ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1971

Externalized Risk Management in AI-Based Video Games: Evidence from Caregivers for Human-Centered System Design

Rosa Aurora Orellana Castillo[1]; Cristian Manuel Calderón[2]; Lesly Nohemy Miranda[3];
Elvia Yamileth Izaguirre[4]; Karla Elizabeth Melgar[5]; Estefany Paola Hernández[6]; Ada Abimael Antúnez[7]
[1,2,3,4,5,6,7]Universidad Tecnológica de Honduras, Honduras, *aurora.castillo@uth.hn, cristian.calderon@uth.hn, lesly.miranda@uth.hn, eyizaguirre@uth.hn, kemelgar@uth.hn, ephernandez@uth.hn, aaantunez@uth.hn*

Abstract*– Video games based on artificial intelligence pose significant challenges for risk management in socio-technical systems, especially when the system design shifts control, monitoring, and decision-making functions to end users. This study analyzes parental mediation as an externalized risk management mechanism, examining how the design of AI-based video games distributes responsibility among technical controls, human intervention, and balanced use strategies. A quantitative cross-sectional design was used with a sample of 309 caregivers. Based on a structured instrument, system-level indicators were constructed to identify patterns of design externalization. The analysis included descriptive statistics, robust comparison tests, and effect size estimates to assess the stability of these patterns across contextual variables in the usage environment. The results indicate that design externalization patterns remain consistent regardless of family environment characteristics. Additionally, a substantial proportion of participants were unable to identify the type of video game used, suggesting a gap between the stated control mechanisms and the actual visibility of the system's operation. The findings reveal a uniform design logic that externalizes risk management to users, underscoring the need for human-centered design approaches that increase the visibility, interpretability, and governance of algorithmic risks in AI-based video games.

Keywords*-- Artificial intelligence, Human-centered design, Risk management, Socio-technical systems, AI-based video games.

Introduction

The incorporation of artificial intelligence (AI) into video games has transformed these platforms into complex socio-technical systems, in which multiple human and technological agents interact within environments regulated by social, organizational, and technical norms. From a systems engineering perspective, AI-based video games can be understood as dynamic architectures where emergent behavior results from the interaction between adaptive algorithms, interaction design, and human decisions, both by players and other actors in the system [1], [2]. This complexity makes video games an ideal context for analyzing how control, responsibility, and risk management are distributed in advanced digital systems.

The integration of machine learning techniques and algorithmic adaptation has enhanced the ability of video games to personalize experiences, dynamically adjust difficulty, and optimize user interaction through mechanisms such as dynamic difficulty adjustment and symbiotic game agents [3], [4]. While these strategies improve engagement and fluidity of the experience, they also introduce system-level risks that are not always apparent to end users. In particular, the immersive nature of AI-driven video games can amplify problematic usage patterns, especially in children and adolescents, as observed in research on problematic gaming and addictive behaviors [5], [6].

Beyond the risks associated with intensive use, AI-powered video games expose users, especially minors, to potentially inappropriate content, unregulated social interactions, and monetization dynamics that can affect their well-being and development. These risks are exacerbated by rapid technological evolution, which challenges the ability of regulatory frameworks to balance innovation with user protection and ethical considerations [7], [8], [9]. In this context, digital systems tend to transfer part of risk management to end users, a phenomenon widely documented in complex systems where security and functionality compete as design objectives [10].

From a user-centered design perspective, the active participation of human actors has been proposed as a strategy for mitigating risks in complex systems. User-centered design and human-centered design approaches emphasize the need to involve users in system management, reducing the likelihood of failures resulting from misunderstandings or inappropriate use [11]. In the case of AI systems aimed at minors, various authors emphasize the importance of designing technologies that are sensitive to developmental stages and capable of avoiding inappropriate responses, ensuring child safety and well-being [12], [13].

However, risk perception and management are not uniform. The literature distinguishes between visible risks, which are easily identifiable due to their direct manifestations, and invisible risks, which are characterized by their low perceptibility, high complexity, and uncertainty [14], [15]. In AI systems, many of the most significant risks, such as algorithmic biases, unforeseen failures, or opaque personalization dynamics, fall into this second category, making them difficult to detect and communicate effectively among the various stakeholders [16], [17].

