected situations and maintain performance, even in the presence of uncertainties. Robust AI systems are less susceptible to adversarial attacks and can provide more accurate results, reducing the risk of incorrect or biased outcomes [24]. Achieving robustness requires techniques like data augmentation, model regularization, and adversarial training to improve system resilience and performance in real-world scenarios.

• **Ethical Considerations**: TAI involves integrating ethical principles and values into its design, development, and deployment. These considerations aim to align AI systems with human values, respect fundamental rights, and avoid harm to individuals or society. They address issues like fairness, transparency, privacy, accountability, and bias. TAI systems promote fairness, equality, and users’ privacy, and mitigate potential biases. They foster public trust and acceptance of AI technologies. Ethical guidelines and frameworks, developed by organizations like IEEE and the European Commission, provide guidance for incorporating ethical considerations into AI development.

• **User-centrality**: User-centrality in TAI focuses on the needs, preferences, and well-being of users, prioritizing their interests, values, and experiences. It involves considering factors like privacy, control, feedback, and trust in AI systems. By incorporating user-centrality, TAI systems aim to enhance user satisfaction, engagement, and acceptance. It also helps identify potential biases, discrimination, and negative impacts on underrepresented groups, ensuring fairness and inclusivity. User-centric design principles and methodologies, such as human-centered design, participatory design, and user testing, can be employed.

• **Interpretability**: Interpretability in TAI pertains to the capability to comprehend and elucidate how an AI system arrives at decisions or predictions. It provides transparency and insight into the internal mechanism of AI systems, enabling users and stakeholders to trust and validate AI outcomes. Interpretability helps identify potential biases, errors, or discriminatory outcomes, allowing for necessary adjustments or interventions. It promotes accountability by enabling users to comprehend the rationale behind AI decisions and hold the system responsible. Techniques like rule-based models, feature importance analysis, and model-agnostic methods can be employed to achieve interpretability in AI systems [25].

AI maintenance in educational settings is crucial for ensuring the trustworthiness of AI systems. It involves robustness assessment, risk identification, and fortifying models across the entire AI lifecycle. Continuous monitoring of deployed models helps detect and mitigate potential risks, enhancing the reliability and trustworthiness of AI systems. For instance, taking proactive measures such as active model scanning and probing helps identify potential failure modes, thereby ensuring the robustness of AI models. AI maintenance practices, including data preprocessing, model training, and performance validation, bridge the gap between model training and deployment, which can ensure the efficacy and reliability of AI systems in educational metaverses [26].

D. **Generative AI (GenAI)**

Generative models are ML techniques designed to produce realistic data samples by learning patterns from training datasets [27]. They learn the distribution of data points and generate new samples that resemble real data. Generative models excel in modeling high-dimensional probability distributions, proving particularly beneficial in intricate data domains. Two popular approaches for generative models are Generative Adversarial Networks (GAN) and Variational Autoencoders (VAE). The latest evolution in generative models comes in the form of AI foundation models. These models showcase their prowess in synthesizing contextually rich content across a wide range of domains, representing a significant advancement in the field of GenAI. Contemporary GenAI models use deep neural networks for NLP and image reconstruction. GenAI
also finds applications in generating art, music, and other creative content. For instance, advanced language models like GPT-4 and DALL-E demonstrate the capabilities of GenAI by generating human-like text and producing visuals from textual descriptions. Moreover, to ensure high-quality content, reinforcement learning can be employed to fine-tune generative models.

1) GenAI in Education: GenAI holds significant promise for transforming education by providing personalized and interactive learning experiences. Capitalizing on its capacity to produce varied content, including text, images, and videos, GenAI incorporates multi-modal learning strategies, enhancing students’ understanding and retention of educational material. GenAI can also generate educational materials like lectures, lesson plans, and assessments, saving time and effort for educators. These models can enhance student engagement and understanding, while also fostering critical thinking and problem-solving skills. However, ethical considerations must be addressed, such as potential misuse or misleading information. Educators must guide to help students navigate these ethical considerations and responsibly use generative models in their learning.

GenAI models, such as GANs, can significantly improve education by generating interactive content, such as virtual simulations and visualizations, and creating personalized learning experiences. GANs [27] are known for their ability to generate realistic data samples, which could potentially be applied in educational settings or virtual environments to create synthetic examples or enhance the realism of virtual experiences. GANs can improve educational research by generating synthetic data that increases sample size, enhances data quality, and reduces anomalous values. They can create larger and more representative samples for comprehensive analyses and interpretations [28]. GANs also protect information anonymity and ensure data security, encouraging the availability and exchange of datasets for research purposes. The integration of GANs in education aligns with the movement towards open science, facilitating data sharing and enabling researchers to test analytical and methodological techniques. Table I illustrates the role of GenAI in education and its advantages.

Applications such as ChatGPT, which uses GenAI to analyze synthetic data, can enhance engineering education by improving tasks such as drafting papers, working with synthetic data, and creating survey instruments. Additionally, they can assist with the analysis of text, the generation of tabular data, and the analysis of statistical data. However, they cannot check sense and knowledge, so human interpretation is necessary for the output to be valid. Using GenAI applications effectively requires the acquisition of skills and resources by engineers. Engineering education and research need transparency, value-creation, and understanding of GenAI systems. As a result of the lack of explainability and the haphazard stitching of sequences of linguistic forms, using GenAI is a cause for concern. To minimize potential harms and ensure the ethical use of GenAI, best practices should be shared, a framework should be created, and examples should be developed [9]. AI will have a varying impact on engineering education depending on how it is implemented and used, so it is imperative that universities carefully consider both the benefits and limitations of incorporating AI into their teaching and assessment practices [10].

GenAI tools can enhance creativity in art and design education by teaching prompt engineering techniques and encouraging iterative processes. GenAI tools when integrated into the curriculum, enable students to explore aesthetic choices, experiment with new techniques, and broaden their creativity through the generation of diverse visual outcomes. Moreover, AI tools can foster collaboration and peer learning, involving students in interpreting and building upon collaborative work. Ethical considerations, including discussions on copyright issues and potential job displacement, can provide students with a comprehensive awareness of AI-associated implications and responsibilities. Therefore, to ensure the successful integration of GenAI tools in educational landscapes, ongoing evaluation and refinement of the curriculum are crucial to staying abreast of the latest developments in AI and its applications in various domains [12].

