Evaluating Virtual Reality Immersion Models and Ethical Implications

Grace Sun

2024-04-29

Introduction

Imagine waking up one day to find that everything you've ever experienced is part of a sophisticated virtual reality simulation. With rapid advancements in virtual reality (VR) technology, a scenario straight out of "The Matrix" might soon be indistinguishable from our everyday lives. Virtual reality crafts immersive three-dimensional environments that users can interact with through specialized hardware, such as headsets and handheld sensor-equipped controllers.[1] By the end of 2023, the number of VR users in the United States was expected to have increased to 70 million, illustrating a growth rate of over 30% in just three years.[2] This rapidly evolving and expanding field will significantly impact the human experience, yet it raises serious ethical questions about our relationship with reality. As we stand on the brink of this technological evolution, highlighted by innovations like the Apple Vision Pro, which is expected to sell 350,000 units in 2024 alone, it's clear that the future of VR is unfolding faster than we might have anticipated.[3] In their paper, "Immersion Metrics for Virtual Reality", Matias N. Selzer and Silvia M. Castro outline their methodology for a user study to gather data about the immersive qualities of VR headsets and their subsequent work on developing predictive models for assessing immersion levels. This paper critiques their methods and, through the lens of utilitarianism, argues that despite some concerns, occasional VR use greatly enhances the human experience by providing entertainment, educational value, and mental health treatment support.

Analysis of Methods

To frame this discussion, it is essential to note that the research by Matias N. Selzer and Silvia M. Castro is currently in its preprint stage and has not yet undergone peer review or been published in an academic journal. Through their research and analysis, the authors aim to develop predictive models for the immersion levels of virtual reality (VR) headsets. They conducted a comprehensive user study featuring a single-subject design, where one participant completed 401 successful trials to generate a robust dataset for analysis. Using this data, Selzer and Castro crafted 17 preliminary models (only 11 of which were unique) through direct methods and feature selection techniques, followed by cross-validation of each unique model. In this section, I will leverage the publicly available dataset from their study to conduct a novel analysis in an attempt to validate their methods and results. Subsequently, I will offer critiques of their approach and suggest improvements.

Novel Analysis

To validate the authors' described methodology, I will follow the three-stage process of model generation, selection, and cross-validation using the R programming language. To begin, I generated five initial models as specified by the authors – simple linear, simple with interactions, complete without interactions, complete, and manual. The distinction between simple and complete models lies in the inclusion of second-order terms in the latter. The manual model incorporates seven hand-selected variables, interactions, and second-order terms. The adjusted R^2 values of my five initial models and the authors' five initial models are comparable, but there are discrepancies in the number of coefficients that will be further discussed in the following section.

---

[1] https://www.britannica.com/technology/virtual-reality
[2] https://www.statista.com/statistics/1017008/united-states-vr-ar-users/
[3] https://www.statista.com/statistics/1398458/apple-vision-pro-shipments/

Table 1 Stage 1 Models Comparison (Selzer and Castro)

Model Name	R^2 Adjusted	Predictors	Coefficients
Simple Linear	0.4121	22	25
Simple with Interactions	0.5647	22	299
Complete without Interaction	0.4423	22	34
Complete	0.5999	22	308
Manual	0.4182	7	14

Table 2 Stage 1 Models Comparison (Sun)

Model Name	R^2 Adjusted	Predictors	Coefficients
Simple Linear	0.4139	22	23
Simple with Interactions	0.5011	22	254
Complete without Interaction	0.4423	22	34
Complete	0.5321	22	265
Manual	0.4361	7	35

Next, I undertook feature selection on the complete model with interactions to derive a more streamlined model. It is difficult to follow the authors' methods here since they do not specify which programming language they utilize in their statistical analysis. Despite extensive exploration of various R functions, it was challenging to replicate the authors' method due to their focus on four specific p-values. The standard stepwise function in R, **stepAIC**, which adjusts variables based on their AIC values, does not include a parameter option for p-value specification. Consequently, I produced three models at this stage (forward, backward, and stepwise) using this standard function, differing significantly from the authors' twelve-model approach based on generating three stepwise models at four different selected p-values (referenced as A, B, C, and D). Although my backward and stepwise models differ greatly from the authors' models, I reached the same conclusion that the model developed using forward selection was identical to the complete model.