In the case of video games used by children and adolescents, the management of these invisible risks often falls to caregivers, who act as external components of the control system. Various design frameworks have proposed conceptualizing the caregiver–child relationship as a functional unit within the system, emphasizing collaboration and shared decision-making rather than purely restrictive approaches [18], [19]. These approaches suggest that parental mediation tools should not be understood solely as surveillance mechanisms, but rather as extensions of the system design aimed at supporting education, trust, and shared responsibility.

Despite these conceptual advances, there is limited empirical evidence analyzing how the design of AI-based video games externalizes risk management to caregivers and how caregivers respond to this externalization. In particular, it is important to examine whether parental mediation strategies are adapted to the characteristics of the family context or whether, on the contrary, they reflect a uniform design logic that delegates control in an undifferentiated manner. In response to this gap, this study analyzes patterns of parental mediation as indicators of externalization in the design of video games with artificial intelligence, adopting a socio-technical systems perspective and human-centered design.

It should be noted that the term “AI-based video games” is used throughout this study in an operational sense, referring to commercially available titles that, based on existing literature and documented game mechanics, incorporate adaptive algorithmic components such as dynamic difficulty adjustment, personalized recommendation systems, or AI-mediated interaction features [3], [4]. The analysis does not rely on formal technical verification of the underlying architectures of each title; rather, it is grounded in caregiver perceptions of the video games their children use and, in the system,-level risk indicators derived from the instrument. This operational framing is consistent with socio-technical research approaches that examine how users experience and respond to algorithmically driven systems, regardless of their degree of technical literacy [1], [2].

MATERIALS AND METHODS

A quantitative cross-sectional study was conducted, with data collected between September and December 2025. The study was conducted entirely online as part of academic research activities linked to the psychology degree program in Honduras. The objective was to analyze patterns of parental mediation in the context of the use of artificial intelligence-based video games from a socio-technical systems and human-centered design perspective.

The study population consisted of fathers, mothers, or legal guardians who reported having at least one child. No additional criteria were established regarding the age of the caregiver or children, beyond the requirement of performing a caregiving role. Participation was voluntary and unpaid. Participants were recruited through convenience sampling via digital social networks, including Facebook, WhatsApp, and Instagram.

Recruitment was carried out by research team members affiliated with the psychology degree program, who shared the survey link through their personal and institutional social media accounts. No probabilistic selection procedure was applied, and participation was self-selected. As a result, the sample should be considered non-probabilistic and its representativeness with respect to the broader caregiver population in Honduras cannot be guaranteed. The final sample comprised 309 participants who completed the survey in full, with no missing data in the composite variables used for analysis. Given the online and self-selected nature of the recruitment, potential response biases related to digital access, platform use, and caregiver awareness of video game issues should be considered when interpreting the findings.

The data collection instrument was an original structured questionnaire, developed based on a theoretical review of literature related to risk management in digital systems, user-centered design, and parental mediation in technological environments. The content of the instrument underwent a validation process by experts prior to its application. The final version of the questionnaire consisted of 13 items, with a five-point Likert-type response format, aimed at assessing three dimensions associated with technical supervision, human mitigation, and balanced use strategies. Based on these items, composite indices and design externalization patterns were constructed, in accordance with the criteria defined in the analysis section.

The validity and internal consistency of the instrument were evaluated using a confirmatory psychometric approach. Reliability indices showed high values in all three dimensions, with Cronbach's alpha coefficients between 0.877 and 0.898 and omega (ω) coefficients above 0.87 in all estimates. Convergent validity was supported by average variance extracted (AVE) values above 0.63, while discriminant validity was assessed using the Heterotrait–Monotrait (HTMT) criterion, with values below the recommended critical thresholds. The overall fit of the confirmatory model showed satisfactory indicators (CFI = 0.964, TLI = 0.954, RMSEA = 0.070, SRMR = 0.039), supporting the structural consistency of the measurement model.