2) Gen AI in Metaverses: GenAI in the metaverses can perform various tasks such as natural language generation, image generation, 3D model generation, interactive storytelling, art generation, video content generation, and personalized messaging [15]. One of the fascinating applications of GenAI in education metaverses is its application in the development of educational games. For instance, developers can employ GenAI techniques to generate player avatars and non-player characters (NPCs), to enhance the learning experience through gamification. It can personalize learning by adjusting difficulty levels based on proficiency, providing real-time feedback, and suggesting personalized study strategies.
Analyzing student data, GenAI can pinpoint knowledge gaps and generate customized remedial materials. GenAI integrated into NPCs and dialogues enables more organic and tailored interactions, with NPCs dynamically responding to player actions and expressing authentic emotions. This empowers players with greater control over the game and characters, the ability to customize the game’s appearance, comprehend character emotions, and optimize game difficulty, enhancing the overall player experience.

GenAI technologies, including ChatGPT and GPT-3, can enhance conversational interfaces with AI-generated characters, contributing to the dynamic and immersive nature of the metaverse. Image generation models like DALL-E and MidJourney can create visually stunning and diverse content, while 3D model generation technologies like Point-E and Lumirithmic can be used to create realistic virtual objects. These technologies use transformer architectures, autoregressive models, variational autoencoders, and GANs to perform tasks like natural language generation, chatbot responses, text completion, interactive storytelling, art generation, style transfer, landscape generation, video content generation, virtual actor creation, personalized video messaging, 3D modeling, shape completion, and enhanced object detection. The transformative power of GenAI lies in creating an interactive, immersive, and diverse virtual world [35].

### E. AI Foundation Models (AIFMs) and LLMs in Education

#### 1) An Overview of AIFMs: AIFMs are pre-trained models that excel in vision and language tasks, designed for real-world interactions like dialogue with humans or autonomous navigation [36]. They are trained using large datasets for multimodal, multitask, and generalist interaction. Foundation models are powerful tools for decision-making in various applications such as dialogue, autonomous driving, healthcare, education, and robotics. These models can be adapted to specific tasks by interacting with external entities and receiving feedback. However, challenges in foundation models include grounding them in practical applications and addressing issues like conditional generative modeling, optimal control, prompting, planning, and reinforcement learning.

AIFMs such as BERT, DALL-E, and GPT-4, can improve AI-assisted education in the metaverse by enhancing learning experiences, generating high-quality content, and aiding in data-driven decision-making. They can generate interactive simulations, virtual experiences, and adaptive learning plans. However, they also pose challenges like bias, privacy, and data security. Additional research is required to comprehensively understand their limitations and potential societal impact [29]. Table III shows recent research on AIFMs in education and their contributions to education.

#### 2) LLMs in Education: LLMs are powerful AI systems that can recognize and generate text, making them useful in the learning and teaching of languages. In addition to providing personalized instructions, they can generate learner-centric materials, and streamline the teaching and learning processes. They can generate creative ideas, reduce teacher workloads, and create tailor-made educational experiences. LLMs can analyze data, recommend topics, summarize texts, and provide feedback. They can also assist teachers in correcting errors and adapting texts for learners. AI is already being used in intelligent tutoring systems and educational data mining, and LLMs have immense potential in education [31]. An illustration of different applications of LLMs in education is provided in Figure 4.

LLMs can be utilized in language teaching and assessment technology for content creation, automated grading, and chatbot applications [32]. They offer improvements in text generation, enabling the creation of language learning materials that were previously unattainable. However, careful prompting and reshaping of the model’s outputs may be necessary before use. LLMs do not outperform state-of-the-art results in automated grading and grammatical error correction tasks, but they may provide alternative feedback styles not captured by existing methods. Chatbot applications can use LLMs to create interactions with language learners, such as Duolingo Max, a subscription service facilitated by OpenAI and Duolingo. Multilingual and open-source alternatives to LLMs, such as Open Assistant and StableVicuna, are also available. Nevertheless, ethical considerations are crucial when incorporating LLMs in education technology, as they should be experimented with to understand their capacities and limitations, including mitigating risks like misinformation and harmful bias.

LLMs can be utilized in various educational disciplines, including computer science education, math education, language learning, medicine, history, literature, and social sciences [34]. They provide personalized assistance and guidance in problem-solving, programming, understanding complex mathematical concepts, and improving language skills through interactive conversations and feedback. In medicine, they help students learn medical terminology, understand medical literature, and provide clinical decision support. LLMs can also be integrated into other disciplines like history, literature, and social sciences to offer additional resources, explanations,

---

### Table III

<table>
<thead>
<tr>
<th>References</th>
<th>Domain</th>
<th>Major Discussions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bommasani et al. [29]</td>
<td>General Ed.</td>
<td>The potential of foundation models in education, personalized learning, automation, language support, equity, and access.</td>
</tr>
<tr>
<td>Bommer et al. [31]</td>
<td>Language Ed.</td>
<td>The potential of Language Model Models like OpenAI’s ChatGPT for reducing teacher preparation time.</td>
</tr>
<tr>
<td>Caines et al. [32]</td>
<td>Language Ed.</td>
<td>The use of LLMs in language teaching and assessment technology, the need for ethical considerations and experimentation.</td>
</tr>
<tr>
<td>Jeon and Lee [33]</td>
<td>General Ed.</td>
<td>The interaction between ChatGPT and teachers in education, emphasizing the importance of pedagogical expertise.</td>
</tr>
<tr>
<td>Kumar et al. [34]</td>
<td>General Ed.</td>
<td>The student interactions with Conversational Learning Machines (LLMs) in educational settings, revealing patterns in confidence and trust, suggesting the need for additional learning support.</td>
</tr>
</tbody>
</table>
and interactive learning experiences. Institutions are advising instructors to clarify their stance on AI use in course syllabi.

LLMs come with inherent limitations, including a lack of reasoning abilities, less expressive logical language, and challenges in producing accurate results due to reliance on training data [37]. These factors collectively contribute to making LLMs less trustworthy and reliable. Moreover, LLMs may encounter speed limitations, affecting real-time performance during the inference process. The inference engine’s efficiency is contingent on the expressiveness of the logical language, and if it falls short, the system may experience delays. Therefore, understanding the limitations and potential pitfalls associated with the use of GenAI is essential for ensuring their responsible use.

III. EDUCATIONAL METAVERSES

In the era of technology and digitization, education has been imparted in innovative ways. AI together with virtual technologies such as VR, AR, MR, and XR can be used to create artificial environments that simulate the real world. This technology seamlessly bridges the gap between the physical and the digital worlds as a fusion of physical, augmented, and virtual worlds [38]. There is a growing interest in metaverses and it has been anticipated that they will occupy an important role in the future of the Internet. In addition to transforming the gaming and entertainment industry, metaverse promises to make learning more fun, engaging, and accessible. Below we discuss the prospects of educational metaverses in transforming traditional educational and teaching practices.

