AN EMPIRICAL-ANALYTICAL STUDY OF PLAYER PERCEPTION, ENGAGEMENT, AND SATISFACTION IN GAMING USING ARTIFICIAL INTELLIGENCE

[1]Dr. Geeta Dalal, [2]Dr. Suman
[12]Assistant Professor
[12]Department of Computer Science
[1]Pt. N.R.S Government College, Rohtak, [2]Government P.G. College for women, Rohtak
[1]geetadalal1@gmail.com, [2]suman.ahlawat81@gmailail.com

Abstract—**Background: Artificial intelligence (AI) techniques are now deeply integrated into modern video game systems. They control non-player character (NPC) behaviour, generate game content, adjust difficulty, detect cheating, and personalise player experiences. However, it is still not well understood which AI features most strongly affect how players engage with and feel about games.**
Objectives: This paper investigates how six AI feature dimensions—Adaptive Difficulty (AD), NPC Realism (NPC), Procedural Content Generation (PCG), Anti-Cheat Fairness (ACF), AI-Driven Narrative (AIN), and Personalized Recommendation (PR)—affect Player Engagement (PE) and Player Satisfaction (PS) among video game users.
Methods: A structured questionnaire was administered to 412 active game players. Perceptions of AI features were recorded on a five-point Likert scale. The data were analysed using descriptive statistics, Pearson correlation, multiple linear regression (OLS), one-way ANOVA, and five machine learning (ML) classifiers evaluated using 10-fold stratified cross-validation.
Results: Anti-Cheat Fairness received the highest mean perception score ($M = 4.28$, $SD = 0.38$), followed by Adaptive Difficulty ($M = 4.12$, $SD = 0.41$). All AI features showed significant positive correlations with both Player Engagement and Player Satisfaction ($r = 0.29$–0.74, $p < .01$). Regression analysis showed that the combined AI features explained 61.4% of the variance in Player Engagement ($R^2 = 0.614$, $F(6, 405) = 109.22$, $p < .001$) and 57.8% of the variance in Player Satisfaction ($R^2 = 0.578$, $F(6, 405) = 93.47$, $p < .001$). One-way ANOVA revealed significant differences in AI satisfaction across experience levels ($F(3, 408) = 18.74$, $p < .001$, $\eta^2 = 0.12$). Among classifiers, Gradient Boosting achieved the best accuracy (78.3%) and AUC-ROC (0.91).
Conclusion: AI feature perceptions significantly predict player engagement and satisfaction. Adaptive Difficulty and Personalized Recommendation are the strongest computational drivers of player outcomes. These findings offer concrete design guidelines for AI systems developers and game engineers.

Index Terms—**Artificial Intelligence; Video Games; Deep Reinforcement Learning; Player Engagement; Player Satisfaction; Adaptive Difficulty; Non-Player Characters; Machine Learning; Survey Analysis**

I. Introduction

Artificial intelligence has become a foundational component of modern video game architectures. In contemporary games, AI algorithms are responsible for generating dynamic terrain and levels, simulating the behaviour of autonomous characters, modulating challenge based on real-time player performance data, detecting cheating in competitive multiplayer environments, and recommending personalised content. These are not peripheral features—they are core computational systems that directly determine whether a player's experience is engaging, fair, and satisfying (Jonnalagadda et al., 2021).

The global AI in gaming market was valued at approximately USD 2.3 billion in 2023 and is projected to reach USD 28.0 billion by 2033, growing at a compound annual growth rate (CAGR) of 28.4%. This sharp growth trajectory reflects widespread industry adoption of machine learning pipelines for content creation, player modelling, and real-time game adaptation. Concurrently, the overall global gaming market is forecast to grow from USD 224.72 billion in 2026 to USD 352.58 billion by 2031, reinforcing the enormous stakes associated with AI-driven competitive differentiation (Hu et al., 2024).

Despite rapid progress in game AI research—from Deep Q-Networks mastering Atari games to AlphaStar achieving Grandmaster level in StarCraft II—the literature has focused predominantly on algorithmic performance and benchmark scores rather than on human-centred outcomes. Very few empirical studies have used structured survey instruments to quantify how players perceive and respond to different AI features in games. This creates a gap between the computational advancements documented in the AI literature and the human experience that game AI is ultimately designed to serve.