Data management included predefined cleaning and quality control procedures. Sample size was determined by availability, based on the scope of recruitment during the collection period. Statistical analyses were performed using Jamovi software (version 2.7.12.0). Descriptive analyses, robust comparison tests, and association tests were used, selected according to compliance with the corresponding statistical assumptions. Effect sizes were reported systematically to complement interpretation beyond statistical significance, setting a significance level of $\alpha = 0.05$. Before completing the questionnaire, all participants provided their informed consent digitally. The study did not receive external funding.

RESULTS

Based on the items in the instrument, four operational indicators were constructed at the system level: the System Risk Exposure Index (SREI), the External Control Load Index (ECLI), the Human Mitigation Dependency Index (HMDI), and the Design Externalization Gap (DEG). Descriptive analyses were performed on a sample of 309 caregivers, with no missing values in the composite variables (Table 1).

The results show consistently high levels across the three main indicators. The SREI had a mean score of 4.21 (95% CI: 4.12–4.30), indicating high perceived exposure to risk signals associated with the use of artificial intelligence-based video games. Similarly, the ECLI (M = 4.17; 95% CI: 4.08–4.27) and HMDI (M = 4.15; 95% CI: 4.06–4.25) reflect a high burden of control and a strong reliance on human mitigation strategies, respectively, suggesting that risk management falls primarily outside the system design.

The DEG showed a mean value close to zero (M = −0.02; 95% CI: −0.09–0.05), indicating an aggregate balance between technical control and human mitigation. Nevertheless, the wide dispersion observed in this variable reveals heterogeneous patterns of design externalization, pointing to substantive differences in how the system transfers responsibility for risk management to caregivers across different contexts of use.

TABLE 1
DESCRIPTIVE STATISTICS OF SYSTEM-LEVEL INDICATORS DERIVED FROM CAREGIVER-REPORTED DATA

	SREI	ECLI	HMDI	DEG
N	309	309	309	309
Missing	0	0	0	0
Mean	4.21	4.17	4.15	-0.0178
95% CI mean lower bound	4.12	4.08	4.06	-0.0899
95% CI mean upper bound	4.30	4.27	4.25	0.0543
Median	4.40	4.25	4.25	0.00
Standard deviation	0.801	0.846	0.819	0.644
Minimum	1.00	1.00	1.00	-4.00
Maximum	5.00	5.00	5.00	3.75
25th percentile	3.80	3.75	3.75	-0.250
50th percentile	4.40	4.25	4.25	0.00
75th percentile	5.00	5.00	5.00	0.250

Note. SREI = System Risk Exposure Index; ECLI = External Control Load Index; HMDI = Human Mitigation Dependency Index; DEG = Design Externalization Gap. Values represent means, 95% confidence intervals, medians, standard deviations, and observed ranges based on a sample of 309 caregivers. All indices were computed as mean scores of their respective item sets. The confidence intervals for the means assume a t distribution with N − 1 degrees of freedom.

Distribution of perceived risk and externalized control burden

The analysis of the distribution of the indicators showed a high concentration of elevated values across the three main indices (Table 1). For the SREI, the 25th percentile was located at 3.80, the median reached 4.40, and the 75th percentile corresponded to the maximum value of the scale (5.00), indicating that exposure to risk signals associated with the use of artificial intelligence–based video games is a widely prevalent condition in the sample. A virtually identical pattern was observed for the ECLI and the HMDI, whose percentiles indicate that most caregivers assume high levels of external control and rely on human mitigation strategies to manage the use of these systems.

The distribution of the Design Externalization Gap (DEG) exhibited a more heterogeneous pattern. Although the median was located at zero, the percentiles revealed a symmetric dispersion around this point, with one quarter of the sample showing a predominance of technical control (DEG < 0) and another quarter showing a greater reliance on human mitigation (DEG > 0). This pattern suggests that the balance observed at the aggregate level masks substantive differences in how the system externalizes risk management to caregivers, pointing to the need for more adaptive design approaches.

Design externalization gap

To examine how responsibility for risk management is distributed between the system and caregivers, the Design Externalization Gap (DEG) was categorized into a three-level Design Externalization Pattern (DEP): predominance of technical control, balance between technical control and human mitigation, and predominance of human mitigation. The distribution of these patterns is presented in Table 2.