Table 3 Stage 2 Models Comparison (Selzer and Castro)

Model Name	R^2 Adjusted	Predictors	Coefficients
A,B,C,D Forward	0.5999	22	308
A Backward/A Stepwise	0.7014	22	177
B Backward	0.5704	19	40
B Stepwise	0.5925	18	42
C Backward	0.492	13	24
C Stepwise	0.5741	18	39
D Backward/B Stepwise	0.4362	9	15

Table 4 Stage 2 Models Comparison (Sun)

Model Name	R^2 Adjusted	Predictors	Coefficients
Forward	0.5321	22	265
Backward	0.7014	22	126
Stepwise	0.7072	22	133

The authors' final stage of model generation is validation. The models that have been developed in the first two stages are mostly quite large, and it's important to mention that overfitting is a concern here that will

be further discussed in the following section. Cross-validation will help determine how well our models are able to generalize to new data. Following the methodology described by the authors, I performed k-fold cross-validation (k = 10 and 10 iterations for each model) on the seven unique models that I generated from the first two stages of model creation (results in Table 6). This approach highlighted substantial variances between my results and those of the authors (Table 5), even for the initial models that closely followed their methods. The adjusted R^2 values that I manually calculated for each model following cross-validation are all below 0.4, revealing the significant role that overfitting was playing in all of the models. After cross-validation, it's clear that none of these models have a good balance of predictive power and model complexity. The adjusted R^2 values that I calculated for the simple model with interactions and the complete model are both below -1, which means that any variability in the immersion level explained by the predictors is overshadowed by the large number of coefficients in the model.

Table 5 Stage 3 Cross Validation Results (Selzer and Castro)

Model Name	RMSE	R^2 *Adjusted*	MAE
Simple Linear	14.86	0.3869	12.16
Simple with Interactions	30.56	0.1393	24.23
Complete without Interaction	14.71	0.4037	12.01
Complete/A,B,C,D Forward	32.18	0.1403	25.47
Manual	14.55	0.4091	11.94
A Backward/A Stepwise	12.71	0.5973	10.23
B Stepwise	12.62	0.5542	10.29
B Backward	12.99	0.5297	10.59
C Stepwise	12.94	0.5314	10.68
C Backward	13.78	0.4691	11.36
D Backward/D Stepwise	14.36	0.4235	11.75

Table 6 Stage 3 Cross Validation Results (Sun)

Model Name	RMSE	R^2 *Adjusted*	MAE
Simple Linear	14.78	0.3557	12.11
Simple with Interactions	24.94	-1.283	19.69
Complete without Interaction	14.71	0.3266	12.00
Complete	25.61	-1.660	20.01
Manual	14.76	0.3371	12.05
Backward	12.66	0.3940	10.05
Stepwise	12.80	0.3693	10.17

Critique of Methodology

A significant issue with the paper's methodology is the discrepancy between the number of predictors the authors claim to have used in their initial models and the number actually listed in their cited dataset. The paper states that 22 predictors were utilized, yet the dataset contains 23. This discrepancy became apparent during my initial feature selection attempt, which failed to eliminate any predictors, prompting further investigation into the data's integrity. Upon running the **vif** function in R to assess collinearity, I found an extraordinarily high vif value between the variables *width* and *height*—exceeding 3×10^{12}. This level of collinearity, which the authors do not discuss but seemingly acknowledge by excluding the *height* variable in their models, poses a significant concern. Addressing collinearity prior to feature selection is crucial and should be clearly documented in the methods section. The mismatch in the number of coefficients reported by the authors raises questions about the clarity and accuracy of their descriptions, complicating efforts to replicate their findings.

Another troubling aspect of the authors' paper that emerged during my attempt to replicate their findings is the inconsistency between their reported number of coefficients and what their methodology would logically dictate. For instance, the authors describe a manual model with 7 predictors and assert that it includes both second-order variables and interactions between each pair of independent variables. Logically, this should result in 7 coefficients for the first-order predictors, another 7 for the second-order predictors, and 21 coefficients for the interactions, considering there are 21 possible combinations for choosing 2 elements from 7 distinct elements. This totals to 35 coefficients. However, the authors write that this model only includes 14 coefficients as seen in Table 1. Their method description does not clarify how they could account for all these variables with so few coefficients, raising questions about either an oversight in their explanation or a calculation error.

