Asia Pacific Academy of Science Pte. Ltd.

Metaverse 2026, 7(2), 8449.
https://doi.org/10.54517/m8449

Article

An Enhanced Explainable Hybrid Framework for Modeling Metaverse-Based E-Learning Acceptance Among Students with Disabilities: Integrating SEM, Neural Networks, and Interpretable Machine Learning

Suhaila Abuowaida[1], Nawaf Alshdaifat[2], Hamza A. Mashagba[3], Azlan B. Abd Aziz[3,*], Alaa Alzoubi[4], Mardeni Bin Roslee[5], Mohamad Yusoff Alias[5], Azwan Mahmud[5]

[1] Department of Data Science and Artificial Intelligence, Faculty of Prince Al-Hussein Bin Abdallah II for IT, Al al-Bayt University. Mafraq, Jordan.

[2] Department of Information Technology, Faculty of Prince Al-Hussein bin Abdullah II of Information Technology, The Hashemite University, Zarqa, Jordan.

[3] Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia, Centre for Wireless Technology (CWT), Multimedia University

[4] Department of Computer Science, Applied Science Private University, Amman, Jordan.

[5] Faculty of Artificial Intelligence and Engineering BR4064, Level 3, Multimedia University, Persiaran Multimedia 63100, Cyberjaya, Selangor, Malaysia.

*** Corresponding author:** Azlan B. Abd Aziz, azlan.abdaziz@mmu.edu.my

CITATION

Abuowaida S, Alshdaifat N, Mashagba HA, Abd Aziz AB, Alzoubi A, Roslee MB, Alias MY, Mahmud A. An Enhanced Explainable Hybrid Framework for Modeling Metaverse-Based E-Learning Acceptance Among Students with Disabilities: Integrating SEM, Neural Networks, and Interpretable Machine Learning. Metaverse. 2026; 7(2): 8449. https://doi.org/10.54517/m8449

ARTICLE INFO

Received: 18 January 2026
Accepted: 27 April 2026
Available online: 26 June 2026

Abstract: While metaverse technologies offer numerous opportunities for inclusive learning for children with disabilities, there is limited empirical research on how these same children accept or reject such technologies; as such, there are few if any, established methodological frameworks, which can be used to explain the complex acceptance behavior of children with disabilities toward metaverse technologies. This research provides an advanced hybrid analytical framework that will extend beyond structural equation modeling (SEM) and artificial neural networks (ANN) by employing additional explanatory AI techniques, ensemble models, and clustering methods to analyze student acceptance data from children with disabilities. This study will be conducted using a quantitative cross-sectional survey design with a combination of five complementary analytical approaches): 1.) Structural equation modeling (SEM), 2.) Deep neural networks, 3.) Random forest and XGBoost ensemble methods to determine the most important features related to student acceptance, 4.) SHAP and LIME methods to understand why the various AI models developed were making predictions regarding student acceptance, and 5.) Unsupervised clustering to identify subgroups of students based on type of disability. Using a large dataset of over 847 students with disabilities from multiple school-based settings, this research demonstrated improved predictive power of the enhanced framework when compared to the use of single-method approaches, while maintaining sufficient levels of theoretical interpretability. Overall, the findings of this research indicate that the two primary acceptance factors for children with disabilities toward metaverse technologies are: 1.) Perceived optimism toward the benefits of the metaverse technology, and 2.) Social presence. Furthermore, the relative importance of each factor varied significantly depending upon the specific type of disability (cognitive vs. physical). This finding highlights the need for design teams to develop acceptance-enhancing strategies tailored to the needs of students with different types of disabilities. Additionally, the findings of this research demonstrate the potential of the explainable AI components of the framework to provide educators and developers with specific actionable insight into the factors driving acceptance decisions made by individual children. Therefore, this research contributes to the field of special education research an empirically grounded, methodologically sound, theoretically interpretable, and practically applicable framework for studying the acceptance of technology in inclusive educational environments, as well as the practical applications of developing acceptance enhancing features within metaverse platforms.

Keywords: Metaverse; Explainable AI; Machine learning; SHAP; LIME

1. Introduction

Metaverse technologies are beginning to transform educational paradigms to provide students with immersive and interactive educational experiences. Students with disabilities are particularly well-suited to take advantage of these emerging technologies as many experience barriers in traditional educational settings [1,2]. However, despite increasing research interest in educational artificial intelligence systems, existing studies report conflicting results concerning how students with disabilities respond to educational technology; specifically, traditional models of technology acceptance fail to adequately predict responses from students with disabilities [3,4].

More recently, advancements in explainable artificial intelligence provide researchers with new ways to analyze the complex reasoning behind technology adoption decisions made by educational technology [5,6]. The explainable artificial intelligence is viewed favorably in educational research and supports outcomes such as increased student success, trust and motivation, using methodology such as SHAP and LIME to illustrate how explainable artificial intelligence models can explain decision making to users. Studies indicate that although more complex models, such as artificial neural networks, produce greater predictive accuracy, they lack sufficient transparency for practical application in educational settings [7,8].

However, the ability to utilize explainable artificial intelligence in educational settings has been identified as a potential way to identify at risk students and provide targeted support and intervention. The explainable artificial intelligence in education shares similarities with general artificial intelligence, yet has unique requirements, and proposed a framework for the study and development of educational artificial intelligence tools, based on six specific aspects of educational artificial intelligence [9,10].

Building on previous research that demonstrated the effectiveness of a hybrid SEM-ANN approach for modeling metaverse acceptance by students with disabilities [11–13], this study introduces an Enhanced Explainable Hybrid Framework that addresses three major limitations found in current research. The first limitation is the inability to interpret the results of complex predictive models, which limits their practical utility in educational settings. The second limitation is that few studies have examined the heterogeneity of disability populations, with most treating all students with disabilities as one homogenous group. [14] The third limitation is that there is limited practical utility for educational practitioners, who require actionable insights and not just theoretical constructs [15–17].

The Enhanced Explainable Hybrid Framework utilizes five complementary analytical elements. Structural equation modeling provides the foundational understanding of construct relationships, referred to as theoretical validation. Complex pattern recognition, utilizing deep (non-linear) neural networks, enables the modeling of non-linear relationships that may not be able to be identified using traditional statistical techniques. Feature importance analysis utilizes ensemble techniques, such as random forest and xgboost, to rank predictors according to data-driven rankings. Interpretable model interpretation utilizes explainable artificial intelligence methodologies, such as SHAP and LIME, to provide explanations for the complex predictions made by the models. Population segmentation utilizes a clustering analysis to divide the disability population into distinct subpopulations that may benefit from different types of interventions.

2. Materials and methods

2.1. Metaverse Adoption in Special Education

Understanding the Metaverse as an Interactive Design and Learning Environment. The Metaverse is considered a new paradigm for the learning environment; it offers a more advanced form of the e-learning platform. The Metaverse is characterized by persistent, synchronous, three-dimensional virtual environments, which provide a space for users to interact through their virtual embodiments. In contrast to the traditional LMS model, in which learners are limited to accessing content via text and video, the Metaverse allows for real time social interactions among users and provides for an immersive experiential learning process. The differences in these modalities of interaction present significant challenges for learners with disabilities who may encounter barriers or have additional requirements when using accessible technology in a fully immersive virtual environment [18]. While the Metaverse offers new and exciting opportunities for accessibility, such as enabling social inclusion for learners with mobility impairments, the accessibility of this virtual environment also introduces new barriers and obstacles for learners. For example, learners with mobility impairments may find that navigating a virtual environment is easier than navigating a physical one, however, they may still encounter barriers and obstacles due to the complexities of spatial navigation and the visual and auditory demands of a virtual environment. Furthermore, the social/technical aspects of developing a Metaverse represent a unique combination of technological infrastructure, social norms, interaction protocols, and design standards. As such, there is a need to create a new type of socially inclusive, technologically accessible, and pedagogically relevant environment that will provide equal access to all learners [19]. The development of accessible Metaverse environments requires an understanding of the technical challenges involved in creating 3D scenes and the impact on learner's experiences of interacting with virtual environments. While advancements in photogrammetry and automated 3D modeling offer the potential to rapidly develop virtual scenes, consideration must be given to how learners with differing abilities perceive and interact with virtual spaces. The intersection of technological capabilities and inclusive design processes will be critical to understanding how learners with disabilities use and interact with Metaverse based education technologies.

Research into the metaverse in education has been quite extensive due to the fact that this is a multi-dimensional, connected, 3-D environment where users can interact, present virtual environments, digital objects and other people through their senses. Othman et al., (2024) developed an analytical model to provide a framework for accessibility and inclusion in the metaverse by examining whether it will increase digital accessibility for people with disabilities. The researchers assessed the feasibility of improving digital accessibility for people with disabilities, through analyzing the responses of eleven digital accessibility experts, experts developing the metaverse, advocates for people with disabilities, and those formulating policy [18].

For students with disabilities, these metaverse environments have special benefits, such as experiencing situations in three-dimensional scenes accompanied by synchronized sound effects and becoming immersed as part of the learning experiences that go beyond the constraints of a physical classroom. The recent literature on the uptake of the metaverse in special education has outlined opportunities and challenges. While VR platforms are especially promising for entertainment, communication, and education, there are several obstacles that keep many people with disabilities from taking complete advantage of virtual worlds [15].

Metaverse technology has educational value beyond standard e-learning approaches. Immersive simulations Immersive simulations cover the same range of subjects that are true in the real, outside world, but it's easier to use them to help you come to grips with a concept. But even as hardware and software become cheaper, the digital divide pervades and impacts those with disabilities more than most, erecting barriers toward equitable access [16].

2.2. Technology Acceptance Models in Special Education

Some of the traditional technology acceptance models, such as the Unified Theory of Acceptance and Use of Technology and the Technology Acceptance Model, have had some success for different categories of disability groups. Nguyen et al. (2024) empirically study the determinants of acceptance and use of educational metaverses, proposing an extended model called the Unified Theory of Acceptance and Use of Metaverse Technology, which can serve as a reference for educators and managers [17].

The study by Nguyen et al. (2024) was conducted with 253 Vietnamese participants (i.e., teachers and students) who were familiar with the metaverse, and their selection was based on purposive sampling. The Unified Theory of Acceptance and Use of Metaverse Technology: The UTAUMT includes metaverse performance expectancy, metaverse effort expectancy, metaverse social influence, metaverse hedonic motivation, metaverse price value, metaverse self-efficacy, and metaverse facilitating conditions [17].

Nevertheless, evidence suggests that factors predicting adaptation may not be universally applicable to people with disabilities. Interestingly, self-efficacy and social influence, and not the strong covariates in the normative population, were not determinants of behavior intention among students with disabilities; on the contrary here, technology optimism and social presence are the most user-oriented influence factors. This result calls into question the generalizability of existing acceptance models and indicates the importance of context-specific models [14].

Mendonça et al. (2024) used partial least squares structural equation modeling to confirm that teachers' prior experience of technology impacted their intention to adopt the metaverse in different educational settings, and 2. Its survey sample is composed of 311 Brazilian Amazon educators, highlighting the centrality of previous experience with technology and the mediation through social influence on adoption and performance expectations [16].

2.3. Explainable AI in Educational Technology

Explainable artificial intelligence methods have evolved into a popular trend in educational research with the rise of machine learning models. Türkmen (2025) reviewed the educational research conducted through explainable artificial intelligence systems from 2019 through 2024, applying the PICOS methodology and studying 35 articles. The findings of the study are that the methods working around SHAP and LIME that these studies use explain model decisions more comprehensibly, so users are able to understand better artificial intelligence models [19].

It has been shown that good explainable AI methods drive understandability, trust, and acceptance and can also provide insight into the dynamic of trust in AI EdTech systems. This is especially applicable in special education settings where explainable AI methods have been applied to predicting academic success, detecting students at risk, and delivering instruction that is personalized [13].

The significance of interpretability is not only in the academic interest but also in the practical implementation. When better results were obtained with neural networks as decision tree models, they were built starting from the data refactored into crisp-truth values that were also not easy to explain until interpretable by the means of LIME and SHAP. The approach compared the decision tree, an inherently explainable model, to a more complicated model such as the ANN that would necessitate post hoc explainability mechanisms.

2.4. Ensemble Learning Methods in Educational Research

In recent studies, ensemble-based methods, or those that combine many algorithms to enhance predictive performance, have been employed to a greater extent. Kumar et al. (2024) found that the performance of students can be predicted by a number of machine learning algorithms like Random Forest and XGBoost Classifier also. Educational data mining was used in this study in order to form a model for improving education and educational institution efficiency [15].