The results show a nearly balanced distribution across the three categories. A total of 33.3% of cases were classified under the predominant technical control pattern, indicating contexts in which regulation of use relies primarily on system-level mechanisms. In contrast, 31.1% of the sample exhibited a predominant human mitigation pattern, reflecting a substantial reliance on external interventions by caregivers. The remaining 35.6% fell into a balanced pattern, in which technical control and human mitigation coexist without a clear predominance.

This distribution demonstrates that design externalization does not follow a uniform pattern but varies substantially across contexts of use. Taken together, the findings indicate that the system does not implement a consistent risk management strategy, but rather transfers responsibility differentially to caregivers, suggesting limitations in the design's ability to adapt to diverse interaction scenarios.

TABLE 2
DISTRIBUTION OF DESIGN EXTERNALIZATION PATTERNS (DEP)

DEP	Counts	% of Total	Cumulative %
System Control	103	33.3%	33.3%
Human Mitigation	96	31.1%	64.4%
Balance	110	35.6%	100.0%

Note. DEP = Design Externalization Pattern, derived from the Design Externalization Gap (DEG). Categories represent predominance of system-centered control, balanced control, or human-dependent mitigation. Percentages are based on a total sample of 309 caregivers.

Design externalization patterns by caregivers' age

The association between caregivers' age groups and design externalization patterns (DEP) was examined using a χ^2 test of independence (Table 3). The results did not show a statistically significant association between the two variables, $\chi^2(10) = 12.8$, $p = 0.235$, with a small effect size (Cramér's V = 0.144).

The distribution of externalization patterns was similar across the different age groups, with the coexistence of predominant technical control, balance between technical control and human mitigation, and predominant human mitigation observed in all of them. These findings indicate that the externalization of risk management in artificial intelligence–based video games does not vary systematically by caregivers' age, suggesting that the observed patterns are primarily driven by system design characteristics and contexts of use rather than by generational differences among users.

Table 3
Design Externalization Patterns (DEP) by Caregiver Age Group

Caregiver Age Group	System Control n (%)	Human Mitigation n (%)	Balanced n (%)	Total
18–24 years	14 (34.1)	12 (29.3)	15 (36.6)	41
25–34 years	25 (38.5)	15 (23.1)	25 (38.5)	65
35–44 years	35 (26.5)	48 (36.4)	49 (37.1)	132
45–54 years	26 (41.3)	16 (25.4)	21 (33.3)	63
55–64 years	2 (33.3)	4 (66.7)	0 (0.0)	6
≥65 years	1 (50.0)	1 (50.0)	0 (0.0)	2
Total	**103 (33.3)**	**96 (31.1)**	**110 (35.6)**	**309**

Note. DEP = Design Externalization Pattern. Percentages are calculated within caregiver age groups (row percentages). Association between caregiver age group and DEP was tested using a χ^2 test of independence, $\chi^2(10) = 12.8$, $p = 0.235$, with a small effect size (Cramér's V = 0.144).

Design externalization patterns by caregivers' educational level

The association between caregivers' educational level and design externalization patterns (DEP) was examined using a χ^2 test of independence (Table 4). The results did not indicate a statistically significant association between the two variables, $\chi^2(12) = 17.4$, $p = 0.137$, with a small effect size (Cramér's V = 0.168).

The distribution of externalization patterns was consistent across the different educational levels, with the coexistence of predominant technical control, balance between technical control and human mitigation, and predominant human mitigation observed in all groups. These findings suggest that the externalization of risk management in artificial intelligence–based video games does not vary differentially by caregivers' educational level, pointing to a system design that transfers responsibility in a relatively homogeneous manner regardless of users' educational capital.