A. An Introduction to Educational Metaverses

1) Educational Potential of the Metaverse: The metaverse offers a revolutionary approach to traditional teaching and learning [39]. It provides immersive virtual environments that enable students to understand complex concepts and theories, allowing them to perform experiments that may be too dangerous or costly in real-world settings. This technology also streamlines remote learning, enabling students to participate in virtual classes and engage with peers and instructors regardless of their geographical location. The metaverse also fosters collaborative learning, enabling students to collaborate on various projects and assignments. This collaborative approach contributes to the development of crucial skills such as communication, problem-solving, and leadership. It also allows for personalized learning experiences, allowing tutors to provide feedback based on student performance. Furthermore, the metaverse plays a pivotal role in enhancing educational accessibility for students with disabilities or special needs. Virtual classrooms within the metaverse can be equipped with accommodations such as closed captions and sign language interpretation, ensuring a more inclusive learning environment. Wang et al. [40] presented a theoretical framework for the development of educational metaverses that encompasses four major domains namely knowledge hub, infrastructure design, talent and training, and research and technology.

By creating innovative and immersive learning environments, educators can benefit from the use of the metaverse. In addition to enabling remote learning and collaboration, it can be used as a platform for virtual classrooms, enabling students and teachers to interact in a virtual environment. Using simulations and virtual laboratories, students can engage in hands-on learning experiences that are not feasible in the physical world by utilizing the metaverse [41]. Metaverse integration of AI can improve educational experiences through personalized learning, intelligent tutoring, and adaptive assessment. Blockchain technology can securely protect educational data in the metaverse, including student records and certifications. Figure 5 illustrates different ways of improving education through the metaverse.

In addition to offering cost-effectiveness and flexibility, the metaverse also enhances communication, preventing academic misconduct and facilitating communication [42]. Through VR, students can manipulate objects, watch demonstrations, and receive expert guidance through gamification [43]. A significant role can be played by non-player characters (NPCs) as tutors, peers, and tutors within the metaverse in providing educational services, arbitration, simulation, and decision-making. The role of NPC peers in social constructivism is crucial, as they enhance learning experiences when real users are unable to participate. Table IV shows recent literature on the educational metaverse.

2) Educational Metaverses: Opportunities: Metaverses are virtual worlds where humans interact with non-human avatars using customizable digital objects. With technological advancements, this imaginative vision has become a viable reality, with major corporations such as Facebook (Meta), Microsoft, and Apple leveraging the opportunity. In the educational sphere, the significance of the Metaverse is particularly noteworthy. There is a strong sense of mutual presence in the metaverse, which combines physical and XR. The use of AR, VR, and MR technologies allows large groups to experience a persistent and immersive simulated world from a first-person perspective.
The immersive, multi-sensory nature of metaverse reduces the need for physical travel and infrastructure. Moreover, the educational metaverse can promote equal opportunity and reduce discrimination by incorporating TAI. Personalized learning experiences in the metaverse can be enhanced by integrating TAI that adapts to the needs and preferences of individual students, thereby enhancing student learning outcomes. Furthermore, it can support educators in designing engaging and effective virtual learning experiences through the use of AI-enabled tools and resources. For the educational metaverse to prosper, AI technologies must be designed and implemented in a way that benefits learners and protects their rights.

3) Enabling Multi-modal Smart Classrooms: In the educational metaverse, education can be transformed through multi-modal smart classrooms, immersive technologies such as AR, VR, and MR, and XR (such as holograms), which enhance the representation of teaching content and the interaction between...
Incorporating hybrid, mobile, collaborative, personalized, and project-based approaches into metaverse-based education represents a paradigm shift toward dynamic and interactive learning. As a result, the system offers immersive virtual-physical experiences that overcome the limitations associated with current online and physical education systems [47]. The metaverse, a 3D virtual environment, is being used in education as an alternative space for teaching and learning. It offers customizable platforms, aligns with Universal Design for Learning, and allows for a hybrid approach. The metaverse aims to address traditional video conferencing fatigue [44], fostering community and personalized learning, and improving digital literacy, problem-solving, and collaboration skills.

The learning environment within the education metaverse is characterized by a seamless integration of virtual and real elements, comprising the learner, temporal and spatial dimensions, and the educational event. The learner’s identity in the educational metaverse is versatile, fluctuating between their actual identity and a virtual avatar. This identity transformation is context-dependent, adapting to the specific educational scenario. The metaverse offers a seamless integration of past, present, and future time through the virtual avatar [45]. The metaverse in education is a computer-generated universe that allows learners to engage in social activities, meetings, discussions, and virtual spaces. The metaverse in higher education significantly impacts student motivation, learning attitudes, and self-efficacy. The metaverse enhances experiential learning through hands-on practice and observation, while low motivation levels may result in static learning activities. Therefore, it’s crucial to consider and support students’ motivation levels when designing metaverse-based learning activities, as it significantly impacts their growth mindsets, attitudes, and self-efficacy [49].

**B. Educational Metaverses: Enabling Technologies**

The metaverse is a shared virtual space that combines AR, VR, and the Internet. It offers a digital environment for interaction with virtual elements. The metaverse holds transformative potential in education, offering immersive learning experiences that seamlessly merge the physical and digital worlds, where the students can delve into intricate concepts and immersively engage with virtual environments. In the educational metaverses, the integration of XR augments perception and interaction by overlaying digital information onto the real world. Whereas, VR provides immersive experiences, where users are fully immersed in simulated environments. The integration of the Internet allows access to digital content, applications, and services within the virtual space. Interaction and socialization enable communication, sharing experiences, and collaboration among users. Advanced technologies like 3D graphics and haptic feedback enable immersive experiences, while virtual objects and environments allow users to explore and manipulate them. The metaverse continues to evolve with technological advancements and user experiences.

The metaverse is a virtual world facilitated by XR and the Internet of Everything (IoE). XR combines VR, AR, and MR for immersive experiences, while IoE enables seamless communication and data exchange through interconnected devices. These technologies are crucial for educational services in future metaverses, offering rich multimedia streaming, immersive user experiences, and high-speed communication. XR allows users to be independent of their physical location and interact with virtual and real-world objects. IoE infrastructure, including Web 3.0, 5G, and 6G communication services, enhances user experiences and reduces latency [11].

1) **AR/VR/MR in Educational Metaverses:** An illustration of VR, AR, and MR is shown in Figure 6 and below we
discuss these technologies within the education context.

**AR & Education:** AR is a powerful tool that enhances student learning by providing immersive and interactive experiences that connect the real and virtual worlds. It enables students to visualize and manipulate virtual objects within the real world, fostering the development of critical thinking and problem-solving skills. AR empowers students by providing a platform for independent learning and fostering engagement through interactive activities, enhancing their understanding and retention of knowledge, and creating dynamic and personalized learning experiences. It positively influences students' attitudes toward science and technology, even among those who may initially show less interest. AR can be used in various educational settings, including K-12 education, STEM education, higher education, professional training, and lifelong learning. In higher education, AR creates immersive learning experiences with interactive simulations and games, improving cognitive and motivational aspects, especially in STEM subjects. It can also provide realistic simulations and hands-on experiences for skill acquisition and practice in fields like healthcare, engineering, and vocational training. Distance education can benefit from AR by creating virtual classrooms and web-based learning environments, bridging the gap between remote learners and instructors. AR can also be integrated into Massive Open Online Courses (MOOCs) to enhance learner engagement and provide interactive content and simulations.