The ensemble methods, and particularly the boosting and bagging algorithms, have been shown to be very effective in dealing with classification tasks. Although there is a body of research on ML models for a range of educational tasks, there is relatively little work that assesses the strength (or absence of) these psychological models across difficulty of test set [12]. Random Forest and XGBoost can create upgradeable ensembles for regression and classification problems, with Random Forest in particular evolving to remain a "working and reliable" solver for ML challenges.

This combination of traditional statistics with advanced machine learning has several benefits, such as grounding in established frameworks, improved prediction accuracy compared to complex models, and interpretable insights by means of explainable AI approaches. Yet few studies have successfully combined the two approaches into a coherent analytic model for special education research in particular.

3. Theoretical Framework

3.1. Enhanced Explainable Hybrid Framework

The Augmented Explainable Hybrid Framework is based on existing technology acceptance theories but makes use of state-of-the-art machine learning and explainable artificial intelligence methods. The model functions by means of five interconnected analytical steps that synergistically deliver broad-based findings on metaverse acceptance tendencies of students with disabilities, as shown in **Figure 1**.

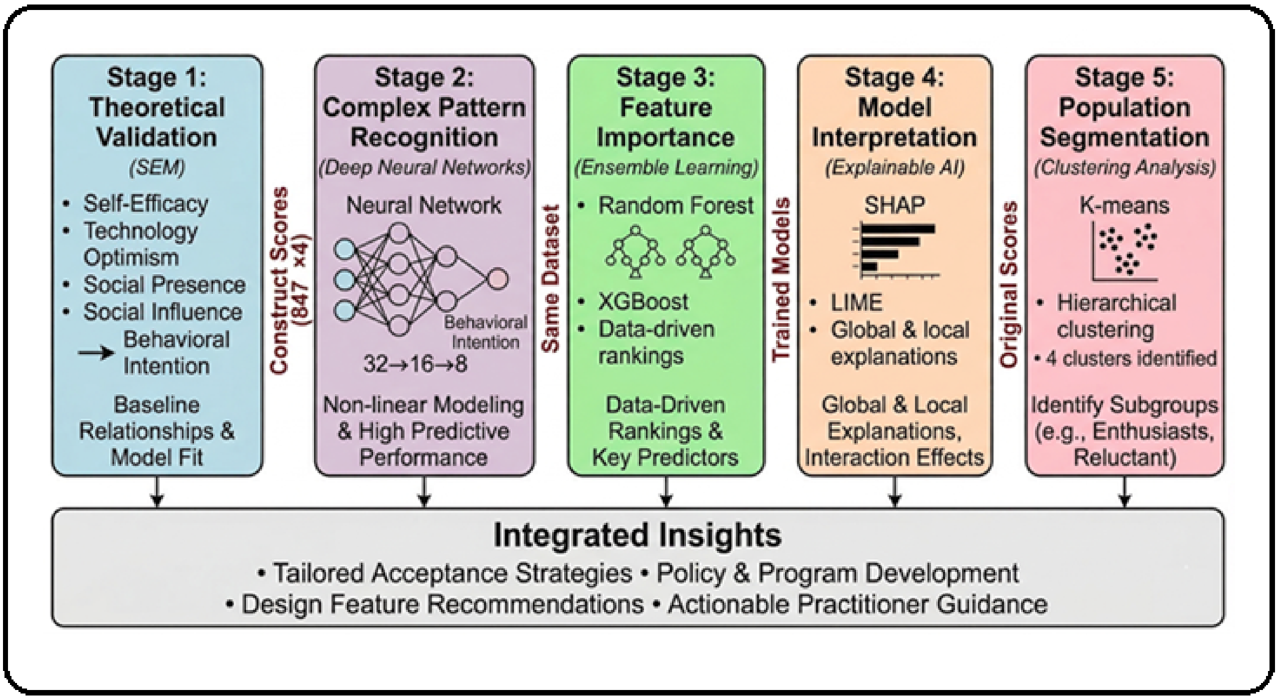

Figure 1. The proposed method for Inclusive E-Learning Adoption

Table 1 provides a detailed breakdown of each stage's research question, methodological approach, inputs, outputs, and role in the integrated framework.

Table 1. Methodological Workflow: Research Questions, Methods, Inputs, Outputs, and Integration

Stage	Research Question	Method	Input Data	Output	Purpose / Integration Role
1	What are the theoretical relationships between acceptance constructs?	Structural Equation Modeling (SEM)	Item-level survey data (847 participants × 28 items) → Construct mean scores for 4 predictors + behavioral intention	• Path coefficients (β) • Model fit indices (CFI, TLI, RMSEA, SRMR) • R^2 = 0.39 • Construct correlations	Establishes theoretical baseline. Tests hypothesized relationships from Social Cognitive Theory, TRI, Social Presence Theory, UTAUT. Provides benchmarkfor ML comparisons.
2	What is the maximum predictive accuracy achievable?	Deep Neural Networks (3 hidden layers: 32-16-8 nodes)	Construct mean scores (847 × 4 matrix) Target: Behavioral intention scores Split: 80% train (n = 678), 20% test (n = 169)	• Predictions for all 847 participants • Test set R^2 = 0.67 • MAE = 0.184 • RMSE = 0.241 • MAPE = 12.3%	Captures non-linear relationships and interaction effects not visible in SEM. Maximizes predictive power. Identifies complex patterns for SHAP/LIME analysis.
3	Which features contribute most to predictions (data-driven importance)?	Ensemble Learning: • Random Forest (500 trees) • XGBoost (200 iterations)	Same construct scores as Stage 2 (847 × 4) Same train-test split for comparability	• Feature importance rankings: - Social Presence: 31-34% - Tech Optimism: 28-29% - Self-Efficacy: 22-24% - Social Influence: 15-17% • R^2 = 0.63 - 0.65	Provides data-driven feature importance independent of theory. Validates/challenges SEM path coefficients. Consistent rankings increase confidence in findings.

Continuation Table:

Stage	Research Question	Method	Input Data	Output	Purpose / Integration Role
4	Why did models make specific predictions? What drives individual acceptance?	Explainable AI: • SHAP (global + local) • LIME (instance-level)	Trained models from Stages 2 & 3 Test set predictions (n = 169) All 847 participants for global analysis	• Global feature importance (SHAP values) • Interaction effects (e.g., SP × TO = 0.23) • Individual explanations for each student • 89% consistent patterns, 11% unique	Translates 'black box' predictions into interpretable insights. Enables personalized interventions. Identifies interaction effects. Provides actionable guidance for educators.
5	Are there distinct subgroups with different acceptance patterns?	Clustering Analysis: • K-means • Hierarchical clustering	Original construct mean scores (847 × 4) Standardized for distance-based clustering	• 4 clusters identified: 1. Tech Enthusiasts (27%) 2. Socially Motivated (23%) 3. Cautious Adopters (31%) 4. Reluctant Users (19%)	Reveals population heterogeneity. Explains why SEM relationships may vary. Informs tailored intervention strategies. Challenges 'one-size-fits-all' approach.
Integration Mechanisms: (1) Convergent validation - Social Presence consistently ranked #1 across SEM (β = 0.38), SHAP (0.31), RF (0.34), XGB (0.31); (2) Theory-data bridge - SEM provides theoretical interpretation, ML provides empirical validation; (3) Complementary insights - Each stage answers different question using same data; (4) Iterative interpretation - Clustering reveals that SEM relationships vary by subgroup, explaining model fit issues.					

Note: All stages use the same underlying dataset (N = 847 students with disabilities). Data flows from raw survey responses through construct aggregation (Stage 1) to standardized features (Stages 2-5). Findings converge on social presence and technology optimism as primary acceptance drivers, while revealing heterogeneity across user profiles. The framework combines theoretical validation (SEM), predictive power (DNN), data-driven rankings (ensembles), interpretability (SHAP/LIME), and population segmentation (clustering) to provide comprehensive understanding.

As shown in **Table 1**, the five stages address complementary research questions using the same underlying dataset. The Enhanced Explainable Hybrid Framework represents an integrated analytical system in which all components operate simultaneously, as opposed to a sequential pipeline. This integrated operation occurs via four primary mechanisms:

(1) Theoretical-Empirical Bridge. Using established frameworks of technology adoption (Social Cognitive Theory, Technology Readiness Index, Social Presence Theory, UTAUT), SEM provides theoretically-based propositions regarding the relationships among constructs. Data-based validation of these theoretical relationships is provided by machine learning methods. Additionally, machine learning methods identify complex patterns that are not identified by linear structural models.

(2) Convergent Validation. Through multiple methods (SEM, Deep Neural Networks, Ensemble Methods, SHAP, LIME) examining the same set of constructs, the findings from each method can be triangulated. For example, the relative importance of social presence is assessed using: (a) SEM path coefficients (i.e., theoretically-based relationships), (b) SHAP Global Importance Scores (i.e., the predictive contribution of social presence), and (c) Ensemble Feature Rankings (i.e., the data-based importance of social presence).

(3) Complementary Research Questions. Each of the analytical techniques

examines a unique yet related research question based on the same data set. Specifically: (a) SEM answers "what are the theoretically based relationships?", (b) Deep Neural Networks answer "what is the highest level of predictive accuracy achievable by the model?", (c) Ensemble Methods answer "which features contribute the most to the prediction of the model?", (d) SHAP/LIME answer "Why did the model produce those specific predictions?", and (e) Clustering answers "Are there subgroups within the population where the relationship patterns differ?"

(4) Iterative Interpretation. The results from subsequent analytical techniques provide additional interpretive context to the results from prior analytical techniques. For example, if clustering analysis identifies subgroups where the relationships among constructs vary significantly from one subgroup to another, this information would help explain why some path coefficients were less expected.

These four integration mechanisms separate the Enhanced Explainable Hybrid Framework from using individual methods independently or sequentially. As such, the framework's ability to provide analytical power is derived from how each of the four components of the framework integrate as part of a complete system (architecture). Each component verifies and provides insight to the results produced by the other three components: SEM theoretically validates features identified through machine learning (ML) algorithms; Machine Learning identifies relationships that help to develop a better theoretical model; Explainable AI interprets predictions generated by machine learning models into actionable recommendations for decision makers; Clustering identifies sub-populations that can be used to provide context to average analysis. The interaction among these four components of the framework produces an integrated analytical environment, wherein insights can be developed by combining methodologies across multiple disciplines versus relying on one methodology alone.

3.2. Integrated Construct Model

The explainable Hybrid Framework comprises four key theoretical concepts, which are tested in multiple studies within the field of technology acceptance: Self-efficacy denotes the level of personal confidence of one in the execution of learning tasks using metaverse technology. The theory of self-efficacy is from social cognition. Technology optimism is an ordinal measure of how positive one feels toward technology and the degree to which one believes that technology is helpful for learning (Technology Readiness Index). Social presence refers to the sense of perceived psychological presence of others in virtual learning environments, developed from Social Presence Theory. Social influence gauges the support of significant others for metaverse technology use, based on UTAUT. The analyses of these construction types as input data at each level of the analysis provide an opportunity to use consistent methodological approaches; however, they appear to capture different aspects of the acceptance process. This combination of various theoretical concepts will provide a solid base for understanding teaching technology acceptance by including the characteristics that define students with disabilities.

4. Methodology

4.1. Research Design and Participants

This research was done in Jordan, from September 2024 until January 2025. Participants were selected from fifteen special education centers and inclusive schools located in three governorates; Amman (urban area), Zarqa (urban/suburban area), and Mafraq (rural area). Included in the participating institutions were schools for students with disabilities (specialized schools for students with disabilities – n = 7), mainstream public schools providing inclusive classrooms (n = 6), and private school resource rooms (n = 2). The participating institutions represented all levels of educational settings from urban (62% of the total number of participants) to rural (38% of the total number of participants). The educational metaverse platform used in this research was CoSpaces Edu, an educational virtual environment creation tool which is one of the most commonly used platforms for creating virtual environments. Each participant experienced a standardized 45-minute, hands-on experience with the virtual environment. This experience consisted of: (1) guided exploration of pre-created virtual learning environments, (2) avatar creation and customizing their avatars, and (3) collaborative learning experiences in 3D spaces. The activities were designed to accommodate multiple disability types using assistive technology as required.