TABLE 4
DESIGN EXTERNALIZATION PATTERNS (DEP) BY CAREGIVER EDUCATIONAL LEVEL

Educational Level	System Control n (%)	Human Mitigation n (%)	Balanced n (%)	Total
No formal education	7 (70.0)	2 (20.0)	1 (10.0)	10
Primary incomplete	7 (53.8)	1 (7.7)	5 (38.5)	13
Primary complete	18 (38.3)	12 (25.5)	17 (36.2)	47
Secondary incomplete	8 (30.8)	6 (23.1)	12 (46.2)	26
Secondary complete	31 (32.0)	36 (37.1)	30 (30.9)	97
Bachelor's degree	30 (29.1)	33 (32.0)	40 (38.8)	103
Master's degree	2 (15.4)	6 (46.2)	5 (38.5)	13
Total	**103 (33.3)**	**96 (31.1)**	**110 (35.6)**	**309**

Note. DEP = Design Externalization Pattern. Percentages are calculated within educational level (row percentages). Association was tested using a χ^2 test of independence, $\chi^2(12) = 17.4$, $p = 0.137$, with a small effect size (Cramér's V = 0.168).

Number of children and design externalization patterns

Whether the number of children differed across design externalization patterns (DEP) was examined using a one-way analysis of variance with Welch's correction, given violations of the assumptions of normality and homogeneity of variances. The results did not show statistically significant differences between groups, $F(2, 200) = 0.826$, $p = 0.439$ (Table 5). Mean numbers of children were similar across the technical control pattern ($M = 2.56$), the human mitigation pattern ($M = 2.33$), and the balanced pattern ($M = 2.47$).

Post hoc comparisons conducted using the Games–Howell procedure confirmed the absence of significant pairwise differences among patterns. These findings indicate that the degree of externalization of risk management in artificial intelligence–based video games does not vary as a function of family size, suggesting that the system design does not adapt to the complexity of the domestic environment.

TABLE 5
NUMBER OF CHILDREN BY DESIGN EXTERNALIZATION PATTERN (DEP)

Design Externalization Pattern	N	Mean	SD	SE
System Control	103	2.56	1.59	0.157
Human Mitigation	96	2.33	1.03	0.105
Balanced	110	2.47	1.43	0.136

Note. DEP = Design Externalization Pattern. Statistical test: Welch one-way ANOVA, $F(2, 200) = 0.826$, $p = 0.439$. Differences between groups were evaluated using Welch's ANOVA due to violations of normality and homogeneity of variances. Post hoc comparisons were conducted using the Games–Howell procedure; no statistically significant pairwise differences were observed.

Children's age and design externalization patterns

The association between children's age and design externalization patterns was examined using a chi-square test of independence. The results indicated no statistically significant association between the two variables, $\chi^2(10) = 12.2$, $p = 0.274$, with a small effect size (Cramér's V = 0.140).

The distribution of children's age ranges was similar across the technical control, human mitigation, and balanced patterns (Table 6). These findings suggest that the degree of externalization of control and responsibility in artificial intelligence–based video games does not vary as a function of the child's developmental stage.

TABLE 6
DISTRIBUTION OF CHILDREN'S AGE GROUPS BY DESIGN EXTERNALIZATION PATTERN (DEP)

Children's Age Group	Control (n)	Human Mitigation (n)	Balanced (n)	Total
1–4	8	1	7	16
5–7	8	7	8	23
8–11	24	18	28	70
12–17	50	47	53	150
18–21	9	12	7	28
>21	4	11	7	22
Total	103	96	110	309

Note. DEP = Design Externalization Pattern. Values represent absolute frequencies. The association between children's age groups and design externalization patterns was examined using a chi-square test of independence. No statistically significant association was observed, $\chi^2(10, N = 309) = 12.2$, $p = 0.274$. The effect size, assessed using Cramer's V, was small (V = 0.140), indicating limited practical significance.

Types of video games used by children according to parental report

Table 7 presents the distribution of video game types that, according to parental reports, are used by their children. The most frequent category was "None," with 115 cases (37.2%), followed by sports video games with 73 cases (23.6%) and simulation games with 55 cases (17.8%). Smaller proportions were reported for action/battle games (8.4%), open-world games (7.4%), adventure and role-playing games (2.3%), survival games (2.3%), and strategy games (1.0%).

Overall, these results show a high concentration of responses within a limited number of categories, particularly highlighting the proportion of parents who did not identify a specific type of video game used by their children.