**VR & Education:** VR offers various educational benefits, including immersive learning experiences, visualization of complex concepts, hands-on learning, virtual field trips, collaborative learning, personalized learning, accessibility and inclusivity, and motivation and engagement. VR provides realistic and interactive environments, enhancing students' understanding and retention of knowledge. It allows students to visualize abstract or complex concepts through 3D models and simulations, making it easier to grasp difficult concepts. VR also allows students to visit places that may be difficult or impossible to access in real life, enhancing safety, promoting teamwork, communication, and problem-solving skills. It can be customized to meet individual student needs, providing personalized learning experiences. VR also ensures equal learning opportunities for students with disabilities or those in remote areas, promoting adaptive learning experiences. The immersive nature of VR enhances student motivation and engagement, creating a sense of presence and excitement and making learning more enjoyable and impactful.

**MR & Education:** MR is a powerful tool when it comes to education that can enhance students' learning by providing them with immersive and interactive experiences. It allows them to engage with virtual objects and environments, enhancing their understanding and retention of knowledge. MR can create immersive learning experiences by allowing students to visualize complex concepts, interact with virtual objects, and engage in hands-on simulations. It also allows students to visualize complex concepts tangibly and intuitively, allowing them to manipulate virtual objects and explore simulations. This experiential learning approach promotes critical thinking, problem-solving, and skill acquisition. MR also facilitates collaborative learning by allowing students to work together in a shared virtual space, fostering teamwork and communication skills. The immersive nature of MR boosts student motivation and engagement, making learning more enjoyable and impactful. It also allows students to apply their knowledge in real-world contexts, bridging the gap between theory and practice. Lastly, MR can be tailored to individual student needs, providing personalized learning experiences, and allowing students to learn at their own pace, receive immediate feedback, and access customized content.

2) **AI in Educational Metaverses:** By enabling various functionality and improving user experiences, AI plays an important role in the metaverse. In the metaverse, ML/DL, and reinforcement learning are used to achieve human-level learning and intelligence. It is through ML that systems in the metaverse are able to reach or exceed human learning capabilities [41]. As a result of the use of AI in the metaverse, operational efficiency and intelligence are boosted, improving the overall functionality of the virtual world. Voice recognition and communication-powered by AI enhance the metaverse's user experience.

The literature suggests that AI has played a pivotal role in creating the metaverse's foundation and its potential to enhance users' immersive experiences in virtual worlds [50]. AI has significant potential to strengthen the metaverse's infrastructure, enhance 3D immersive experiences, and broaden the range of services provided in virtual environments. As part of the paper, the author explored various AI techniques, including NLP, machine vision, blockchains, networking, digital twins, and neural interfaces, in order to demonstrate their applicability and potential in the metaverse.

AI is increasingly poised to become a driving force behind the metaverse, facilitating its creation, maintenance, and seamless integration of the real and virtual worlds, facilitating avatar participation in a wide range of activities [51]. Several companies are already utilizing AI to develop avatars for the metaverse. Developing digital humanoids using AI for interaction with users within VR environments. AI is instrumental in language processing in the metaverse, enabling natural language interactions by dissecting languages, analyzing content, and generating real-time responses. In the metaverse, the ability of AI to learn from both data and human feedback is paramount for continuous improvement and scalability. In summary, AI plays an essential role in the metaverse, contributing to avatar generation, human-like interactions, language processing, and data-driven learning.

AI, along with IoT, digital twins (DT), big data, and communication networks, is one of the technical enablers in the metaverse ecosystem. By creating a high-quality 3D environment, reproducing real-time responses based on real-world features, and interconnecting events and transactions in the real and virtual worlds, it helps create a high-quality virtual environment. By integrating AI and blockchain in the metaverse, both technologies are enhanced [52]. Using AI, decision-making processes can be simplified, blockchain systems can be automated and optimized, and data security can be enhanced. For IoT networks and to enhance blockchain scalability in the metaverse, AI-enabled blockchain consensus
protocols have been proposed. AI algorithms and techniques are used in these protocols to detect anomalies, generate valid proofs, and adapt the consensus approach to a variety of systems and attacks.

It is possible to incorporate AI with the metaverse into physical education, with an emphasis on mobile internet-based football instruction [55]. Through the use of 360-degree panoramic VR football teaching videos integrated with the metaverse and the K-means algorithm, the authors proposed a strategy for improving football teaching quality. Simulation experiments conducted on the strategy revealed that it was superior in terms of the proxy server hit ratio, the byte hit ratio, the mean response time, and the student experience. A simulation experiment on the proposed strategy was also conducted using CDN Simulator (CDNSim). A number of video content delivery strategies were tested, including a proposed K-means-based optimized 360-degree panoramic VR football teaching video delivery strategy. The paper also highlighted the importance of content distribution technology and the need for efficient use of network resources to provide better services and improve students' quality of experience in football teaching.

Based on fuzzy ontology concepts, AI, and FML, the paper proposes an intelligent agent for high-school student learning, which combines fuzzy ontology concepts, AI, and FML. In order to evaluate students’ interest and performance in learning content, the intelligent agent interacts with the learning environment [54]. In order to present students’ performance, a thermometer was used to predict the learning temperature of the students. In English and CI applications, the implementation of AI-FML Metaverse improved students’ learning outcomes. The program has been evaluated in Taiwanese elementary schools, junior high schools, and senior high schools. In addition, the paper described the steps involved in the construction of the domain ontology using Protege, the creation of a dictionary file for the CKIP Tagger, and the extraction of key concepts using the CKIP Tagger. In addition, a summary of the extracted term sets of student actions and the time spent by students in various learning activities is provided.

AI is crucial in the metaverse for enhancing the learning experience and providing personalized education. It analyzes user data, creates personalized learning paths, and provides real-time guidance. AI-powered virtual assistants offer real-time support, and intelligent content creation and curation generate engaging educational materials. AI also facilitates automated assessment and personalized recommendations [11].