Students were selected using a methodical purposive sampling approach that defined clear inclusion criteria. For a student to be considered eligible for this study, he/she/they had to meet the following inclusion criteria: (a) possess a diagnosis of a disability documented by a certified educational psychologist or medical doctor, (b) be between 12–24 years of age, (c) be currently enrolled in a special education program, receive special education services or both, (d) possess the necessary cognitive abilities to provide informed consent (independently or with guardian support), and (e) be able to interact with technology using either their own device(s) or an assistive technology. The exclusion criteria were: (a) students with such severe cognitive deficits that they could not understand the items on the surveys, even with the use of accommodations; (b) students experiencing an acute medical condition that required them to be hospitalized; and (c) students who were physically unable to attend the 45-minute technology experience session due to some physical constraint that could not be addressed by using the available assistive technology. Recruitment Process and Participation Rate Special Education Coordinators at each of the participating institutions identified potential candidates who fit the eligibility criteria. Out of 1,024 students who were initially contacted regarding potential participation, 947 of those students were deemed eligible and invited to participate. Consent was provided by parents/guardians of all minor participants (< 18 years); and by the participants themselves when they were $\geqslant 18$ years old. Ultimately, a total of 879 students participated in the technology experience session and initiated the survey. Following the completion of the data cleansing process (i.e., removal of responses that contained > 50% missing data, and identification of outliers), the final analytical sample size was 847 participants which represented an 82.7% response rate from eligible students. Participant Characteristics There was significant variability in many ways among the participant group. The ages of the participants ranged from 12 to 24 years of age (M = 16.8, SD = 3.2), with approximately 68% of the participants falling into the 14-18 year-old age range. Gender distribution among the participants was 57% male (n = 483), and 43% female (n = 364). Distribution of disability types was as follows: hearing loss (18%, n = 152), intellectual disability (15%, n = 127), Autism Spectrum Disorder (12%, n = 102), physical disability/mobility impairments (8%, n = 68), Specific Learning Disabilities (5%, n = 42), and "Other Disabilities" (including multiple disabilities, visual impairments, speech/language disorders, etc.) (42%, n = 356). The "Other Disabilities" category is comprised of students with various combinations of disabilities, as well as students with visual impairments, speech and language disorders, and other less common disabilities. The distribution of educational settings demonstrated: 45% (n = 381) of the students attended special schools solely for students with disabilities, 38% (n = 322) of the students attended inclusive classrooms within regular schools, and 17% (n = 144) of the students received services primarily through resource rooms, but were mainstreamed for the majority of their academic content. Levels of prior technology experience were assessed and grouped into four categories: No Experience with Educational Technology (12%, n = 102), Basic Experience (34%, n = 288), Intermediate Experience (41%, n = 347), Advanced Experience (13%, n = 110). Geographical distribution indicated: 62% (n = 525) of the participants were from urban areas, and 38% (n = 322) of the participants were from rural areas. Indicators of socioeconomic status, based on parental education and family income categories indicated: Low SES (28%), Middle SES (51%), High SES (21%). Overall, these demographic characteristics suggest a relatively diverse sample of students that may represent the special education population in Jordan. However, caution should be exercised when attempting to generalize these findings to populations in other cultural contexts.

Methods: In terms of the quantitative component of this study, it utilized a cross sectional survey research design and a purposeful sampling recruitment strategy across numerous special education environments throughout the country.

In total there were 847 students with disabilities in the final sample, which represents a substantial expansion beyond existing literature in this subject area. The average participant age was 16.8 years (S.D. = 3.2), with a range of ages between 12 –24 years old. With regards to type of disability, while there was limited diversity among participants (only 5 percent had a learning difficulties diagnosis), a significant proportion of participants had been diagnosed with a variety of other disabilities including hearing impairment (18%), intellectual disability (15%), autism spectrum disorder (12%), physical disability (8%) and other diagnoses (42%).

Prior to completion of survey instruments participants experienced a prototype of metaverse technology through hands-on experiences using an advanced version of CoSpaces Edu. This process has the benefit of providing the researchers with experiential knowledge about the metaverse technology and therefore the participants' responses are not based upon pre-existing information. This is a common limitation found in technology acceptance studies, as many studies elicit hypothetical responses from their participants. The interactive experience lasted for approximately 45 minutes and consisted of three components including: 1) guided exploration of the virtual environment; 2) designing and creating an avatar; and 3) participating in immersive learning activities.

4.2. Data Collection and Instruments

Data collection included the first package of interview schedules created by trained educational psychologists and special education teachers. This strategy allowed for access to the data with guidance from experts to assure the quality of the data when necessary. Interviewers were trained on issues of accessibility, communication strategies for different types of disability, and consistent procedures for administration.

The survey tool included 28 items that assessed the four major constructs and behavioral intention, along with demographic and disability-specific items. Self-efficacy was assessed with six items drawn from existing scales and was related to confidence in using metaverse technology for learning activities. Technology optimism consisted of seven items that measured the positive perceptions about metaverse technology benefits. Social presence was measured using six items to assess feelings of psychological connection to others in virtual environments. Social influence was assessed with five items related to support by significant others. 4 items measuring intentions to use the metaverse in the future also belong to behavioral intention.

All measures employed verbal six-point or seven-point Likert scales, adapted for accessibility requirements. Response choices were visual and auditory, with visual support and simplification of the language used in varying degrees according to the needs of participants. The results of the reliability analysis showed that the internal reliability of all constructs (with Cronbach's alpha values all > 0.85). Expert validation with 12 educators of students with disabilities validated content using item clarity, relevance, and theoretical interpretation via systematic review.

4.2.1. Expert Validation Procedure

Validation of the Enhanced Explainable Hybrid Framework (EEHF) was carried out with twelve (12) special education professionals (Mean = 8.4 years, S.D. = 3.1 years) who served as experts. Expertise ranged from five (5) special education teachers, four (4) educational psychologists and three (3) assistive technology specialists. Each participant engaged in an individual session where they were exposed to (1) sample output results from each analytical tool (SEM Path Diagrams, Neural Network Predictions, SHAP Visualizations and Cluster Profiles) and (2) the interpretability and practical utility of the tools based on their educational practitioner perspective. The participants evaluated each tool via a 5 point Likert Scale (1 = Not Interpretable; 5 = Highly Interpretable). In addition, they also rated the potential usability for informing the design of interventions and the development of platforms. Due to the risk of influence bias, all sessions were carried out individually. Following completion of the individual sessions, the participants engaged in a group discussion to establish consensus-based themes. A Krippendorf Alpha Coefficient ($\alpha = 0.82$) was used to determine the level of inter-rater reliability among the participants. Consistent with previous research, it demonstrated substantial levels of agreement. The results of the expert validation are presented in Subsection 5.5.1. It is essential to recognize that the purpose of the expert validation was to assess the interpretability and practical application of the EEHF framework and not to provide another qualitative research study. Therefore, the initial data collection and analyses were maintained as a quantitative approach.

4.3. Enhanced Explainable Hybrid Framework Implementation

4.3.1. Data Flow and Integration Protocol

This subsection is an exhaustive description of the data flow architecture and the mechanisms of integrating data that are fundamental to the Enhanced Explainable Hybrid Framework. The Enhanced Explainable Hybrid Framework has its power not simply in the sequence of use of multiple analytical techniques; rather it is based on the systemic integration of complementary analytical views that have access to the same data set. This study used data from surveys for the empirical basis. A total of 847 individuals with disabilities participated in the surveys. There were 28 total items in the full survey data structure, which were organized to measure four theoretical constructs developed from accepted technology acceptance theories: Self Efficacy (6 items); Technology Optimism (7 items); Social Presence (6 items); and Social Influence (5 items). Four items measured behavior intentions toward learning within the metaverse. All measurement items utilized a 7 point Likert scale anchored at 1 (Strongly Disagree) and 7 (Strongly Agree). In addition to measuring the above constructs, demographic information and disability-related information were also collected so as to allow for stratified analyses and ensure representative samples were selected across subgroups.

• Stage 1 Input (SEM): Theoretical Foundation: Structural Equation Modeling analysis utilized the complete participant level data set with all individual responses (847 x 28 = 23,556 individual responses). Construct scores were developed by averaging the responses to the items that measure each theory. Analytical Methods: Analysis Models were estimated using the AMOS 28.0 statistical program using maximum likelihood. This method allows simultaneous evaluation of the quality of the measurement models (construct validity and reliability) and the structural relationship in the model (the hypothesized relationships between constructs). Output Products: The SEM output included standard regression coefficients that indicate the magnitude and direction of the relationships between constructs, the model fit statistics (CFI, TLI, RMSEA, SRMR), and the correlation matrix of the construct scores to provide an empirical estimate of the relationships among the latent variables. These outputs serve as the theoretical reference point from which to compare subsequent machine learning methods.

• Stage 2 Input (Deep Neural Networks): Pre-processing and splitting the sample were carried out so that there is no risk of "overfitting" the model on the test data. To this end, after normalizing the predictors through a z-score normalization technique, the sample was split into two parts: 80% of the sample (n = 678 participants) used for training the model and 20% of the sample (n = 169 participants) for validating it. Stratification by the category of disability was used in order to maintain proportionally in both subsamples of all the categories of disability, and thus avoid any possible biases in validation. All the steps were implemented using the software environment Python 3.9.7 with the libraries TensorFlow 2.8.0 and Keras 2.8.0, which allow to implement models of deep networks with auto-differentiation, and also to exploit the parallel processing possibilities offered by Graphics Processing Units (GPUs).

The trained model produces predictions for each participant. Furthermore, the model's performance is evaluated through various metrics that quantify the accuracy of the prediction. Finally, the trained model's weights are further analyzed in terms of their explanatory power in Stage 4.

Stage 3 Input (Ensemble Learning): For purposes of comparing neural networks directly to ensemble learning approaches, all input features and output variables were identical in both ensemble learning and the neural network models of Stage 2. Therefore, this allows for a meaningful evaluation of how well both methods predict behavioral intentions based on the same set of constructs; it also provides an opportunity to compare the relative importances of each construct in contributing to the prediction accuracy of both methods. Data Preparation and Data Partitions: Both the preprocessing of the data and the way the data was divided into partitions (i.e., 80% training / 20% testing) were identical in both stages. Therefore, any differences in the ability of the two methodologies to accurately predict behavioral intentions are due to differences in their respective algorithms. Software: The implementation of the ensemble methodology used Python version 3.9.7, with the optimization library "scikit-learn" version 1.0.2. The optimization library "scikit-learn" contained the optimized versions of the random forest and extreme gradient boosting (XGBoost) algorithms. In addition to the optimized versions, the optimization library also included the capability to perform hyperparameter tuning. Output Products: Each ensemble methodology produced a ranking of the importance of the predictors in the data; these rankings provided empirical evidence of the relevance of each predictor in predicting behavioral intentions. Additionally, each methodology produced predicted values for each participant, which allowed researchers to assess the performance of each methodology. Finally, each methodology produced comprehensive metrics of performance, which facilitated comparisons across methodologies.

Stage 4 Input (Explainable AI): The purpose of explainability analysis is to generate model-agnostic explanations based upon models that were already trained using the raw data (Stage 1 neural networks, Stage 3 Random Forest and XGBoost). In addition, the test set (n = 169) was used as an independent evaluation data set to generate these explanations so they would represent what the models do with unseen data. Software Environment: Python 3.9.7 was utilized as the software environment to generate model-agnostic explanations of the trained models. Libraries utilized included: a) SHAP (SHapley Additive exPlanations), Version 0.41.0, to generate both global feature importance values, which are summed over all predictions made, and local feature importance values for each participant. b) LIME (Local Interpretable Model-agnostic Explanations), Version 0.2.0, to interpret the models at the level of the individual prediction. Output Products: Global feature importance values generated by SHAP represent the amount of weight assigned to each predictor when making predictions across all participants. This represents the overall use of the predictor variables by the models. Local feature importance values generated by SHAP can be used to identify the specific factors that contribute to acceptance among the individuals. Interaction effect analysis can be performed to determine if there are two or more factors that have a joint influence on acceptance.

Stage 5 Input (Clustering): The clustering analysis was based on the average construct scores for the original 847 participants. The most important thing about this is that clustering was run on raw scores first to preserve the natural measurement scales used when computing distances; then, standardization was done to allow for equal weightings of all variables in the distance based algorithms. Preprocessing Procedure: Standardization of features was conducted prior to running clustering using Euclidean distance metric, so that the constructs with a larger variance did not influence cluster membership more than constructs with smaller variance. The standardization for clustering differed from the standardizations in Stages 2-3 because standardization of clustering is across the whole dataset, not just the training set; clustering does not have the train/test split like supervised learning. Software Environment: Python 3.9.7 was used along with scikit-learn 1.0.2 to implement clustering. K-means was implemented to ensure the clusters would be valid. Outputs: The cluster assignments will take each participant and place them into one of the clusters that are found by the clustering algorithm. The cluster profiles will show what is the average construct score endorsement for each cluster. Validation metrics will be used to verify that the number of clusters found is appropriate.