Table 7
Types of Video Games Used by Children According to Parental Report

Type of video game	n	%
Action / battle games (Free Fire, Fortnite, God of War, Devil May Cry)	26	8.4
Adventure and role-playing games (World of Warcraft, Elder Scrolls Online)	7	2.3
Sports games (FIFA, Dream League Soccer, FC Mobile, NBA 2K)	73	23.6
Strategy games (XCOM [TRPG], Civilization [4X])	3	1.0
Open-world games (Minecraft, Grand Theft Auto)	23	7.4
None	115	37.2
Simulation games (Roblox)	55	17.8
Survival games (Age of Empires, StarCraft, Command & Conquer)	7	2.3

Note. Values represent absolute frequencies and percentages of the total sample (N = 309). Categories reflect parental self-reported identification of the primary type of video games used by their children.

DISCUSSION

The results of the study make it possible to introduce a key distinction for understanding parental mediation in contexts of artificial intelligence–based video games: the difference between visible risk and invisible risk. Visible risks are associated with elements traditionally recognized by parents, such as explicit violence, inappropriate language, or intense graphic content. In contrast, invisible risks correspond to less evident features of system design, such as adaptive algorithms, personalization mechanisms, reward dynamics, integrated monetization, and AI-mediated social interaction patterns. This differentiation is consistent with the literature on risk communication and management in complex systems, where risks with low perceptibility tend to be underestimated by end users [14], [15].

The high proportion of parents reporting the use of sports video games, together with the substantial percentage of responses indicating a lack of knowledge about the specific types of video games used by their children, may suggest that parental risk perception is primarily activated in the presence of content displaying clear signals of visible risk. In games perceived as socially acceptable or low risk, such as sports video games, parental attention appears to focus less on the structural characteristics of the system and more on a general trust in the presumed harmlessness of the content. This finding is consistent with prior research indicating that the social acceptance of certain genres may reduce vigilance toward underlying algorithmic mechanisms and their potential emergent effects [7].

This pattern helps explain the coexistence of high self-reported levels of parental mediation with limited knowledge about the actual use of AI-based video games. In these cases, mediation appears to operate within a general preventive framework rather than through informed supervision of the invisible risks associated with the game's algorithmic design. Consequently, elements such as automatic difficulty adaptation, optimization of time-on-task, or AI-mediated social interaction may go unnoticed, even though they have a significant impact on the experience of child and adolescent users [2], [3], [4]. This phenomenon appears to reflect a process of risk externalization, in which the management of complex systems is shifted to users who do not always have sufficient information to assess their implications, as documented in studies on digital risk management [10], [16].

From a human-centered design perspective, these findings have direct implications for developers and designers of AI-based video games. First, it is essential for systems to incorporate mechanisms of functional transparency that make invisible risks visible, enabling parents to understand not only the game genre but also its algorithmic and interaction logics. The literature on responsible AI design emphasizes the need for systems that are sensitive to children's contexts and capable of mitigating inappropriate responses through explicit design decisions [12], [13].

Second, parental mediation tools should be conceived as integrated components of the system, with clear interfaces designed to support informed decision-making, going beyond basic technical controls or time-based restrictions. In this regard, approaches that conceptualize the caregiver–child relationship as a functional unit of the system have shown potential to promote more collaborative and effective mediation [18], [19].

Furthermore, the results suggest that responsible design of AI-based video games should avoid assuming that certain genres are intrinsically "safe." Instead, developers should adopt universal mediation approaches that recognize that risks emerge from system functioning rather than solely from explicit content. This entails offering customization options, understandable feedback, and active support mechanisms that assist families in managing video game use in a balanced manner, aligning with user-centered design principles as a strategy for risk mitigation [11].

From an engineering and system governance perspective, these findings suggest that current AI-based video game architectures implicitly rely on users as active components of risk management without providing adequate system-level transparency or safeguards. This design logic raises relevant AI safety concerns, as critical control functions related to exposure, adaptation, and engagement are externalized beyond the technical boundaries of the system. Incorporating human-centered design principles at the engineering level—such as interpretable algorithmic feedback, adaptive parental control interfaces, and embedded risk signaling mechanisms—could reduce reliance on informal human mitigation and improve accountability within AI-driven interactive systems.