It is possible to utilize AI-powered bots and assistants in the metaverse to enhance the educational experience by providing personalized guidance and assistance [55]. In the metaverse, AI can analyze vast quantities of data in order to uncover information about students’ behavior, preferences, and learning patterns, thereby providing personalized and adaptive learning experiences. Through AI-driven NLP, students can be able to communicate and interact with virtual environments, enhancing their immersion and engagement in the educational experience. In the metaverse, AI can assist in the creation and execution of creative marketing research and customer relationship management techniques, which can also be applied to educational environments. In the metaverse, AI-driven predictive analytics can support educators in gaining a deeper understanding of student behavior and learning outcomes, which will enable them to make data-driven decisions related to education. By enhancing personalization, interactivity, and data-driven insights, AI plays a crucial role in leveraging the potential of the metaverse in education.

C. Augmented and Virtual Educational Metaverses

1) Augmented Educational Metaverse: Augmented metaverse combines the real world with the digital environment, allowing for the simulation of abstract graphics on physical objects. In education, augmented metaverse technology has numerous applications, such as in medical education the research team can create a platform for spinal surgery using AR to project a pedicle screw onto a human body shape in real time. Moreover, AR simulation content can also be used as an educational tool, relating abstract graphics to physical objects and enhancing learning experiences. For instance, an augmented metaverse can be used in science education to visualize complex scientific concepts, create interactive learning environments, and provide real-time translations. Augmented metaverse can also support special education by providing personalized learning experiences and assistive tools for students with disabilities. In vocational training, it can simulate real-world scenarios and provide hands-on practice in a safe and controlled environment. Overall, augmented metaverse in education offers a wide range of applications that enhance learning experiences, promote engagement, and provide interactive and immersive educational content [56].

In addition, the augmented metaverse has the potential to revolutionize various educational levels. For instance, in K-12 education, it can elevate student understanding and engagement by providing interactive and immersive learning experiences in subjects like physics, chemistry, biology, and mathematics. In higher education, it can create interactive learning environments for subjects like engineering, computer science, anatomy, and astronomy. Augmented metaverse can also be used in professional training programs to provide realistic simulations, virtual practice environments, and interactive learning experiences for various industries. It can also support lifelong learning by providing interactive and immersive experiences for individuals of all ages, allowing them to explore and understand complex concepts in various subjects. In computer science, an augmented metaverse can create coding environments, allowing students to practice programming and explore computer science concepts in an immersive manner. It can also be used in environmental science to study phenomena like virtual field trips and data visualizations. In astronomy, an augmented metaverse can offer interactive visualizations and virtual tours of the universe. In anatomy and physiology, AR can provide interactive visualizations and simulations of the human body. Lastly, an augmented metaverse can enhance the study of nanotechnology by providing visualizations and interactive simulations of nanoscale phenomena, allowing students to explore and manipulate materials at the atomic and
molecular levels. In engineering, an augmented metaverse can overlay virtual objects and information onto real-world engineering projects, allowing students to visualize and interact with designs, prototypes, and simulations. It can also enhance mathematical learning by providing interactive visualizations, geometric explorations, and virtual manipulatives. Figure 7 illustrates the synergy among learning sciences, resources, assessment, and recommendation systems in an augmented space realized through an augmented educational metaverse.

2) Virtual Educational Metaverse: Across diverse educational domains, the virtual metaverse could play an instrumental role in K-12 education, STEM education, higher education, and professional training, etc. As part of K-12 education, the virtual metaverse facilitates immersive simulations, historical recreations, and scientific explorations, which enrich students' understanding of intricate concepts. The use of virtual objects and scenarios can facilitate active learning and critical thinking in higher education by providing students with an opportunity to explore and interact with them.

In STEM education, the virtual metaverse plays a crucial role in providing students with hands-on experiences and simulations to enhance their understanding and engagement in subjects such as physics, chemistry, mathematics, astronomy, and biology, etc. Moreover, professional training can be simulated in the virtual metaverse to develop practical skills, including medical procedures, engineering tasks, and safety training. Furthermore, the virtual metaverse has the potential to support lifelong learning by providing opportunities for interactive and experiential learning. By delivering immersive and interactive learning experiences, the virtual metaverse can transform education across multiple levels, enhancing student engagement, comprehension, and skill acquisition. Figure 8 illustrates a virtual educational metaverse.

D. Educational Metaverses: Risks and Challenges

1) Lack of Inclusiveness and Accessibility: The accessibility of the educational metaverse is a key concern due to disparities in communication modes and languages. It is crucial to integrate various communication methods to ensure universal usability. The risk of unequal access to foundational technology is another concern, as the educational metaverse depends on high-speed internet, cloud resources, XR devices, computational power, and AI models. The XR digital divide can deepen existing inequalities, particularly for those with limited access in the global south. Also, there are concerns regarding accountability, oversight, and individual liberties within the educational metaverse due to the concentration of power and corporate monopolies.

2) Ethical Considerations: The integration of AI in the educational metaverse raises ethical concerns about fairness, accountability, transparency, bias, autonomy, agency, and inclusion. There is a lack of training for researchers and divergent interpretations within the community, making it difficult to establish a coherent framework. Despite the importance of transparency, concerns about algorithmic bias persist. Human-centered design approaches are increasingly important to minimize potential harm. In educational context, the proliferation of AR/VR devices has led to an increase in the collection of student data, which poses potential risks to marginalized communities. In the educational metaverse, ethical questions arise about the responsible use of influential technology in the form of manipulative advertising and behavior modification. To ensure that educational metaverses' experiences are shaped by a variety of perspectives, diversity must be represented during system design.

The integration of AI into educational metaverses possesses challenges such as privacy, informed consent, data breaches, algorithmic bias, and lack of explainability and interpretability. The need to develop TAI tools for art education becomes evident when students express ethical concerns about copyright issues and the potential replacement of artists. There is a degree of controversy among students regarding the use of AI tools beyond the classroom, with varying opinions regarding the enhancement of their final artistic works as a result. Using AI-generated images as finished products offers an intriguing prospect worth further exploration. Educators and researchers must address ethical considerations associated with AI-generated art in future curricula. It is imperative to develop effective strategies to harness the full potential of AI tools within the art and design educational frameworks. The intricate nature of AI necessitates risk-intensive implementation procedures. Moreover, a set of predefined
guidelines as well as interdisciplinary research are necessary to address ethical considerations, including issues of user rights and digital copyright division [2].

3) Unintended Harmful and Antisocial Consequences: Educational metaverse settings can cause discomfort or distress for those in proximity, especially if the content is offensive or inappropriate. Unintended harmful consequences include VR Fatigue and "VR Hangover," which can trigger dissociative depersonalizing conditions. Cybersickness, similar to motion sickness, can result in nausea, dizziness, and disorientation. High realism in educational metaverse environments could lead to severe health consequences such as heart attacks and fatal incidents. Further, drawbacks include learner fatigue, disruptions in the virtual classroom environment, and the possibility of inaccurate information [44].