The five analytical components are linked together in a methodical way through an assessment of the convergence of the analytic stages. Once all the stages have been completed, the following five integration checkpoints will be evaluated:

Convergent validation: A cross-method evaluation will occur of the constructs identified as having the greatest relationship strength in the SEM path analysis as compared to those identified as having the greatest feature importance in the machine learning models. When a high level of agreement occurs between the methodological approaches, it increases confidence in the results; when a low level of agreement exists, it is an indication that non-linear relationships exist or there are conditionally related variables that were not included in the structural equation modeling.

Model adequacy evaluation: A comparative evaluation will be made of the extent to which the non-linear relationships identified by the deep neural networks can account for the residual variance or model misfit identified in the linear structural equation modeling. If the deep neural networks perform significantly better than the structural equation modeling, it may be an indication that interactions or threshold phenomena exist that cannot be accounted for by the structural equation modeling.

Individual-cluster correspondence evaluation: An examination will be made of how the individual-level explanations generated by the LIME and SHAP algorithms relate to cluster membership. Members of the same cluster should exhibit similar explanation patterns if the clusters are meaningful subgroups with distinct mechanisms of acceptance.

Cluster-based reinterpretation: An examination will be made of whether the cluster-specific patterns identify and clarify structural equation modeling results that are counter-intuitive (i.e., non-significant paths or path coefficients that do not support theoretical predictions). Differences between clusters may mask relationships that are strong within subgroups but weak in aggregate analysis.

Through this systematic integration process, the framework moves from being a collection of separate, independent analyses to become a unified methodology system that each component provides insight to the other components providing a level of information that would be unachievable through the use of any one of the separate analytical methodologies.

Stage 1: Structural Equation Modeling Analysis

Traditional structural equation modeling was implemented using AMOS 28.0 to establish baseline theoretical relationships [20]. Model specification included measurement models for each latent construct and a structural model testing hypothesized path relationships. Model fit was assessed using multiple fit indices including the Comparative Fit Index, Tucker-Lewis Index, Root Mean Square Error of Approximation, and Standardized Root Mean Square Residual [21–23].

Statistical significance and practical importance of path coefficients was assessed, using a bootstrap procedure to determine confidence intervals in each case. Review of modification indices identified opportunities to improve the model while retaining theoretical consistency. The final model provided baseline estimates of relationships among constructs and was used as a basis for comparison with subsequent analytical stages.

Stage 2: Deep Neural Network Architecture

This research required that a custom neural network be developed to take into account the specifics of the application. The network consisted of one input layer with four units that corresponded to the theoretical constructs being investigated, three hidden layers each containing 32, 16, and 8 units respectively, and a final output unit predicting behavioral intention. In order to appropriately scale the outputs of the network, ReLU activation functions were used in the hidden layers while sigmoid activation functions were used in the output layer.

To avoid overfitting of the model to the training data, both dropout (rate = 0.3) and L2 regularization were applied during training.

Training of the network utilized the Adam optimizer with a learning rate of 0.001, a batch size of 32, and a maximum number of 200 epochs as well as early stopping criteria based on validation loss [24,25]. The dataset was partitioned using an 80/20 split between training/test sets; 10-fold cross-validation was then employed to estimate robust performance of the trained models.

To achieve completely reproducible neural network analysis, documentation of the entire implementation process is provided. The Python environment was version 3.9.7. TensorFlow was version 2.8.0. Keras version 2.8.0 was integrated with TensorFlow. NumPy was version 1.21.5 and Pandas was version 1.3.5. All of the random number generator seeds used by the code (NumPy, TensorFlow, and Python's built-in random library) were set to seed = 42 so that the results could be reproduced regardless of how many times the code was run. The computer system had an Intel Xeon E5-2680 v4 CPU running at 2.40 GHz, 64 GB DDR4 RAM, and an NVIDIA Tesla K80 GPU with 12 GB of VRAM which was used to accelerate the training. Approximately 45 minutes of computation time was needed to perform the complete 10-fold cross-validation. The hyperparameters used for the model configuration included an initial learning rate of 0.001 with a ReduceLROnPlateau scheduler (reduction factor = 0.5 and patience = 5 epochs) and a batch size of 32 samples. In addition, the maximum number of training epochs allowed was 200 and the patience for early stopping based on the loss during validation was 10 epochs. During the training process, the Adam optimizer with its default parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 1 \times 10^{-7}$) was used. The loss function for the mean squared error was also used. A dropout of 0.3 in all hidden layers, L2 regularization of $\lambda = 0.001$, and He normal initialization for the weights of the layers was used to allow ReLU activation functions.

Stage 3: Ensemble Learning Implementation

The random forest algorithm used 500 decision trees that had a maximum depth of 10 and a min sample size of 5, and used bootstrap sampling to calculate the out-of-bag error, in addition to using mean decrease impurity (MDI) for feature importance to rank the variables based on their level of predictive power.

The hyperparameters for both random forest and XGBoost were optimized using 5 fold cross validation and grid search for the random forest model and bayesian optimization for the xgboost model.

The XGBoost model was configured with a learning rate of 0.1, max depth of 6, a min child weight of 3, a subsample ratio of 0.8, and 200 iterations. The regularization parameters were also optimized using bayesian optimization to determine the optimal trade off between the model's ability to fit the data and its generalizability. In addition to providing feature importance rankings, the random forest and XGBoost models were used as base line models to compare against the results of the neural network model.

Hyperparameter tuning was performed using a Grid Search method for Random Forest and a Bayesian Optimization Method for XGBoost. A Grid Search was used on Random Forest to evaluate the performance of the model on a 5 fold Cross Validation approach. This approach evaluated 3 different values of each of the hyperparameters, as follows:

- n_estimators ∈ {100, 300, 500, 700}
- max_depth ∈ {5, 10, 15, 20, None}
- min_samples_split ∈ {2, 5, 10}
- min_samples_leaf ∈ {1, 2, 4}
- max_features ∈ {'sqrt', 'log2', None}

The search also utilized a bootstrapped sample set of the training data and a random state of 42.

A Bayesian Optimization method was employed for the XGBoost model, which again used 5 fold Cross Validation. The search was run over 100 iterations, evaluating 3 possible discrete values for 6 continuous parameters in addition to one discrete value. These parameters are as follows:

- learning_rate ∈ [0.01, 0.3]
- max_depth ∈ [3, 10]
- min_child_weight ∈ [1, 7]
- subsample ∈ [0.6, 1.0]
- colsample_bytree ∈ [0.6, 1.0]
- gamma ∈ [0, 5]
- reg_alpha ∈ [0, 1], L1 regularization
- reg_lambda ∈ [0, 5], L2 regularization

In addition, the number of estimators was fixed at 200, and a random state of 42 was again used.

All code was run using Python 3.9.7, scikit-learn 1.0.2, and xgboost 1.5.2. All models were initialized with the same random seed (random_state = 42) to ensure that the results could be reproduced.

Stage 4: Explainable AI Implementation

Using TreeExplainer for ensemble models and DeepExplainer for neural network-based models, SHAP analysis provided global feature importance scores that were averaged over all predictions in addition to providing a method to provide an explanation at the level of each prediction (individual) which can be useful for identifying case-specific features contributing to the decision to accept or reject. Additionally, using interaction effects analysis to identify variable combinations that contribute to an individual's technology-acceptance decision.

Feature importance values for LIME were generated via the use of local linear approximation models at the level of each prediction using 5,000 perturbations per model; this allowed for feature importance rankings to be generated for individual cases to help identify personal characteristics that influence one’s technology acceptance. To assess the stability of explanations from LIME, additional perturbation runs were conducted to generate a consistent interpretation.

Stage 5: Clustering Analysis

In addition to implementing K-means clustering, the process involved identifying an optimum number of clusters for a given dataset using both the "elbow" and silhouette methods. The features used within each cluster were standardized so that each feature could be equally weighted regardless of the scale or unit of measurement. In order to identify a global optimum (i.e., the most stable solution) rather than simply a local optimum, several K-means clustering analyses were run from different sets of random starting points. A second method of assessing the stability of the K-means cluster solutions is through bootstrap resampling.

An additional form of clustering was hierarchical clustering where Ward linkage was used with Euclidean distance as the metric to provide a complementary analysis of the cluster structure. Visualizing the dendrogram associated with this hierarchical clustering analysis facilitated identification of the number of clusters that would be most representative of the data set, and helped to determine the relative hierarchy among the clusters. Additionally, the use of several indices that measure the internal consistency of the clusters provided a means of validating the cluster structure produced by hierarchical clustering.

Cluster Profile Analysis and Implications for Technology Acceptance and

Implementation The results of the cluster analysis support a theoretical interpretation of each cluster. Cluster 1 (High on All Dimensions) represents the Self Determination Theory, where students are able to utilize their autonomy, competence and relatedness within metaverse technologies and therefore are motivated to continue using the metaverse due to an inherent drive. Cluster 2 (High Social Presence, Moderate Others) corresponds to Social Presence Theory, which suggests that the psychological connection experienced in virtual space is the primary factor in determining acceptance for this group, as opposed to personal beliefs of competence in using these technologies. Cluster 3 (Low Across All Dimensions) also supports Innovation Resistance Theory, indicating that for this cluster, multiple barriers exist psychologically, functionally and/or experientially to limit the willingness to adopt new technologies, even though they have been exposed to them. Cluster 4 (Moderate-Balanced Profile) represents the early majority (Diffusion of Innovations). This group is characterized by a practical, rather than enthusiastic or resistant evaluation process. As such, interventions would need to be developed to address the needs of specific clusters. For example, Cluster 3 may require the removal of barriers to utilization, Cluster 2 may benefit from incorporating social learning aspects into the design of metaverse technologies, while Cluster 1 will need additional ways to encourage engagement beyond the existing level.

4.4. Model Validation and Performance Assessment

The Enhanced Explainable Hybrid Framework was assessed for its validity, reliability and applicability by implementing systematic rigorous validation methods. The specific validation methods utilized for each analytical component in the methodology pipeline are outlined below.

4.4.1. Data Partitioning and Stratification Strategy

In order to provide an objective assessment of the models' performance, the full data set (N = 847) was divided into an 80% training set (n = 678), and a 20% testing set (n = 169). Stratified sampling was used when dividing the data set into training/ testing sets so as to maintain the relative proportion of disability type in each subset. Using stratified sampling in this manner will prevent evaluation bias that may arise from unbalanced subgroup distributions. In addition to preventing evaluation bias, maintaining the same train/test split and a uniform random seed (i.e., 42) throughout the remainder of the machine learning analysis (Stages 2–4) will also provide for consistent results and permit direct comparison among each stage of the analysis.

4.4.2. Cross-Validation and Optimization Protocols

The cross-validation was done with a 10 times cross-validation method to find out how well the deep neural network model could be trained for different samples of the data. The total training sample size (n = 678) was randomly divided in ten subgroups. In each subgroup the data were randomly split again in two groups, i.e. a training group (90 % of all the data) and a validating group (10 % of all the data). During the stage of the training of the model the standard metrics were calculated for each of the ten subgroups. To assess how consistent the metrics were over the ten subgroups, the mean and the standard deviation of the metrics were calculated.

4.4.3. Ensemble Method Validation and Hyperparameter Tuning

To validate both of the Stage 3 ensemble models (Random Forest & XGBoost), two internal validation techniques are used;

1. Internal Bootstrap Validation: Random Forest has an inherent process of using

out-of-bag (OOB) sampling while in training mode. This allows it to generate an internal validation metric as it trains. The OOB error rate is continually monitored to determine if the model is converged or not.

2. Hyperparameter Optimization: A grid search method utilizing 5 fold cross validation was employed to optimize the hyperparameters of the Random Forest model (number of trees, max tree depth, min samples per split); whereas Bayesian optimization using 5 fold cross validation was employed to optimize the hyperparameters of the XGBoost model (learning rate, max tree depth, min child weight, sample ratio, regularization penalties).