This confusion extends to other models as well; for example, a simple model is said to have 22 predictors but mysteriously lists 25 coefficients, with no clear explanation of where the additional coefficients arise from. After rigorously following the authors' methods, even the adjusted R^2 values from my analysis slightly differ from those reported by the authors, which is perplexing since these models are deterministic and should yield consistent results. The authors clearly reference their dataset in their citations and do not note any deviations from this cited dataset in their methods. The absence of details about the software or programming language used for generating the models further complicates replication efforts. The lack of clarity and detail in the methodological descriptions severely hampers the reproducibility of the study, which explains the discrepancies between the models I generated and those reported by the authors.

Selzer and Castro briefly mention overfitting towards the end of their paper, but this critical issue deserves more thorough attention given the complexity of the models involved. Overfitting occurs when a model fits the training data too closely, failing to generalize effectively to new data. This is particularly relevant here, as the complexity and large number of coefficients in the models suggest that they might perform poorly on unseen data. This suspicion is confirmed during cross-validation, where most of the newly adjusted R^2 values are disappointingly low, indicating significant overfitting for most of the models. Overfitting not only compromises the model's accuracy on new data but also its utility in practical applications, making it less reliable for decision-making purposes. To combat overfitting, the authors employ pruning techniques through stepwise regression, but other techniques, such as regularization, could be implemented to further reduce complexity. Regularization techniques such as Lasso or Ridge regression penalize the coefficients of the model to prevent them from becoming too large.

Another notable concern about the authors' methodology is their reliance on stepwise selection. This method has been widely criticized for overly focusing on individual fit at the expense of the overall model fit (Rose and McGuire 2019). Stepwise selection tends to select variables based not only on their contribution to the model but also on arbitrary statistical thresholds, which can result in the inclusion of irrelevant variables or the exclusion of important ones. Moreover, stepwise methods often ignore the problem of multicollinearity among predictors, which can skew the results and diminish the interpretability of the model. Given the wealth of literature advocating for alternative feature selection methods better suited to high-dimensional data, such as Lasso and Elastic Net, which both regularize the coefficients to prevent overfitting and better handle multicollinearity, the authors' choice to use stepwise selection appears particularly ill-advised (Zhang and Shen, 2010; Lima et al., 2021). A different approach might have yielded more effective models.

In their analysis, the authors identify three final models as the "best" for predicting immersion, with adjusted R² values after cross-validation of 0.5542, 0.4691, and 0.424. Only one of these models surpasses the 0.5 threshold, which is relatively low for a model with over 40 coefficients, indicating limited predictive power. My attempts at replicating their work resulted in even lower adjusted R² values (below 0.4 for all seven models), reinforcing my belief that the methodology described was insufficient for producing effective predictive models. While this outcome does not necessarily indicate a weak link between immersion levels and the hardware and software settings of VR headsets, it strongly suggests that the methods employed in developing these models have led to overly complex yet underperforming predictive tools—barely more effective than random chance. This analysis points to a critical need for refining the model-building process to enhance both accuracy and generalizability in predicting VR immersion.

Despite these critiques, the choice of a single-subject design was appropriate for the study's context, where the

response variable—immersion—is subjectively reported. Using a single subject can enhance the reliability and consistency of results since individual perceptions of immersion can vary significantly. Incorporating multiple participants could introduce variability that obscures the true effects being studied. While the single-subject design is a notable strength, this examination of the author's methods, executed as thoroughly as possible with the given data and descriptions, illustrates the need for more rigorous and clear methodologies that enhance both the reproducibility of research and the accuracy and generalizability of the resulting models, especially in the up and coming field of virtual reality technology research.

Analysis of Normative Consideration

The growth in popularity of virtual reality (VR) technology, with its ability to transport users into imaginative digital worlds, has sparked a myriad of ethical considerations. Drawing on utilitarian philosophy, which posits that actions should maximize pleasure and minimize pain, VR can be seen as primarily beneficial when used in moderation to enrich daily life. However, overindulgence to the extent of hindering one's connection to actual reality brings about significant negative outcomes, diverging from utilitarian ideals. In this section, I argue that the occasional use of VR aligns with utilitarian principles by contributing more positive than negative outcomes to the human experience, tipping the utilitarian scales in favor of VR use.