Antisocial side effects include filter bubbles and echo chambers, where personalized online content can isolate individuals within tailored information spheres. AR/VR may also magnify the propagation of false narratives and information, highlighting the need for responsible digital citizenship and ethical guidelines. Empathy used antisocially can be misused in VR-generated emotional experiences, potentially leading to polarization and manipulation. Moreover, students in educational metaverses can be subject to cyberbullying and harassment that could have real impact on their lives.

4) Privacy and Security Issues: Several challenges arise within an AI-driven educational metaverses, such as privacy concerns, algorithmic bias, and transparency concerns [1]. The issue of digital copyright within the educational metaverse is also of pressing concern. Cybersecurity issues can be encountered in the dynamic metaverse landscape [58]. In educational metaverses, these issues can be addressed through the use of trustworthy AI. Plass et al. [59] emphasize the importance of trustworthiness in AI solutions and suggest a user-centered approach for AI design and human interfaces. This involves stakeholder analysis and user testing, focusing on users and use cases throughout AI development. Moreover, for AI to be trusted in the metaverse, several fundamental factors are required including human oversight, technical resilience, privacy and data governance, transparency, non-discrimination, and accountability.

TAI promotes transparency and explainability, making it easier for users to understand how AI systems make decisions in the metaverse. To implement TAI in the educational metaverses, it is necessary to develop and adhere to ethical guidelines and frameworks. Consequently, it is necessary to ensure that student data is protected and secure in the metaverse environment. It is essential to monitor and evaluate AI systems in the metaverse continuously to identify and mitigate any potential biases or unintended consequences. Establishing best practices and guidelines for TAI in the educational metaverses requires collaboration between educators, AI developers, and policymakers [60]. By adopting a modular approach, metaverse applications can be tailored to specific contexts, which enhances both security and trustworthiness [38].

5) Risks Associated with GenAI and AIFMs: While foundation models have proven to be beneficial for educational purposes, they also face a variety of challenges such as bias, privacy, data security, interpretability, limited understanding, the potential for widening academic disparities, and the potential for societal impact [29]. The foundation model is often referred to as a black box system with limited understanding. Additionally, foundation models may inherit biases from other models, potentially impacting AI systems. This raises concerns about information veracity, potential bias, and misinformation, while also posing equity and access issues, consuming significant energy, and contributing to environmental damage. Developing trustworthiness into applications using these models will be critical to ensure the efficacy of the educational metaverses. This necessitates the development of novel framework to evaluate these models [61].

Generative models, LLMs, faces several challenges in the educational metaverse. These include a lack of robust reasoning abilities, unpredictable and difficult-to-interpret outputs, limited control over generated content, potential bias or inappropriate content, ethical considerations, and computational demands. These issues can limit the effectiveness of GenAI in providing accurate and reliable educational content. Additionally, the computational demands of GenAI models, particularly those using highly expressive logic, can impact the speed and performance of educational applications in the metaverse. It is therefore imperative that educators monitor and validate the successful implementation of GenAI in the educational metaverse in order to achieve trustworthiness [37].

IV. Trustworthy AI in Educational Metaverses

In this section, we will elaborate upon various essential methodologies that can be leveraged to attain TAI in educational metaverses. We start by first highlighting the significance of TAI for educational metaverses.

A. Unveiling the Significance of TAI for Educational Metaverses

To ensure TAI in educational metaverses, transparency, accountability, data privacy, user feedback, collaboration, and education are crucial. Transparency involves providing clear explanations for AI systems’ decisions, while accountability involves holding them accountable for their actions and adhering to ethical guidelines. Data privacy ensures data security and responsibility, while user feedback encourages continuous improvement. Collaboration involves involving stakeholders in AI system design, fostering trust and ownership. Education refers to providing training on AI systems’ capabilities, limitations, and potential biases, enabling informed decision-making and critical evaluation. The educational metaverse being a human-centric system should incorporate ethical values and cultural norms into AI systems, prioritizing human well-being, empathy, and understanding. Implementing polite and compassionate AI systems can perpetuate positive social behaviors [23]. Addressing algorithmic biases and imperfections is crucial, as AI systems must understand the complexities of the human condition.

To achieve trustworthy GenAI in the educational metaverse, explicit knowledge and rules of thumb for AI models are essential. Step-by-step reasoning can enhance trustworthiness
and interpretability by providing a full process behind AI-generated answers. Documenting and auditing knowledge sources at each step in the reasoning chain allows users to verify the provenance of the knowledge used in AI-generated outputs. The use of a hybrid approach combining the strengths of GenAI models with more formal approaches is suggested in [37]. TAI should prioritize concise explanations based on user context, prior knowledge, and resource constraints. Future possibilities in trustworthy GenAI and AIFMs include improved reasoning capabilities, enhanced interpretability, ethical considerations, efficient computational methods, integration of formal approaches, and personalized and context-aware content generation. Advancements in GenAI and AIFMs can enhance reasoning abilities, enabling more accurate and reliable educational content. Ethical considerations can be addressed to ensure unbiased and appropriate content generation. Moreover, efficient computational methods can be optimized for real-time performance in educational applications such as education metaverses. The literature argues that hybridizing GenAI with formal approaches can lead to more trustworthy outputs [37].

B. Ethical AI Practices for Trustworthiness

The educational metaverse faces ethical concerns regarding AI’s application, including data and model privacy and security, algorithmic bias, and its impact on student autonomy and agency. To ensure ethical use of AI in educational metaverses, educators must be aware of these concerns and consider transparency, accountability, and fairness. Further research is needed to understand the ethical implications of AI in the metaverse and develop guidelines and frameworks to address these issues [1]. TAI can significantly impact the future of the metaverse, but it requires explicit ethical norms to control user behavior and maintain order [62]. The literature articulates that to address these challenges, AI systems should be developed and implemented with ethical considerations as a cross-cutting discipline [63]. Collaboration between AI and human specialists is recommended to ensure robust and reliable AI decisions, verifying AI-generated knowledge and aligning explanations with domain expertise. Human oversight is also crucial in identifying spurious findings, contextual differences, and reducing bias in AI decisions [64].

Methods to overcome trust, ethics, and privacy issues include integrating diverse data sources into AI models, curating training data to avoid biases and stereotypes, encouraging data and media literacy, and developing policies and regulations to ensure secure, ethical, and responsible use of AI technology in educational contexts [65]. The bottom-up approach in AI Ethics involves collaborative assessment of local education practices to establish professional standards, necessitating increased support and access to development settings in education to facilitate multidisciplinary research and development [66]. Moreover, a recent study shows that the students have raised concerns about the use of GenAI in art, especially copyright issues and job displacement [12]. Also, the authors emphasized the need for ethical considerations when integrating AI tools into art and design education. Integrating LLMs can facilitate discussions on ethical implications, allowing students to understand the implications of AI usage.