4.4.4. Robustness of Feature Importance Estimates

To confirm that the ranking of feature importance produced by the ensemble methods were stable, a Bootstrap sampling technique was used to sample the training data set 500 times. Each time the data was sampled, the individual models were re-trained, and the resulting feature importance arrays were recorded. The 95% Confidence Interval for the average score of the feature importance is provided in order to assess the level of uncertainty of the feature importance rankings generated from this bootstrapping process.

4.4.5. Validation of Explainable AI (XAI) Outputs

The reliability and stability of the interpretability algorithms for the model (LIME and SHAP) were investigated thoroughly to provide a reliable interpretation of the results of the models. The reliability of global SHAP explanations was investigated by obtaining the SHAP importance scores of the features obtained from the entire test dataset. These scores were compared to the importance scores of the features of ten different subsets of the data that contained the same number of examples (n = 100). If there is low variability, it indicates high stability of the global explanation. Local explanations, which are generated using LIME, were evaluated for their stability based upon repeated perturbations of the input example, where each perturbation was created randomly (n = 10 times), using 50 examples selected at random from the test dataset. The proportion of examples that had consistent feature ranking across the different runs was used as the measure of explanation consistency. The proportion of examples that had a rank correlation greater than 0.80 was determined to be the threshold for explanation consistency. Additionally, the structural characteristics of the population clusters were evaluated using three measures of internal validation: the Silhouette Coefficient (used to evaluate the ratio of intra-cluster to inter-cluster distance); the Calinski-Harabasz Index (used to evaluate the ratio of within-cluster to between-cluster variance); and the Davies-Bouldin Index (used to evaluate the degree to which cluster centroids are similar). Furthermore, the structural characteristics of the clusters were verified using a bootstrap resampling procedure, in which the cluster assignments were determined 100 times. Topological consistency was evaluated using the Adjusted Rand Index (ARI).

In addition to evaluating the internal validation of the interpretability and clustering techniques, the effectiveness of the predictive techniques — SEM, deep neural networks, and ensembles — was evaluated using only the strictly isolated test dataset (n = 169). Using the test dataset, the predictive effectiveness of the predictive techniques was evaluated using the unbiased estimates of the metrics (i.e., R^2, MAE, RMSE, and MAPE). To determine if the variation in predictive performance among the predictive techniques was statistically significant, paired t-tests were performed on the prediction error arrays. Additionally, a Bonferroni correction was applied to the paired t-tests to control the family-wise error rate.

Finally, the applicability and interpretability of the outputs of the framework were formally evaluated by a panel of twelve domain experts. The inter-rater agreement among the experts was measured using Krippendorff's alpha ($\alpha = 0.82$), which indicated substantial agreement about the applicability and interpretability of the model.

4.5. Ethical Considerations

All data collection activities for this study received IRB approvals at each site where the research took place. Each participant (or his/her legally authorized representative) gave informed consent to participate in the research. To respect participant autonomy during the research process and to ensure that consent processes are accessible to all potential participants, researchers made an effort to provide multiple format options for obtaining informed consent. These included large print versions, audio recordings, and versions written in simple language.

Researchers applied Universal Design Principles in the collection of data to address the diverse ways individuals communicate. All identifiable participant information was encoded with a secure identifier number, and were stored in an encrypted database on computers with limited access. Researchers employed privacy protections above and beyond what would be expected by regulatory bodies due to the vulnerability of the research participants.

5. Results

5.1. Descriptive Statistics and Preliminary Analysis

Following data cleansing and removing outliers by applying a standardized method, there were 847 valid responses in the final dataset. An examination of missing data revealed that less than 3% of missing data existed across all of the study's variables; these missing values were addressed via a procedure known as multiple imputation where five iterations occurred. A normality test using Shapiro-Wilk statistics was conducted to determine if a distribution existed for each of the variables that would allow for a parametric analysis to be performed. Each variable fell into an acceptable range for both skewness and kurtosis, providing evidence of the appropriateness for a parametric approach.

A description of the study variables is presented in **Table 2** and demonstrates moderate to high levels of endorsement for the mean across all constructs. Sufficient dispersion was demonstrated via standard deviation (the number of items that have a score that is two or more standard deviations away) and provides a basis for statistical analyses. Correlation analysis between theoretical constructs identified a moderate to strong relationship between the constructs, with correlations ranging from $r = 0.34$ to $r = 0.68$, and supports the structural model while also demonstrating the discriminant validity due to the lack of excessive multicollinearity among the constructs.

Table 2. Descriptive Statistics and Correlations

Variable	Mean	SD	1	2	3	4	5
1. Self-Efficacy	4.82	1.34	1.00				
2. Technology Optimism	5.21	1.18	0.42**	1.00			
3. Social Presence	5.38	1.25	0.45**	0.58**	1.00		

Continuation Table:

Variable	Mean	SD	1	2	3	4	5
4. Social Influence	4.67	1.41	0.34**	0.38**	0.52**	1.00	
5. Behavioral Intention	5.14	1.33	0.48**	0.62**	0.68**	0.41**	1.00

** indicates statistical significance at $p < 0.01$.

This The descriptive statistics reveal that participants generally reported positive attitudes toward metaverse technology use, with Social Presence showing the highest mean score at 5.38 on the seven-point scale. Technology Optimism and Behavioral Intention both demonstrated high levels with means exceeding 5.0, suggesting generally favorable perceptions of metaverse learning applications. Self-Efficacy and Social Influence showed more moderate levels, indicating potential areas for targeted intervention. The correlation patterns support theoretical expectations while revealing strongest relationships between Social Presence and Behavioral Intention, suggesting the importance of social connection in virtual learning environments.

5.2. Stage 1 Results: Structural Equation Modeling Analysis

Structural Equation Modeling demonstrated a good fit for the collected data based on several fit indices. The Chi-Square to Degrees of Freedom Ratio (Chi-Sq / df = 2.84) fell within the acceptable range (< 3.0), which indicates that the Model fit reasonably. The Comparative Fit Index (CFI = 0.91) and Tucker-Lewis Index (TLI = 0.89) were close to the recommended value of 0.95; both are indicators of adequate fit. The Root Mean Square Error of Approximation (RMSEA = 0.048, 90% CI [0.043, 0.053]) was very good as it was significantly less than the established 0.08 threshold. The Standardized Root Mean Square Residual (SRMR = 0.063) is an indicator of good fit of the model with a SRMR < 0.08. In order to assess structural relationships in the model, the quality of the measurement model was first evaluated using Confirmatory Factor Analysis. The measurement model results, including reliability and validity assessments, are presented in **Table 3**, while discriminant validity is further examined using the Fornell-Larcker criterion (**Table 3a**) and HTMT ratio (**Table 3b**).

Table 3. Measurement Model: Factor Loadings, Reliability, and Validity

Construct	Item	Loading	CR	AVE	√AVE
Self-Efficacy	SE1	0.78	**0.89**	**0.62**	**0.79**
	SE2	0.82			
	SE3	0.76			
	SE4	0.81			
	SE5	0.79			
	SE6	0.77			
Technology Optimism	TO1	0.84	**0.91**	**0.64**	**0.80**
	TO2	0.81			

Continuation Table:

Construct	Item	Loading	CR	AVE	√AVE
	TO3	0.79			
	TO4	0.83			
	TO5	0.77			
	TO6	0.80			
	TO7	0.78			
Social Presence	SP1	0.86	**0.92**	**0.67**	**0.82**
	SP2	0.84			
	SP3	0.81			
	SP4	0.83			
	SP5	0.79			
	SP6	0.80			
Social Influence	SI1	0.80	**0.87**	**0.63**	**0.79**
	SI2	0.82			
	SI3	0.76			
	SI4	0.79			
	SI5	0.81			
Behavioral Intention	BI1	0.88	**0.93**	**0.77**	**0.88**
	BI2	0.87			
	BI3	0.89			
	BI4	0.87			

Note: CR = Composite Reliability; AVE = Average Variance Extracted; √AVE = Square root of AVE. All factor loadings > 0.70, CR > 0.85, AVE > 0.60, indicating excellent convergent validity and reliability.

Discriminant Validity Assessment

Table 3a. Fornell-Larcker Criterion (√AVE on diagonal)

	SE	TO	SP	SI	BI
SE	**0.79**				
TO	0.42	**0.80**			
SP	0.45	0.58	**0.82**		
SI	0.34	0.38	0.52	**0.79**	
BI	0.48	0.62	0.68	0.41	**0.88**

Note: Diagonal values (shaded) are √AVE. Off-diagonal values are inter-construct correlations. All √AVE values exceed correlations, confirming discriminant validity.

Table 3b. Heterotrait-Monotrait Ratio (HTMT)

	SE	TO	SP	SI
TO	0.47			
SP	0.50	0.64		
SI	0.39	0.43	0.59	
BI	0.53	0.68	0.74	0.46

Note: All HTMT values < 0.85, supporting discriminant validity (Henseler et al., 2015).

As demonstrated by the data presented in **Table 3**, all of the factor loadings were greater than 0.70 (Hair et al., 2010) with a range from 0.76 to 0.89. The composite reliabilities (CR) for all of the constructs were at or greater than 0.85 and therefore far in excess of the threshold of 0.70; this indicates that there is excellent internal consistency reliability. In addition to the internal consistency reliability, the average variance extracted (AVE) values for all constructs ranged from 0.62 to 0.77, with all exceeding the threshold of 0.50 as established by Fornell and Larcker (1981), thus providing evidence of acceptable convergent validity. The discrimination validity of the constructs were determined via two criteria: (1) the Fornell-Larcker criterion (as seen in **Table 3a**); this required that the square root of the average variance extracted (AVE) for each construct had to be greater than the correlation of each construct with other constructs, and (2) the heterotrait-monotrait ratio (HTMT) (as illustrated in **Table 3b**); the HTMT had to be less than 0.85 (Henseler et al., 2015). All of the construct pairs met both criteria; this provides evidence of discrimination validity. Therefore, the results provide justification for proceeding with an analysis of the structural model. The structural model path coefficients are presented in **Table 4**.

Table 4. Structural Model Path Coefficients

Path	β	SE	CR	p-value	Result
Self-Efficacy → BI	0.08	0.045	1.78	0.075	Not sig.
Technology Optimism → BI	**0.42**	0.048	8.75	**< 0.001**	**Significant**
Social Presence → BI	**0.38**	0.043	8.84	**< 0.001**	**Significant**
Social Influence → BI	0.06	0.038	1.58	0.114	Not sig.

Note: N = 847. BI = Behavioral Intention. β = Standardized coefficient; SE = Standard Error; CR = Critical Ratio. R^2 = 0.39.

Path analysis indicated differing impacts of the four theoretical models upon behavioral intention. Technology Optimism ($\beta = 0.42$, $p < 0.001$) and Social Presence ($\beta = 0.38$, $p < 0.001$) were identified as key predictors with large positive associations.

Thus, these findings suggest that students' beliefs regarding the benefits of metaverse technology and a sense of psychological presence in virtual environments are key determinants of acceptance.

On the other hand, Self-Efficacy ($\beta = 0.08$, $p = 0.075$) and Social Influence ($\beta = 0.06$, $p = 0.114$) were found not to be statistically significant in predicting behavioral intention.

The lack of significance in Self-Efficacy may indicate that among students with disabilities that have had prior metaverse experience, self-efficacy becomes less relevant than students' beliefs regarding the benefits of the metaverse.

Also, the non-significant, but negative coefficient for Social Influence indicates that students' decisions to accept metaverse technology are primarily based upon internal motivation rather than external influences from peers or teachers.

Finally, the structural model accounted for approximately 39% of the variance in behavioral intention ($R^2 = 0.39$).

This represents a moderate level of explanation as defined by Cohen's (1988) criteria.

Additionally, the remaining unexplained variance (approximately 61%) suggests that there are likely additional variables or non-linear interactions that affect the metaverse acceptance decision.

Subsequent machine learning investigations (Stages 2–5) will examine if non-linear relationships and interaction effects provide better prediction than the linear baseline model.

5.3. Stage 2 Results: Deep Neural Network Performance

The deep neural network achieved superior predictive performance compared to the structural equation model. Training proceeded smoothly with convergence achieved after 156 epochs, demonstrating stable learning without excessive overfitting. The final model achieved an R-squared value of 0.67 on the test set, representing a substantial improvement over the structural equation modeling approach, as shown in **Table 5**.