Additionally, the system-level indices proposed in this study (SREI, ECLI, HMDI, and DEG) may serve as practical analytical tools for engineering audits and design evaluation processes in AI-based interactive systems. By operationalizing the degree of risk externalization and human dependency, these indicators could inform internal compliance assessments, user-centered safety reviews, and iterative design improvements aligned with emerging AI governance frameworks. Their application may support evidence-based decision-making during system development, deployment, and post-market evaluation, particularly in technologies designed for vulnerable populations.

Taken together, the study reinforces the need to move toward design models that integrate parental mediation as a structural principle in the development of AI-based video games. By making invisible risks visible and facilitating informed mediation, designers can contribute to the creation of safer, more balanced digital experiences aligned with parental expectations, promoting the responsible use of emerging technologies across diverse family contexts.

LIMITATIONS

Several limitations of this study should be acknowledged. First, the cross-sectional design prevents causal inferences; the patterns observed reflect associations between variables at a single point in time and cannot establish directional relationships between system design features and caregiver behavior. Second, the data are based entirely on self-report, which introduces potential biases related to social desirability, recall accuracy, and subjective interpretation of mediation practices. Third, the sample was obtained through non-probabilistic convenience sampling via digital social networks in Honduras, which limits the generalizability of the findings to other populations, cultural contexts, or countries with different technological infrastructures and regulatory frameworks. Fourth, the classification of video games as "AI-based" relied on caregiver reports and literature-informed criteria rather than direct technical verification of game architectures, which may introduce classification uncertainty. Future research should address these limitations by employing longitudinal designs, objective measures of game use, and probabilistic sampling strategies across diverse cultural settings. Mixed-methods approaches that combine caregiver surveys with technical analysis of game systems could further strengthen the evidence base for human-centered design recommendations in this domain.

CONCLUSIONS

This study analyzed patterns of parental mediation in the use of artificial intelligence–based video games, showing that strategies of supervision, communication, and balance are expressed consistently among caregivers, regardless of their sociodemographic characteristics. The absence of significant differences by age, educational level, family size, or children's age suggests that parental mediation responds to shared functional criteria rather than to structural conditions of the household.

In addition, the results reveal a gap between self-reported mediation and specific knowledge about the video games used by children, particularly in contexts where the content is perceived as low risk. This gap highlights the importance of distinguishing between visible risks associated with content and invisible risks derived from the algorithmic design of AI-based systems.

Taken together, the findings reinforce the need for human-centered design approaches that integrate parental mediation as a structural component, promoting greater system transparency and supporting informed decision-making for the responsible use of artificial intelligence–based video games.

REFERENCES

[1] A. P. Shah and A. R. Pritchett, "Agent-Based Modeling and Simulation of Socio-Technical Systems," Organizational Simulation, pp. 323–367, Jun. 2005, doi: 10.1002/0471739448.CH12.

[2] J. Zhu et al., "Player-ai interaction: What neural network games reveal about ai as play," Conference on Human Factors in Computing Systems - Proceedings, May 2021, doi: 10.1145/3411764.3445307.

[3] Q. Mi and T. Gao, "Novel Dynamic Difficulty Adjustment Methods for Niche Games," Lecture Notes on Data Engineering and Communications Technologies, vol. 214, pp. 189–200, 2024, doi: 10.1007/978-3-031-64766-6_19.

[4] S. P. Sithungu and E. M. Ehlers, "Adaptive Game AI-Based Dynamic Difficulty Scaling via the Symbiotic Game Agent," IFIP Adv. Inf. Commun. Technol., vol. 581 AICT, pp. 107–117, 2020, doi: 10.1007/978-3-030-46931-3_11.

[5] K. Paschke and R. Thomasius, "Digital media use and mental health in adolescents—a narrative review," Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz, vol. 67, no. 4, pp. 456–464, Apr. 2024, doi: 10.1007/S00103-024-03848-Y.