C. Using Explainable AI to Establish Trust in Educational Metaverses

Explainability in AI refers to the ability of AI models and their outcomes to be easily understood by humans, ensuring transparency and trustworthiness. Explainable AI (XAI) is a crucial tool for AI-empowered education that could enhance students’ understanding of AI-driven systems and provide teachers with insights into AI-powered educational tools. However, XAI techniques face challenges in explaining complex models in a clear and comprehensive manner [25], potentially limiting their use in educational settings. There is a lack of uniform standards in XAI, necessitating the establishment of standardized metrics and frameworks to assess and evaluate explanations generated by various XAI approaches [67]. Educational initiatives using XAI focus on exploring objective aspects of students’ learning processes and using XAI tools to explain opaque ML models and rule-based systems. The explanations are predominantly visual-based and symbolic-based, with some researchers exploring the integration of neural-symbolic computing to address interpretability issues.

Langer et al. [68] presented a stakeholder perspective on XAI and proposed a conceptual model that guides interdisciplinary research on XAI, emphasizing the need for explainability methods to establish trustworthiness in AI systems. In order to reinforce trust, it may be helpful to provide insight into the system’s resilience, the conditions under which its outputs are reliable, and scenarios in which its outputs could be misleading. Furthermore, trustworthiness requires accountability. The determination of responsibility for errors is an essential component of the use of AI systems. It requires collaborative efforts across various disciplines to develop explainability approaches that are aligned with stakeholders’ preferences. By understanding stakeholder perspectives and addressing their objectives and requirements, AI systems can be made more trustworthy. Consideration of the preferences of diverse stakeholders, such as those who are seeking user-friendly systems or those who are interested in learning how to utilize them, can facilitate the development of trust in AI.

An approach to establishing trustworthiness in AI systems is presented in [69]. Where the objective is to enhance trust by providing personalized and on-demand explanations. It emphasizes the need for human involvement in decision-making and transparency. The study explores various explanation types (i.e., rule-based and example-based) to understand their impact on trustworthiness. It highlights the interconnections between controllability, transparency, fairness, trustworthiness, and understanding. The authors also attempted to address situational factors like surprise and confusion, aiming to enhance trustworthiness in AI systems. In the metaverse, XAI empowers system engineers and data scientists to gain insight into AI models, assess their outcomes, and mitigate potential legal and security risks, while maintaining a highly reliable user experience [50].

D. Secure and Private AI for Educational Metaverses

The immersive experience in the metaverse relies on wearable devices like VR headsets, which can collect fine-grained
Fig. 9. Towards TAI in the Educational Metaverse.

16

personal data. To ensure privacy-preserving AI, data can be
anonymized and encrypted using for various data encryption
techniques such as homomorphic encryption, garbled circuits,
etc. Moreover, the privacy of the training data can be ensured
using other methods such as differential privacy, and federated
learning [70]. Furthermore, to enhance the robustness of AI
systems against adversarial attacks, various defenses can be
leveraged such as adversarial training, input reconstruction,
defensive distillation, and adversarial detection [71], etc.

AI computation within the metaverse can be made more
secure and responsible with the help of blockchain technology.
By enhancing privacy and data integrity, it effectively ad-
dresses critical data security concerns for AI. The fragmented
and isolated nature of AI-related data makes it difficult to share
it efficiently between organizations. Due to the highly sensitive
nature of the data collected in the metaverse, the privacy
problem gets exacerbated. Since blockchains are decentralized,
anonymous, and immutable, they provide a secure and private
mechanism for sharing data. As a result, not only is responsible
AI development made possible, but also users are allowed to
share data without fear of violating their privacy, enabling
customized and reliable services to be provided [72]. In [62],
the authors emphasized that utilizing blockchain to store users' 
private data can enhance data security and transparency.

Privacy regulations can impede the practical application
of AI methods in education metaverses. To address these
concerns and maintain the confidentiality of sensitive user
data, privacy-preserving techniques, such as differential pri-

E. Controlling Content Generation in GenAI

GenAI has shown promising results in tasks like text and
image generation, making it a promising solution for augment-
ing AI-driven education systems. Retrieval-based approaches
in GenAI aim to provide more information, enhancing their
generative capabilities. This includes retrieving grounding
contexts from external sources like websites, books, images,
codes, tables, graphs, and audio. GenAI models can serve as
general-purpose learners in different formats, which can be
beneficial in educational settings. Retrieval-augmented gener-
ation addresses important concerns in education, such as factu-
ality, reasoning, interpretability, and robustness [73]. Retrieval-
augmented modeling in education can enhance the reliability
and trustworthiness of AI-generated content by allowing mod-
els to access external memory and retrieve relevant documents
[74]. This allows for accurate and up-to-date knowledge,
making the decision-making process more understandable and
trustworthy. Moreover, the ability of GenAI models to interact
with the external world and retrieve knowledge in different
forms can enhance their learning capabilities in educational
contexts. GenAI and retrieval-augmented generation have the
potential to significantly impact education by improving gen-
erative models’ ability to generate relevant and accurate edu-
cational content. Recent advancements in pre-trained text rep-
resentations and attribute discriminators can further enhance
the trustworthiness of controlled text generation by achieving
conditional generation without re-training [75], [76]. Figure 9
delineates various methodologies for attaining TAI within the educational metaverse.

Controlled content generation is crucial in the realm of TAI in education, as it ensures that AI-generated content meets specific constraints such as accuracy, appropriateness, and educational standards. This approach promotes trust and credibility in educational AI applications, generating reliable, unbiased, and aligned materials with curriculum guidelines. It also allows for personalized, engaging learning experiences, ensuring transparency, explanation, and understanding of the content. Controlled text generation in AI-powered educational systems also helps mitigate the risks of misleading or inappropriate content, ensuring the delivery of accurate and trustworthy information to students [77]. Educational AI systems can generate explainable and faithful explanations and trustworthy information to students [77]. Educational AI systems can generate explainable and faithful explanations and predictions, providing transparent reasoning based on retrieved documents. In this regard, retrieval-augmented modeling can generate knowledge-intensive outputs, such as accurate and contextually relevant image and text generation, contributing to the development of reliable and informative content for learners.

V. OPEN RESEARCH ISSUES

A. Enhancing Inclusivity in Educational Metaverses

The metaverse offers a transformative platform for assistive technology in inclusive education, overcoming challenges such as lack of universal solutions, cost issues, and limited awareness is crucial to realize true potential of metaverse. By integrating VR, AR, and MR technologies, the metaverse facilitates immersive and interactive learning experiences, allowing personalized and adaptive learning through AI and ML algorithms. This allows students with learning difficulties to engage in collaborative, inclusive environments in real-time, addressing limited research and awareness among educational stakeholders. The metaverse also offers assistive tools like text-to-speech, speech recognition, and visual aids, enhancing accessibility for students with diverse learning needs. For instance, a virtual public space can be customized for users with visual impairments, emphasizing other senses with haptic or audio interfaces, e.g., text-to-speech features can be used for alternative communication. AR/VR can also promote disability rights and workplace equity, ensuring equal access to experiences and opportunities for individuals with physical disabilities. However, the integration of assistive technologies within the metaverse, will lead to unique challenges.