Virtual reality, much like traditional gaming platforms, offers a unique blend of entertainment and social interaction. Users can engage in a range of activities from collaborative puzzle-solving and competitive gaming to virtual tours of distant landmarks, enriching their social interactions and broadening their cultural exposure without leaving their living space. This is especially valuable for those unable to meet friends in person, allowing them to enjoy three-dimensional shared experiences from the comfort of home. However, the socialization benefits of virtual reality depend on moderate use. The immersive nature of VR can make it particularly captivating, leading users to spend extended periods in virtual environments. Excessive engagement can lead to potential addiction and neglect of real-life responsibilities, transforming a beneficial tool into a source of isolation and discomfort. Unlike in-person interactions, virtual connections lack the depth of physical social bonds, emphasizing the importance of balance in VR usage.

When used as an educational tool, virtual reality technology has the potential to enhance training across various professions and revolutionize learning. By offering realistic, immersive training environments, VR helps develop skilled professionals in high-risk fields, significantly benefiting the broader community. Traditional hands-on training can be costly and challenging to simulate accurately, but VR provides a cost-effective solution by enabling repeated practice in controlled, three-dimensional settings. For instance, surgeons can refine their techniques on high-risk procedures using VR simulations that mimic real-life operations. Similarly, first responders such as firefighters, EMS personnel, and police officers can rehearse emergency scenarios that are difficult to replicate safely in real life. Beyond professional training, VR also transforms traditional education, enabling students of all ages to explore and interact with complex subjects in ways that textbooks and lectures cannot match. By immersively experiencing diverse locations and scenarios, students can ignite new interests and deepen their passions for learning. The widespread use of VR in education not only produces more competent professionals but also fosters a broader passion for learning and innovation, creating a lasting impact on society.

Virtual reality has the unique potential to revolutionize mental health treatment and therapy. With its immersive capabilities, VR serves as a powerful medium for exposure therapy, a psychological treatment that helps individuals confront and overcome fears in a controlled, safe environment. For instance, individuals with phobias can gradually face the objects or scenarios that trigger their anxiety within the comfort of a virtual setting, effectively reducing fear responses without real-world risks. VR can also simulate calming environments to aid in stress reduction and mindfulness practices, offering a safe space from the hectic pace of everyday life. Therapists are also exploring its use in treating conditions like PTSD, enabling patients to process traumatic memories through carefully constructed virtual experiences (Rizzo and Shilling 2017). The customization available in VR means that therapeutic settings can be tailored specifically to each individual's needs, potentially increasing the efficacy of the treatment. This innovative use of VR in mental health not only broadens access to effective therapy but also represents a significant leap in integrating technology with compassionate care.

The use of virtual reality also raises substantial ethical concerns, particularly around data privacy and security. VR systems are uniquely equipped to collect a wide range of personal data from users, capturing detailed information about their behaviors, preferences, and interactions within virtual environments. This capability prompts critical questions about how users are informed about their data being collected and the measures in place to ensure their consent is genuinely informed. Moreover, VR technology distinguishes itself from other digital tools by its ability to gather unique personal data. This includes tracking eye movements and analyzing voice tones, which can reveal more about a person's emotional and cognitive states than traditional data inputs. The potential misuse of this sensitive data could have expansive consequences, making it imperative that there are stringent regulations and ethical guidelines governing the use and storage of data collected through VR. Ensuring the privacy and security of user data in VR platforms is crucial not only for protecting individuals but also for maintaining trust in this rapidly advancing technology.

Excessive use of VR can result in health issues like eye strain, headaches, and motion sickness. While occasional use can mitigate these risks, the effort to improve the immersive nature of VR also presents an important ethical concern. Highly immersive VR, while controlled and idyllic, can lead to an unhealthy addiction that detracts from real-life experiences and responsibilities, blurring the line between reality and simulation. There is a significance to living a life rooted in the real world, however difficult that may be, that cannot be experienced in virtual worlds. The unpredictability and challenges of real life are essential for fostering human qualities such as empathy and resilience that ultimately contribute to fulfillment and long-term joy.