[6] C. Montag, H. Yang, A. M. S. Wu, R. Ali, and J. D. Elhai, "The role of artificial intelligence in general, and large language models specifically, for understanding addictive behaviors," Ann. N. Y. Acad. Sci., vol. 1548, no. 1, pp. 5–11, Jun. 2025, doi: 10.1111/NYAS.15337.

[7] G. Valença, "What Risks and Opportunities Games Bring to Children? A Consultation on Platform Regulation," Communications in Computer and Information Science, vol. 2623 CCIS, pp. 102–111, 2026, doi: 10.1007/978-3-032-01426-9_7.

[8] P. Koukal, "THE RELATIONSHIP BETWEEN THE DIGITAL SERVICES ACT AND THE ARTIFICIAL INTELLIGENCE LIABILITY DIRECTIVE," Jusletter IT, vol. 2025, pp. 63–71, Feb. 2025, doi: 10.38023/84F925EA-4655-40C7-A0FE-06F8E80E0EA4.

[9] D. Saluzzo, "'BIG TECH' RESPONSIBLE BUSINESS CONDUCT Transparency and due diligence obligations for online platforms and safer space for online users' fundamental rights, now and in the Metaverse," Journal of Law, Market and Innovation, vol. 2, no. 1, pp. 59–76, Mar. 2023, doi: 10.13135/2785-7867/7439.

[10] J. Wallner, "Cyber Risk Management," Encyclopedia of Quantitative Risk Analysis and Assessment: Melnick/Risk, pp. 1–11, Jan. 2008, doi: 10.1002/9780470061596.RISK0233.

[11] D. J. Williams and M. Groen, "User-centred design as a risk management tool," Intelligent Systems Reference Library, vol. 80, pp. 107–127, 2015, doi: 10.1007/978-3-662-45704-7_5.

[12] N. Kurian, "Designing Child-Safe Conversational AI: Three Dilemmas for Responsible Design," CUI 2025 - Proceedings of the 2025 ACM Conference on Conversational User Interfaces, Jul. 2025, doi: 10.1145/3719160.3737638.

[13] N. Kurian, "'No, Alexa, no!': designing child-safe AI and protecting children from the risks of the 'empathy gap' in large language models," Learn. Media Technol., vol. 50, no. 4, pp. 621–634, 2025, doi: 10.1080/17439884.2024.2367052.

[14] D. Carbone, F. Marcatto, F. Mistichelli, and D. Ferrante, "Perceiving Digital Threats and Artificial Intelligence: A Psychometric Approach to Cyber Risk," Journal of Cybersecurity and Privacy, vol. 5, no. 4, Dec. 2025, doi: 10.3390/JCP5040093.

[15] R. Skotnes, K. Hansen, and A. V. Krøvel, "Risk and Crisis Communication about Invisible Hazards," Journal of International Crisis and Risk Communication Research, vol. 4, no. 2, pp. 413–438, 2021, doi: 10.30658/JICRCR.4.2.9.

[16] A. Schreiber and I. Schreiber, "Bridging knowledge gap: the contribution of employees' awareness of AI cyber risks comprehensive program to reducing emerging AI digital threats," Information and Computer Security, vol. 32, no. 5, pp. 613–635, Nov. 2024, doi: 10.1108/ICS-10-2023-0199.

[17] D. Showkat, L. Wang, L. Chan, and A. To, "Equality Engine: Fostering Critical Machine Learning Bias Literacy Through a Transformational Game," Proc. ACM Hum. Comput. Interact., vol. 9, no. 2, May 2025, doi: 10.1145/3710966.

[18] C. A. Abel and T. D. Grace, "Designing dyadic caregiver-child personas for interactive digital media use," Proceedings of the Interaction Design and Children Conference, IDC 2020, pp. 198–204, Jun. 2020, doi: 10.1145/3392063.3394391.

[19] S. Moqbel, "Family-Based Persuasion in Digital Parenting Tools: A Research Proposal," Communications in Computer and Information Science, vol. 2542 CCIS, pp. 148–159, 2026, doi: 10.1007/978-3-031-97177-8_14.