B. Multi-level Authentication Strategies for Enhanced Cybersecurity

The AI-based metaverse, which integrates VR, AR and MR could face cybersecurity challenges, particularly in user identification [58]. Biometric methods like fingerprint, face, and voice recognition are vulnerable to security threats like spoofing attacks, and detection errors. Sharing biometric data with social network operators also poses a threat to privacy and security. False-acceptance and false-rejection rates are also significant issues, as they can compromise user interactions and user experience. Addressing these cybersecurity challenges is crucial to ensure the trust and safety of users using these immersive technologies. Implementing robust security measures, encryption protocols, and continuous advancements in biometric authentication methods is essential to mitigate risks associated with these immersive technologies. Improved accuracy and robustness in biometric authentication methods are essential for the effectiveness of these technologies in VR, AR, and MR settings. ML-based biometric systems should be made robust to adversarial attacks to enhance user security within the educational metaverse.

C. Advancing Data Security with Federated Learning and Blockchain Integration

The integration of Federated Learning (FL) and blockchain technology in educational metaverses like VR, AR, and MR offers significant benefits in enhancing data privacy and security. FL enables on-device ML, reducing the risk of data breaches and ensuring the security of personal information and learning data. Blockchain technology contributes to decentralization and security of data storage and transactions, protecting user data and digital assets. The combined integration of FL and blockchain can create a robust, privacy-preserving environment for users while safeguarding personal information and digital interactions. This will foster trust in the educational metaverse and encourages active participation in immersive and collaborative learning experiences.

Future research in VR, AR, and MR-based educational metaverses can possibly focus on improvising robust privacy-preserving techniques. This includes investigating privacy-preserving AI techniques, ensuring secure communication between central orchestrators and end devices, optimizing consensus algorithms, and integrating different privacy techniques. The research agenda also can include analyzing trade-offs between privacy, computation cost, and communication overhead. Thereby establishing a resilient framework that balances user privacy, performance, scalability, and security for successful immersive learning experiences.

D. Optimizing Educational Outcomes with Multi-Modal AI Foundation Models

In the realm of the metaverse, multi-modal language models, AI bots, and ambient intelligence, underpinned by AI foundation models such as GPT, BERT, and CLIP, are revolutionizing multi-modal learning [29]. These technologies can be used as an augmentation tool for language acquisition and spatial understanding. Similarly, advancements in multi-modal pre-training methods are vital for aligning features across 3D and 2D modalities and language, thereby improving the understanding of spatial information in virtual environments. Challenges in assembling scalable, high-quality multi-modal data for 3D applications are being addressed by innovative frameworks like ULIP-2 [78], which leverage large multi-modal models for generating comprehensive language counterparts for 3D objects. Future research on models like BLIP-2 promises further advancements in zero-shot classification, enhancing interaction accuracy with 3D objects in virtual
spaces [72]. Ultimately, the integration of these multi-modal AI/FRs is key to unlocking the full potential of educational technologies in the metaverse, leading to optimized learning outcomes and richer, more immersive educational experiences.

E. Techno-Pedagogic Design for Human-Centered Process-Oriented Learning

In the rapidly evolving landscape of educational technology, particularly with the advent of educational metaverses that leverage GenAI and extended reality (AI-XR), there’s an urgent need to recalibrate our focus towards the human-centered nature of learning. Tools like ChatGPT have revolutionized the way educational content, such as homework and reports, is generated, streamlining and automating tasks that were traditionally the remit of human educators. However, it’s crucial to recognize that education is fundamentally a transformative process, not merely a production line for outcomes. This shift in perspective calls for innovative technopedagogical approaches that prioritize process-oriented learning, ensuring that the focus remains on enhancing, developing, and augmenting human capabilities [80]. While GenAI and extended reality are powerful tools, their true potential lies not in churning out “products of learning”, but in creating nurturing and empathetic environments that foster genuine human learning [81]. through modalities that support personalized learning and well-being [83]. As we delve deeper into the realm of educational metaverses, the techno-pedagogic design of these systems emerges as a critical area for research and development, underscoring the need to harmonize technological advancements in educational metaverses with the intrinsic human-centric ethos of education.

F. Ethical Use of AI-XR in Educational Metaverses

As we embrace the burgeoning potential of AI-XR technologies, it is imperative to address the inherent risks associated with their use, particularly in the realms of exploitation, manipulation, and persuasion [38], [43]. AI-XR environments have the capacity to create highly immersive and convincing experiences, which, if misused, can lead to significant ethical dilemmas and manipulative scenarios. The power of AI-XR to influence perception and behavior underscores the need for robust technical safeguards, political oversight, and comprehensive regulatory frameworks. Such measures are essential to prevent the abuse of these technologies, ensuring they are used in ways that are beneficial and respectful of individual autonomy and societal norms. Drawing from the insights of Rosenberg [85] in his work on the regulation of XR technologies, it becomes clear that without these critical checks and balances, the potential for harm is significant. Therefore, it is not just desirable but essential to implement multi-layered strategies encompassing technical, political, governance, and regulatory aspects to govern the use of AI-XR technologies. This approach will help in maintaining ethical standards, protecting user privacy, and ensuring that XR serves as a tool for positive and constructive engagement rather than a means for deceptive or coercive practices.

VI. CONCLUSIONS

Metaverse holds significant promise for transforming education through immersive experiences, with Artificial Intelligence (AI) technologies, particularly Generative AI (GenAI), playing a pivotal role. However, the human-centric nature of educational metaverses necessitates a focus on ensuring the trustworthy application of AI. To bridge the gap between Trustworthy AI (TAI) and educational metaverses, we present a comprehensive study focusing on educational metaverses, enabling technologies, the need for credibility, and proactive measures to achieve TAI in educational metaverses. We navigate through challenges encompassing data privacy, accessibility, cybersecurity, economic considerations, and ethical issues. Furthermore, we discuss possible solutions that can be leveraged to attain TAI in educational metaverses. Finally, we highlight various open research issues that require further investigation to advance the integration of TAI in the evolving landscape of educational metaverses.

VII. ACKNOWLEDGEMENTS

The authors would like to acknowledge support from Qatar University High Impact Internal Grant QUHI-CEG23/24-127 and Open access funding provided by Qatar National Library.

REFERENCES