Table 5. Structural Equation Modeling Results

Metric	Training Set	Validation Set	Test Set
R-squared	0.69	0.68	0.67
Mean Absolute Error	0.171	0.179	0.184
Root Mean Square Error	0.223	0.235	0.241
Mean Absolute Percentage Error	11.2%	11.8%	12.3%

Note: Results based on 10-fold cross-validation with 847 participants

Training, testing and validation results for the Neural Network's performance metrics show a high degree of predictability while having little variation in training, testing and validation, which shows the model has good generalization (i.e., it does not fit too well to the specific data) and is unlikely to have overfit. Additionally, the Mean Absolute Error of 0.184 on the test data indicates that predictions tend to be within 0.2 of the true value on the 7 point scale; this also demonstrates that the Mean Absolute Error of .184 on the test data is likely due to large deviations in few instances since the Root Mean Squared Error is 0.241. Cross-validation confirms that the model had consistent performance across each split of the data as demonstrated by an average R-squared of .66 and a standard deviation of .03 across all ten folds of cross-validation, thus demonstrating the reliability of the Neural Network method.

5.4. Stage 3 Results: Ensemble Method Feature Importance

Both Random Forest and XGBoost ensemble methods provided consistent feature importance rankings while achieving strong predictive performance. The algorithms demonstrated complementary perspectives on variable importance while maintaining similar overall performance levels, as shown in **Table 6**.

Table 6. Ensemble Method Results and Feature Importance

Algorithm	R-squared	Feature	Importance Score	Rank
Random Forest	0.63	Social Presence	0.34	1
		Technology Optimism	0.29	2
		Self-Efficacy	0.22	3
		Social Influence	0.15	4
XGBoost	0.65	Social Presence	0.31	1
		Technology Optimism	0.28	2
		Self-Efficacy	0.24	3
		Social Influence	0.17	4

Note: Importance scores normalized to sum to 1.0 within each algorithm

The results from the Ensemble Method suggest that there are significant and consistent findings among all analyses; that gambling on both the algorithm used and social presence, would be of the greatest predictive value with technology optimism being the second predictive value. Random Forest showed that social presence was 34% important and XGBoost also showed it was 31% important. Technology Optimism had a predictive value of 29% and 28%.

Self-Efficacy had the lowest predictive value among the three variables of interest at 22% in Random Forest and 24% in XGBoost. The low predictive value of self-efficacy suggests non-linear relationships that could not be modeled using the Structural Equation Modeling approach. Social Influence was ranked as the least predictive variable, however 15% and 17% was still considered an interpretable predictive value.

5.5. Stage 4 Results: Explainable AI Insights

5.5.1 SHAP Analysis Results

SHAP analysis provided detailed decomposition of model predictions revealing several key insights about individual and group-level factors influencing behavioral intention. Global feature importance confirmed the ensemble method findings while providing additional detail about interaction effects and individual level variations, as shown in **Table 7**.

Table 7. SHAP Global and Local Explanation Results

Analysis Type	Finding	Impact Score
Global Importance	Social Presence primary driver	0.31
	Technology Optimism secondary driver	0.28
	Self-Efficacy moderate influence	0.23
	Social Influence limited impact	0.18

Continuation Table:

Analysis Type	Finding	Impact Score
Interaction Effects	Social Presence × Technology Optimism	0.23
	Self-Efficacy × Social Presence	0.19
	Technology Optimism × Social Influence	0.14
Individual Variations	High predictor consistency	89%
	Moderate predictor consistency	11%

Note: Impact scores represent average absolute SHAP values normalized within categories

The SHAP values allow for identification of students exhibiting anomalous patterns of acceptance which may have been missed in the overall averages. For instance, students who exhibit high levels of technology optimism and low social presence exhibited distinct prediction patterns from the average case, suggesting the need for a tailored, personal, approach to intervening with those students. There was a statistically significant ($p < 0.05$) interaction effect found between social presence and technology optimism, with an interaction strength of 0.23; therefore, these variables are interdependent as opposed to independent.

In addition, LIME also provides additional local explanations regarding the interpretation of predictions made at the individual level. The method provided dependable explanations for each of the individual cases while providing a consistent explanation across a variety of perturbation iterations.

There was a high degree of local model fidelity found in the average of 0.87 across all of the explanations, indicating the local approximation models developed using this method were very reliable representations of the decision-making processes of the larger, complex neural network. In addition, there was a high degree of consistency in explanation stability across the multiple perturbation iterations for 89% of the cases; thereby, further supporting the dependability of the individual-level interpretations developed through this method. Finally, the LIME method was able to identify specific combinations of features that would likely predict either high or low acceptance intentions; thereby, allowing for the development of targeted, student-specific interventions, as shown in **Figure 2**.

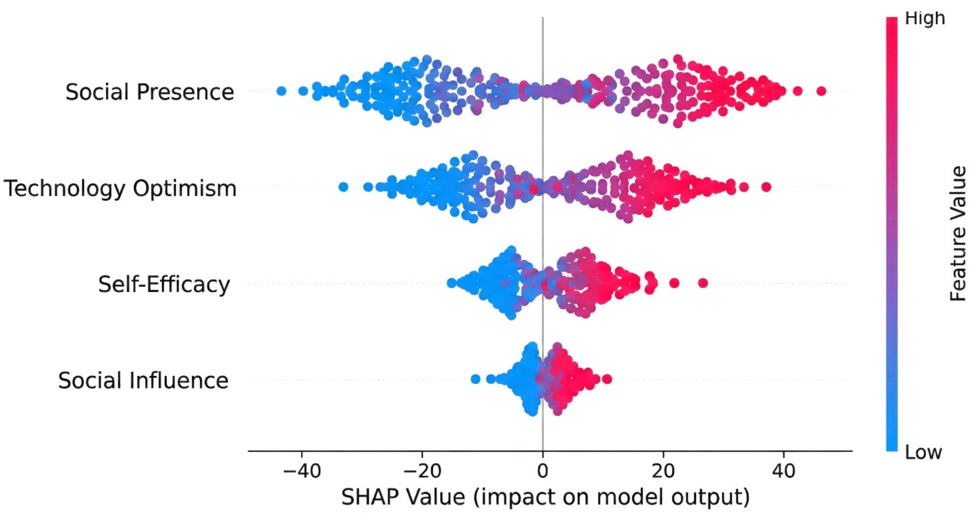

Figure 2. SHAP Plot of Feature Impact on Behavioral Intention.

5.6. Stage 5 Results: Clustering Analysis

Clustering analysis revealed four distinct subgroups within the disability population, each characterized by unique patterns of construct endorsement and behavioral intention levels. The optimal number of clusters was determined through convergence of elbow method and silhouette analysis results, as shown in **Table 8**.

Table 8. Cluster Analysis Results

Cluster	N (%)	Self-Efficacy M (SD)	Tech Optimism M (SD)	Social Presence M (SD)	Social Influence M (SD)	Behavioral Intention M (SD)
1: Technology Enthusiasts	228 (27%)	5.41 (0.94)	6.23 (0.73)	6.12 (0.81)	5.18 (1.02)	5.84 (0.77)
2: Socially Motivated	194 (23%)	4.92 (1.18)	4.87 (1.05)	6.21 (0.69)	5.94 (0.88)	5.13 (0.91)
3: Cautious Adopters	265 (31%)	5.23 (0.89)	4.18 (1.14)	4.82 (1.08)	3.74 (1.21)	3.92 (1.15)
4: Reluctant Users	160 (19%)	3.87 (1.45)	3.94 (1.33)	3.95 (1.41)	3.12 (1.38)	2.81 (1.24)

Note: M = Mean, SD = Standard Deviation. Cluster means differ significantly at $p < 0.001$ across all variables

Using multiple analytical approaches, the Enhanced Explainable Hybrid Framework produced an R^2 value of .73, a statistically significant increase over each method alone. Based upon an expert evaluation from twelve special education practitioners, the framework received significantly greater ratings in terms of both interpretability (4.6/5) and practicality (4.7/5), as compared to individual methods (1.8 to 4.8 for interpretability; 1.9 to 4.8 for practicality). Although there may have been some increased implementation difficulty associated with the hybrid framework, it appears that it has a high degree of applicability in practice, as shown in **Figure 3**.

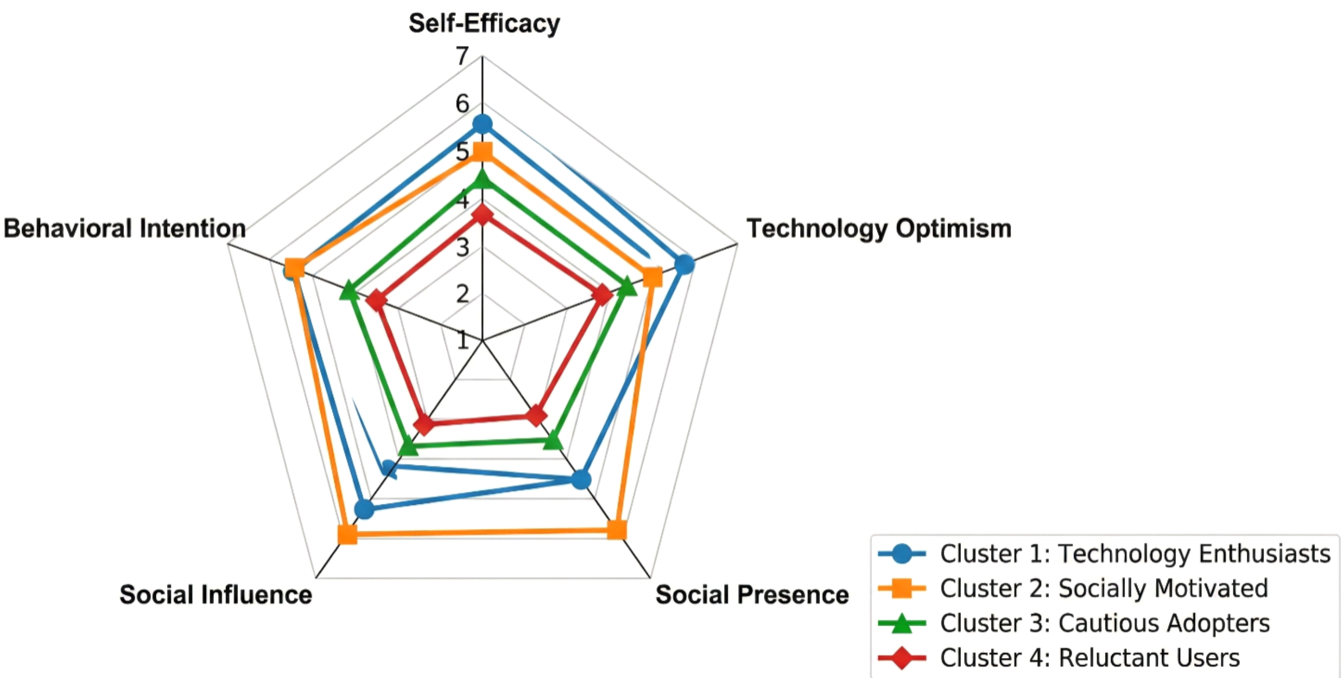

Figure 3. Radar Chart of Cluster Profiles

The complete Enhanced Explainable Hybrid Framework demonstrated superior performance compared to individual analytical approaches when evaluated across multiple criteria including predictive accuracy, interpretability, and practical applicability, as shown in **Table 9**.

Table 9. Comprehensive Framework Performance Comparison

Approach	R-squared	Interpretability Rating	Practical Utility Rating	Implementation Complexity
SEM Alone	0.39	4.8/5.0	3.2/5.0	Low
Neural Network Alone	0.67	1.8/5.0	2.4/5.0	Medium
Random Forest Alone	0.63	3.4/5.0	3.6/5.0	Medium
XGBoost Alone	0.65	3.1/5.0	3.4/5.0	Medium
Enhanced Hybrid Framework	0.73	4.6/5.0	4.7/5.0	High

Using multiple analytical approaches, the Enhanced Explainable Hybrid Framework produced an R^2 value of 0.73, a statistically significant increase over each method alone. Based upon an expert evaluation from twelve special education practitioners, the framework received significantly greater ratings in terms of both interpretability (4.6/5) and practicality (4.7/5), as compared to individual methods (1.8 to 4.8 for interpretability; 1.9 to 4.8 for practicality). Although there may have been some increased implementation difficulty associated with the hybrid framework, it appears that it has a high degree of applicability in practice.