While the use of virtual reality presents both benefits and challenges, it's clear that the advantages significantly surpass the drawbacks when VR is used moderately and not as a primary reliance. The potential negative impacts, such as social isolation, addiction, and a detachment from reality, are mitigated by occasional and purposeful use. In contrast, the benefits of VR, including its applications in entertainment, education, and mental health treatment, significantly outweigh concerns related to data collection and usage. These complex issues can be addressed through the implementation of stringent regulations and industry standards. When used responsibly, VR emerges as a transformative tool that not only expands human capabilities but also fosters innovation. By carefully navigating the potential positive impacts and negative limitations of this immersive technology, we can ensure that VR enhances human experiences in a way that aligns with utilitarian ideals, promoting the greatest good and minimizing harm.

Conclusion

Impact of Paper

Matias N. Selzer and Silvia M. Castro's paper makes a significant contribution to the field of virtual reality by establishing a novel framework for quantifying VR immersion. Their work addresses a critical gap in existing methodologies by developing metrics that account for both hardware and software attributes, providing a comprehensive means to measure user-perceived immersion. The impact of this research is profound as it not only advances our understanding of VR technology but also sets a benchmark for future developments. By enabling a standardized approach to evaluate and compare the immersion levels of different VR systems, this paper lays foundational work that could influence VR system design, user experience optimization, and the broader application of immersive technologies in various industries. The metrics introduced have the potential to foster innovation across educational, entertainment, and therapeutic fields by offering developers and researchers quantifiable data to improve VR interfaces and interaction models, ultimately enhancing the end-user experience.

Future Work & Wrap-Up

Building upon the foundational insights from Selzer and Castro's study, there are several promising directions for future research in the field of VR technology. Firstly, refining and testing the existing predictive models with a broader and more diverse participant base would enhance their accuracy and generalizability. This should include adjustments for the variability in users' self-reported immersion levels. Longitudinal studies could offer valuable information on how immersion levels evolve with repeated VR use, potentially highlighting

issues like desensitization. Previous discussions have pointed out the limitations of stepwise regression for feature selection, so further research should explore alternative feature selection techniques better suited to handling large models. Efforts to develop smaller, more precise models that maintain accuracy and generalizability will ultimately yield more practical and effective tools for assessing VR immersion.

In conclusion, the research conducted by Matias N. Selzer and Silvia M. Castro in "Immersion Metrics for Virtual Reality" serves as a pivotal exploration of how virtual reality immersion can be quantified and enhanced. However, the critique of their methodology reveals significant challenges in model validity and reproducibility that need to be addressed. The overfitting and ambiguity in model construction, particularly in the choice and analysis of variables, highlight the need for more robust statistical techniques and transparency in research processes. Additionally, the normative considerations emphasize the double-sided nature of VR technology. While it offers substantial benefits for entertainment, education, and mental health, these advantages are contrasted by ethical concerns regarding data privacy, user health, and the potential for social isolation. Thus, ensuring the responsible use of VR technology is essential. By recommending occasional use and refining the ethical standards governing its use, VR can truly enhance human capabilities without compromising individual well-being or societal values. Future efforts should aim not only to refine the technical aspects of predicting VR immersion but also to integrate ethical considerations into the development of VR technologies, ensuring they are used to enrich human life rather than serve as an escape from reality.

References

Lima, E., Hyde, R., & Green, M. (2021, January 11). Model selection for inferential models with high dimensional data: Synthesis and graphical representation of multiple techniques. Nature News. https://www.nature.com/articles/s41598-020-79317-8

Rizzo, A., & Shilling, R. (2017, January 16). Clinical virtual reality tools to advance the prevention, assessment, and treatment of PTSD. European journal of psychotraumatology. https://pubmed.ncbi.nlm.nih.gov/29372007/

Rose, S., & McGuire, T. G. (2019). Limitations of P-values and R-squared for stepwise regression building: A fairness demonstration in Health Policy Risk Adjustment. The American statistician. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6602076/

Selzer, M. (2022, February 22). Virtual reality immersion dataset. Mendeley Data. https://data.mendeley.com/datasets/kj79vpcsc5/2

Selzer, M. N., & Castro, S. M. (2022, June 15). Immersion metrics for virtual reality. arXiv.org. https://arxiv.org/abs/2206.07748

Zhang, Y., & Shen, X. (2010, October 1). Model selection procedure for high-dimensional data. Statistical analysis and data mining. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2992390/