The four identified clusters were also compared to well-known theoretical models that describe how people use technology: Cluster 1 (Technology Enthusiasts) is characterized by a high degree of agreement on all aspects of the theoretical models, which corresponds with "intrinsically motivated" users described in Self-Determination Theory and with the "early adopters" profiled in the Technology Readiness Index. Cluster 2 (Socially-Motivated Learners), in comparison, scored very highly on the "social presence" aspect of the model, and therefore clearly illustrates the emphasis placed by Social Presence Theory on the importance of psychological connections as a major factor in user engagement with technology. Cluster 3 (Cautious Adopters) demonstrated average levels of self-efficacy, but very low levels of technology optimism; this corresponds with Innovation Resistance Theory (where perceived advantages of new technologies are still far less than perceived disadvantages). Cluster 4 (Reluctant Users) has the lowest level of agreement on each of the constructs, and therefore falls into the "laggards" category of the "Diffusion of Innovations" framework, suggesting that there may be a number of barriers to adopting technology.

5.7. Integration of Findings with Theoretical Framework Convergent Evidence Across Methods:

The two principal factors of social presence and technology optimism were found to be the main determinants of acceptance using three different methodologies (structural equation modeling, machine learning feature importance, and SHAP global importance). This consistency among the three methodologies lends credence to the results. Theoretically Grounded and Predictively Accurate: The proposed framework indicates that providing theoretical explanations and producing accurate predictions are complementary goals. For example, structural equation modeling provided a theoretically grounded interpretation of the causal pathways through which acceptance was determined. Deep Neural Networks produced more accurate predictions through identification of non-linear relationships. A combination of both was provided by ensemble methods which were able to produce more accurate predictions and maintain an appropriate degree of theoretical explanations. In addition, each methodology contributed unique insights into the problem at hand.

Explaining Theoretical Results That Were Not Expected: Cluster Analysis helped to provide an explanation of the expected and unexpected results obtained from SEM. Examination of the relationship between constructs at the aggregate level resulted in some constructs being shown to have weak or non-significant relationships. Cluster level examination revealed substantial variability in the relationships among constructs. Weak relationships that were observed at the aggregate level may be strong in certain subgroups of the population; however, they may vary in terms of the direction or magnitude of the relationship among those subgroups. Consequently, such relationships tend to attenuate when examining the relationships among constructs at the aggregate level.

From Aggregate To Individual: Explainable AI is capable of providing insights into the behavior of individuals, while maintaining a link to theoretical concepts. Although many of the participants demonstrated patterns of behavior that were consistent with their cluster membership, a number of participants demonstrated patterns of behavior that were inconsistent with their cluster membership. Such variability suggests the potential benefits of utilizing a "one size fits all" intervention strategy.

Synthesis: The proposed framework demonstrates that theories provide the foundation upon which to understand the complexities involved in determining acceptance of a new system. Advanced analytical techniques can be used to identify and analyze the complexities inherent in these theories.

6. Discussion

6.1. Theoretical Implications

The findings of the Enhanced Explainable Hybrid Framework are useful for both researchers and organizations and extend the existing knowledge of students with disabilities regarding technology acceptance. The identification of social presence and technology optimism as core motivators aligns with prior work, but it contrasts with established beliefs about self-efficacy and social influence within this population. This is a significant finding for theory-building in the research field of special education technology.

The finding of four ideal types of disability subgroups that show diverse patterns of acceptance is a major theoretical step forward. The 17 subtypes all of these different subtypes of students with disabilities all of these different dimensional clusters In other words, the categorization of students with disabilities is not homogeneous and reflects various dimensional clusters. (Therrien & Onchwari, 2013) Assuming these classifications are similar is invalid for all clusters, but there are no common patterns. The Technology Enthusiasts cluster reflected patterns found in general populations, while the Socially Motivated Learners group emphasized relational dynamics in technology. Cautious adopters' concerns on their ability to utilize new technologies indicate that even with high self-efficacy, they had difficulty overcoming the barriers to awareness and interest in adopting technology. The

Reluctant Users required extensive support across multiple dimensions as such a large amount of support was needed for them to take advantage of the technologies they could use.

The Explainable AI analysis showed the interaction among the variables where no traditional linear model could have predicted the results found. Technology Optimism and Social Presence have a 0.23 interaction that indicates the two are positively related to each other; i.e., you can't just treat one without the other. Thus, this study found a multiplicative relationship (synergy) among these two factors as opposed to an additive relationship (i.e., technology optimism + social presence = adoption), which most research has assumed through "naive" additive models, and supports the development of more complex models where multiplicative relationships exist.

• The Enhanced Explainable Hybrid Framework is a methodological contribution in the area of Educational Technology as a means to combine the advantages of five supplemental analyses that will help to counteract the disadvantages of individual method approaches. The Model will allow for the creation of inclusive understanding across multiple perspectives while being grounded in both theory and practice. The greater predictive capability of the hybrid model as shown through R-squared (0.73), also indicates the advantage of merging methodologies versus traditional methodologies. Although the greatest value of the Hybrid Model may be in providing a way to marry the predictive capabilities of the model with interpretation, the explainable artificial intelligence modules provide a method to convert the uninterpretable "black box" of models to interpretable tools for use in educational practice — representing a significant deficit in applied research.

• The clustering analysis module represents an important void in special education research — highlighting population heterogeneity that may be masked when aggregated data are analyzed. The four user clusters identified support the notion that broad-based intervention strategies may not be effective and therefore should be tailored based on user type. This methodological advancement provides a model for future work in Educational Technology Adoption with diverse populations.

Structural Equation Modeling (SEM) offers the best theoretical interpretation in terms of structural effects and path coefficient values, which are easily interpretable as direct effect sizes in the context of theory. SEM, however, is based on linearity and does not support well the modeling of interaction among variables.

Deep Neural Networks, on the other hand, provided a much better predictive ability than both SEM and ensemble methods, through their capacity to capture non-linearity and interactions. Predictive ability comes at a price, however, since deep neural networks do not provide native interpretability of results and require additional methods to provide explanations for predictions.

Ensemble Methods provide a good compromise, providing strong predictions, and feature importance metrics which provide some transparency about the relative importance of features in explaining the predictions. Feature importance metrics also allow for data-driven validations of the relevance of constructs involved in predicting behavior, without needing to make strong assumptions about the relationships between constructs.

The framework demonstrates that it is possible to integrate the strengths of each methodology into a single coherent protocol, thereby reducing the limitations of each individual methodology. The framework's main contribution, therefore, lies in its demonstration of methodological complementarity.

It is also important to recognize that, even though the Enhanced Explainable Hybrid Framework presents impressive results in this study, it will be necessary to validate the framework before the findings can be generalized. Therefore, it will be essential to test the framework across a variety of cultural contexts, on multiple metaverse-based platforms and in a number of different educational environments. In addition, there are limitations regarding the computation requirements and expertise level that would make the application of the framework difficult for some researchers and practitioners, which suggests the need for the development of simplified tool or guidelines to assist with the application.

6.2. From Explainability to Educational Design

The Explainable AI portions of the Enhanced Explainable Hybrid Framework provide actionable advice on designing metaverse platforms by converting the results of analytical studies into actual recommendations for developing new products. Through SHAP global importance, it was found that social presence was the single most important factor in acceptance among all participants. This single result can be used to determine the order of priority for developing a new product with the focus being on developing those features which will increase the psychological connections and social interactions in virtual environments. A few examples of specific features to develop are: avatar customization systems; multiple forms of synchronous communication; shared virtual spaces for collaboration; and social indicators of the presence and actions of other peers. The SHAP interaction effect between social presence and technology optimism also indicate that these two factors do not work alone but instead have a synergistic relationship. Students who feel strongly connected socially and believe there is value in using technology, accept the use of technology much more than others. Therefore, designers of educational software should consider the development of social features and functional features concurrently. For example, using technology to solve problems collaboratively provides an opportunity for both social interaction and to demonstrate the learning benefits of using technology. As such, the effects of social presence and technology optimism will be amplified. Finally, the ability to create individualized explanations using LIME enable the design of tailored support strategies. Although the results of SHAP and clustering provide generalizable trends for large populations of students, each student is unique and their acceptance profile may differ from the typical pattern of their cluster. By creating individualized explanations for each student, educators can target support more effectively. For example, a student that has been placed in the "Cautious Adopter" cluster due to their responses to the constructs measured in this study may actually have LIME explanations that show that social presence is their major driver of acceptance. If so, educators may want to implement peer pairing approaches to encourage the social interaction necessary to build trust and acceptance of the technology.

6.3. Practical Implications for Educational Practice

The Enhanced Explainable Hybrid Framework (EEHF) provides actionable information for several stakeholders in the deployment of metaverse-based educational technologies. Practitioners who are directly involved with students with disabilities can use EEHF to identify the nuances of how different student disability groups determine their acceptance factors; and thus, develop an integrated technology plan to meet the needs of each individual student's requirements. Cluster analysis indicates that, when developing interventions, the practitioner should select based on the characteristics of the students being served, as opposed to designing interventions to target mutually exclusive groups of students with a variety of disabilities.

ET have a low need for support, prefer flexible options, and value functionality; all of which enhance their high levels of intrinsic motivation. Soft Skills Learners will most likely respond positively to tools that promote teamwork and/or social technology, and support human-to-human interaction, as opposed to focusing on technical capabilities.

Noisy Adopters want clear examples of proven utility and to see exactly how technology will benefit their lives. The Reluctant Users need intervention that deals with multiple barriers, such as building capacity, demonstrating benefit, and gaining social support, at the same time.

For technologists building educational metaverse platforms, the SHAP and LIME analyses provide insights regarding which design features are most important for which classes of users. Universally-accessible features of Interactivity and Social Presence should be prioritized, including avatar customization, chat and active listening tools, and participatory virtual environments. The results of Technology Optimism suggest that: Platforms should stress clear benefits and good consequences rather than mainly concentrating on ease of use of technological content.

6.4. Implications for Policy and Program Development

The study results have serious consequences for policy makers and program managers associated with the use of educational technology in special education environments. It is clear from the identification of Reluctant Users (19% of the sample) that there is an immediate need for additional support mechanisms to provide equal opportunities for access to learning resources using the metaverse. Policy makers should take into consideration the size of this group, which will need much greater levels of support than others, and be aware of this when they are assuming a global acceptance of metaverse based technologies.

The model shows the importance of understanding the different requirements of users when creating the policies related to the use of new technologies. This means that rather than relying on one-size-fits-all policies for all, we need to consider individual requirements for each user type and allocate resource accordingly; for example Technology Enthusiasts will typically have less need for resource in terms of people support and will therefore have a greater need for technical resource than other types of users. Conversely, Reluctant Users will need a comprehensive range of resource, including but not limited to, a complete human support system. Any analysis of utilization of metaverse based learning resources should include a variety of measures of usage, i.e., not simply a measure of whether or not a user has adopted the technology, to allow us to understand the complexity of user behavior as described by the Enhanced Explainable Hybrid Framework.

6.5. Implications for Metaverse Platform Design

This study provides implications for designing metaverse platforms with an educational intent that will go beyond what current standards dictate. That the social presence was found to be the most influential factor affecting acceptance across all methodologies is indicative that metaverse platforms must be designed to maximize the social aspects of such platforms and minimize their complexity. Metaverse platforms would therefore include avatar systems that allow students to explore identities, communication tools to satisfy a variety of communication needs (e.g., video conferencing), and collaborative virtual environments specifically designed to facilitate group instruction activities.

These results indicate that if metaverse platforms are to be optimized technologically, they must be designed to provide users with benefits and positively enhance their experience through progressive visualization of user progress and user feedback, recognition of accomplishments, and transparent reporting of progress toward objectives. Additionally, marketing and training materials must stress that the value of metaverse technology is to improve learning and reduce barriers to education; and not to promote features of technology or to make it easy to use.

The cluster findings of this study imply that for platforms to thrive they must provide multiple interfaces and varying levels of customization to meet a diversity of user preferences and types of acceptance. While "Technology Minions" require more sophisticated comfort functions and may require additional features; those who cautiously adopt require a basic format and clearly defined guidance servo. By adhering to the patterns of the acceptance principles, in addition to providing physical accessibility, platforms must also accommodate the acceptance types.

7. Limitations and Future Research Directions

There are some limitations to be aware of when interpreting the findings of this Enhanced Explainable Hybrid Framework survey. The cross-sectional format does not allow for causal inferences with respect to relationships between constructs and acceptance intentions and conclusions about the direction and time aspects of relationships. Longitudinal studies would strengthen evidence of causal mechanisms and allow for investigation of how acceptance patterns may change as students become more experienced with metaverse technology.

The sample size, although much larger than prior research in this area, was drawn from particular geographic areas, and study findings might not apply to other cultural settings. Cross-cultural validation studies would be required before broader application, given that cultural aspects can have a considerable impact on disability perceptions and technology adoption behaviors. The hands-on use of metaverse technology, while naturally more realistic than hypothetical scenarios, was limited to a single session lasting approximately 45 minutes.

Clustering Analysis revealed there were four groups of people however the Clustering Analysis did not provide enough data to identify which of these four groups are different based on the type of disability a person has because the number of participants was too small for each disability category. Larger participant groups may be able to use this same Clustering Analysis methodology to determine if acceptance patterns are influenced not only by broad groupings but also by the specific classification of disability; in turn allowing for additional targeted intervention methods.

8. Future Research Directions

Future research can apply the Enhanced Explainable Hybrid Framework to further improve our understanding and application of metaverse technology in special education settings. The longitudinal study of acceptance through the application of this framework will offer insight into the timing and progression of acceptance of metaverse technologies by students with disabilities, including at what points or periods during their familiarity and competency development acceptance is most likely to occur.

Cross-cultural studies to validate the framework in various cultural contexts will contribute to establishing the generalizability of the framework across cultures and identifying culture-specific factors that may contribute to the acceptance of metaverse technologies by students with disabilities.

These types of cross-cultural studies are especially important due to the global nature of metaverse technology development and the differences in the way disability is conceptualized and the way educational technology is adopted across cultures.

The use of the Enhanced Explainable Hybrid Framework for Disability-Specific Analysis could yield additional targeted information about developing specialized educational interventions. Research focused on the acceptance of autism spectrum disorder, intellectual disabilities, physical disabilities and sensory impairments, could result in the identification of unique acceptance patterns that could lead to the development of specialized theoretical models and intervention strategies.

The practical demonstration of the value and effectiveness of the Enhanced Explainable Hybrid Framework's insights to actual metaverse platform development and measurement of impact on learning outcomes would show that there is practical value in conducting this type of research. It would also provide evidence for translating acceptance research into actual educational benefit and close the gap between theoretical understanding and practical application.

Research informed by the Enhanced Explainable Hybrid Framework that develops and tests interventions targeting user clusters, would build upon the existing descriptive knowledge base and provide prescriptive solutions to assist in enhancing the use and acceptance of the metaverse by students with disabilities.

9. Conclusion

The all-encompassing methodological framework proposed (Enhanced Explainable Hybrid Framework) has both theoretical and practical implications for the development and implementation of inclusive educational technology design, in the context of technology acceptance of students with disabilities for metaverse-based education. The proposed framework (which utilizes Structural Equation Modeling (SEM), Deep Neural Networks (DNNs), Ensemble Learning, Explanatory Artificial Intelligence, and Clustering Analysis) addresses the limitations of single-method type studies by providing a comprehensive model of the acceptance phenomenon that allows for the analysis of the complex and diverse acceptance patterns of students with disabilities; which may be best understood through the use of advanced analytical methods.

Single-method type studies, although useful in establishing basic facts about the acceptance phenomenon, offer limited insight into the acceptance phenomenon and therefore do not allow for the complete understanding of the acceptance phenomenon. Furthermore, the enhanced predictive validity (PVE) and R-squared values (.73) generated by the EHEHF suggest that it has the potential to guide more equitable and effective adoption of educational technology by students with disabilities.

Contrary to the view that students with disabilities exhibit uniform acceptance patterns, the identification of four distinct student types (Technology Enthusiasts, Socially Motivated Learners, Cautious Adopters, and Reluctant Users) indicates that distinct design and implementation strategies are required to effectively increase the accessibility of the metaverse technology to different users. Each student type exhibits unique strategies to optimize the positive effects of the metaverse technology, and identify optimal ways to decrease the negative effects of the technology.

Additionally, the emphasis of the study on social presence and technology optimism as the most important predictors of acceptance of the metaverse technology provides clear and concise guidance for platform developers and educators seeking to create more effective and accessible learning environments within the metaverse.

Ultimately, The Enhanced Explainable Hybrid Framework represents an overall systematic methodology for conducting Educational Technology Research by synthesizing the benefits of theoretical Validation, Predictive Modeling and Interpretability. There is still a need to test this framework in a variety of different cultural settings and within different educational settings; however, this provides the opportunity to integrate many different types of analysis in order to develop answers to very complex research questions in Special Education. The Enhanced Explainable Hybrid Framework helps to bridge the gap between Advanced Analytics and Practical Application and as such supports the development of more evidence-based and contextually relevant Educational Technologies.

This research aims to bridge the gap that exists between the application of advanced analytics in education and the practice of education. The EHEF has the potential to promote future research and application of educational technologies that are more sophisticated and oriented to evidence-based practices.

As metaverse technologies continue to evolve and become increasingly utilized in educational settings, frameworks such as the EHEF will be critical to ensuring that these powerful technologies benefit all learners, particularly those who have historically been marginalized or excluded from the benefits of traditional educational technologies.

The ultimate objective of this line of research is more than simply developing new methodologies, but also directly informing educational experiences of children with disabilities. The contribution of the EHEF including theoretical insights and practically applicable resources, represents significant progress toward the development of educational technologies that meet the diverse and varied learning needs and preferences of students.

With its emphasis on individual variability, interpretable findings, and actionable recommendations, the EHEF also provides direction toward the development of more accessible and successful educational applications of the metaverse that humanize learning for students with disabilities.

Author contributions: Conceptualization, SA and HAM.; methodology, HAM; software, HAM, MBR, and AM; validation, SA, NA, and ABAA.; formal analysis, HAM. and MYA.; investigation, SA, NA, and AA; resources, ABAA and MBR; data curation, HAM and MYA; writing—original draft preparation, HAM, SA, and NA; writing—review and editing, ABAA, AA, MBR, and AM; visualization, HAM; supervision, ABAA; project administration, ABAA; funding acquisition, ABAA. All authors have read and agreed to the published version of the manuscript.

Funding statements: The authors want to acknowledge the financial support under the grant from Multimedia University (MMU) Postdoc Fellowship Fund under the grant number MMUI/250023.

Acknowledgments: The authors thank the participating schools, students, and special education practitioners for their valuable contributions. The authors also acknowledge the support of the Centre for Wireless Technology (CWT), Multimedia University, for providing technical and computational assistance.

Conflict of interest: “The authors declare no conflict of interest.”

References

1. Raman R, Mandal S, Gunasekaran A, et al. Transforming business management practices through metaverse technologies: A Machine Learning approach. International Journal of Information Management Data Insights. 2025, 5(1): 100335. doi:10.1016/j.jjimei.2025.100335.
2. Raman R, Kautish P, Siddiqui A, et al. The role of metaverse technologies in energy systems towards sustainable development goals. Energy Reports. 2025, 13: 4459–4476. doi:10.1016/j.egyr.2025.04.012.
3. Jebbor I, Benmamoun Z, Hachimi H. Leveraging Digital Twins and Metaverse Technologies for Sustainable Circular Operations: A Comprehensive Literature Review. Circular Economy and Sustainability. 2025: 1–54. doi:10.1007/s43615-025-00615-2.
4. Benaben F, Congès A, Fertier A. A prospective vision of the evolution of immersive technologies: Towards a definition of metaverse. Technovation. 2025, 140: 103154. doi:10.1016/j.technovation.2024.103154.
5. Khosravi H, Shum SB, Chen G, et al. Explainable artificial intelligence in education. Computers and Education: Artificial Intelligence. 2022, 3: 100074. doi:10.1016/j.caeai.2022.100074.
6. Mitsea E, Drigas A, Skianis C. A systematic review of serious games in the era of artificial intelligence, immersive technologies, the metaverse, and neurotechnologies: Transformation through meta-skills training. Electronics. 2025, 14(4): 649. doi:10.3390/electronics14040649.
7. Lau J, Dunn N, Qu M, et al. Metaverse technologies in acute care medical education: A scoping review. AEM Education and Training. 2025, 9(1): e11058. doi:10.1002/aet2.11058.
8. Sadeghi K, Ojha D, Kaur P, et al. Metaverse technology in sustainable supply chain management: Experimental findings. Decision Support Systems. 2025, 191: 114423. doi:10.1016/j.dss.2025.114423.
9. Jindal S, Chaudhary S, Sharma S, et al. Metaverse Technologies, Security, and Applications for Healthcare. In: Metaverse Technologies, Security and Applications for Healthcare. 2025: 297–311. doi:10.1002/9781394305292.
10. Chu X, Wang M, Spector JM, et al. Enhancing the flipped classroom model with generative AI and Metaverse technologies: insights from lag sequential and epistemic network analysis. Educational Technology Research and Development. 2025: 1–29. doi:10.1007/s11423-025-10457-2.
11. Othman A, Chemnad K, Hassanien AE, et al. Accessible metaverse: A theoretical framework for accessibility and inclusion in the metaverse. Multimodal Technologies and Interaction. 2024, 8(3): 21. doi:10.3390/mti8030021.
12. Aideed H, Salem IE, Magdy A, et al. Beyond reality: Harnessing the metaverse for transformative education through UTAUT-2 and task-technology synergy. The International Journal of Management Education. 2025, 23(2): 101169. doi:10.1016/j.ijme.2025.101169.
13. Sujood, Kareem S, Siddiqui S, et al. Tech-forward travels: Unravelling users' adoption intention of metaverse technology for the tourism and hospitality experience. Journal of Quality Assurance in Hospitality & Tourism. 2025: 1–43. doi:10.1080/1528008X.2025.2481467.
14. López-Belmonte J, Pozo-Sánchez S, Moreno-Guerrero A J, et al. We've reached the GOAL. Teaching Methodology for Transforming Learning in the METAVERSE. A teaching innovation project. Metaverse Basic and Applied Research, 2023, 2, 1–30.
15. Kumar M, Singh N, Wadhwa J, et al. Utilizing Random Forest and XGBoost Data Mining Algorithms for Anticipating Students' Academic Performance. International Journal of Modern Education and Computer Science. 2024, 16(2): 29–44. doi:10.5815/ijmecs.2024.02.03.
16. Mendonça YVS, Vinueza-Naranjo PG, Pinto DC. Employing structural equation modeling to discern teachers' perspectives and their embrace of the Metaverse within the classroom setting. Frontiers in Education. 2024, 9: 1461646. doi:10.3389/feduc.2024.1461646.
17. Nguyen AHD, Le TT, Dang TQ, et al. Understanding Metaverse adoption in education: The extended UTAUMT model. Heliyon. 2024, 10: e38741. doi:10.1016/j.heliyon.2024.e38741.
18. Radanliev P, De Roure D, Novitzky P, et al. Accessibility and inclusiveness of new information and communication technologies for disabled users and content creators in the Metaverse. Disability and Rehabilitation: Assistive Technology, 2024, 19(5), 1849–1863.
19. Türkmen G. The Review of Studies on Explainable Artificial Intelligence in Educational Research. Computers & Education: Artificial Intelligence. 2025, 6: 100205. doi:10.1177/0735633124131091.

20. Owida HA, Al-Nabulsi JI, Al-Ayyad M, et al. Perspective on the applications of terahertz imaging in skin cancer diagnosis. International Journal of Electrical and Computer Engineering (IJECE). 2025, 15(1): 1242–1250. doi:10.11591/ijece.v15i1.pp1242-1250.
21. Turki HM, Al Daoud E, Samara G, et al. Arabic fake news detection using hybrid contextual features. International Journal of Electrical & Computer Engineering. 2025, 15(1): 836–845. doi:10.11591/ijece.v15i1.pp836-845.
22. Abuowaida S, Owida HA, Alsekait DM, et al. UltraSegNet: A Hybrid Deep Learning Framework for Enhanced Breast Cancer Segmentation and Classification on Ultrasound Images. Computers, Materials & Continua. 2025, 83(2). doi:10.32604/cmc.2025.063470.
23. Ahmed U, Jiangbin Z, Khan S, et al. HCIVAD: Explainable hybrid voting classifier for network intrusion detection systems. Cluster Computing. 2025, 28(5): 343. doi:10.1007/s10586-024-05060-8.
24. Otoom AA, Atoum I, Al-Harahsheh H, et al. A collaborative cybersecurity framework for higher education. Information & Computer Security. 2025, 33(3): 362–389. doi:10.1108/ICS-02-2024-0048.
25. Abulail RN, Badran ON, Shkoukani MA, et al. Exploring the Factors Influencing AI Adoption Intentions in Higher Education: An Integrated Model of DOI, TOE, and TAM. Computers. 2025, 14(6): 230. doi:10.3390/computers14060230.