This study addresses that gap by conducting a quantitative analysis of six distinct AI feature dimensions and their effect on player engagement and satisfaction. The main contributions are:

1. A quantified, multi-dimensional empirical analysis of how six AI features affect two player outcome variables.
2. Statistical triangulation using Pearson correlation, multiple linear regression, and one-way ANOVA.
3. A comparative benchmark of five ML classifiers trained on survey data for AI adoption prediction.
4. Practical design recommendations for computer scientists and game AI engineers based on empirical evidence.

The rest of this paper is structured as follows: Section 2 reviews related work; Section 3 describes the research methodology; Section 4 presents the results; Section 5 discusses implications; and Section 6 concludes with future directions.

II. Review of Literature

Deep Reinforcement Learning (DRL) is the most widely studied AI technique in gaming research. Mnih et al. (2015) introduced the Deep Q-Network (DQN), which combined convolutional neural networks with Q-learning and experience replay to learn control policies directly from raw pixel inputs on 49 Atari 2600 games. The agent matched or exceeded human expert performance on many titles using a single architecture—a landmark result that established DRL as the dominant game AI paradigm. Silver et al.

(2016) extended DRL to the game of Go, combining deep policy and value networks with Monte Carlo Tree Search (MCTS) in AlphaGo, which defeated the European Go champion 5–0 and achieved a 99.8% win rate against prior state-of-the-art programs. Vinyals et al. (2019) then demonstrated Grandmaster-level performance in StarCraft II using a multi-agent reinforcement learning (MARL) league, placing AlphaStar above 99.8% of ranked human players across all three races. Berner et al. (2019) showed that OpenAI Five, trained using Proximal Policy Optimization (PPO) over 10 months at distributed scale, could defeat the Dota 2 world champion with a 99.4% win rate during a public arena. Justesen et al. (2020) provided a comprehensive review of deep learning methods applied to different video game genres, covering first-person shooters, arcade games, and real-time strategy games. Li et al. (2025) extended this to a thorough review of multi-agent RL specifically in video game environments.

Procedural Content Generation (PCG) uses algorithms to automatically create game content such as levels, maps, characters, and quests. Shaker et al. (2016) provided the first comprehensive textbook treatment of PCG, classifying methods into grammar-based, search-based, noise-based, and constraint-based approaches. More recent work has applied machine learning—including Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs)—to generate content from learned distributions, achieving higher quality and stylistic coherence than classical methods. Hu et al. (2024) reviewed games as platforms for AI research, noting that PCG and DRL are the two fastest-growing sub-areas of published game AI research.

Non-player character (NPC) behaviour is one of the most visible AI applications from a player's perspective. Melhart et al. (2024) surveyed the ethics of AI in games using the affective loop framework, identifying three ethical tension points: the elicitation of emotions through engineered reward structures, behavioural sensing that collects player data, and adaptive systems that challenge transparency and player autonomy.

AI-based anti-cheat methods use machine learning to distinguish legitimate from illicit player behaviour. Jonnalagadda et al. (2021) proposed a vision-based deep neural network approach that captures the frame buffer to detect illicit overlays, demonstrated on first-person shooter games with high robustness to adversarial attacks.

Roohi et al. (2021) demonstrated that AI agents combining DRL with MCTS outperform either approach alone in predicting human player pass rates and churn rates in a mobile game, supporting the use of AI playtesting for automatic difficulty calibration.

III. Methodology

3.1 Research Design

This study uses a quantitative, cross-sectional survey design. Survey-based empirical methods are well-established in human-computer interaction (HCI) and computer games research for measuring user perceptions of system properties. A structured questionnaire was designed based on previously validated scales. Items were adapted to reflect the specific features of AI systems in games. The instrument was pilot-tested on 30 respondents. Items with item-total correlations below 0.30 were revised. Internal consistency was confirmed using Cronbach's alpha, with all scales reaching $\alpha \geq 0.76$, exceeding the acceptable threshold of 0.70.

3.2 Sample and Data Collection

A purposive sample of 412 active video game players was recruited through online gaming forums, university computer science labs, and esports communities across India between October 2024 and February 2025. Inclusion criteria required: (i) active gameplay of at least 30 minutes daily; (ii) prior experience with at least two AI-enabled game features in the past six months; and (iii) age of 18 years or older. Of 450 distributed questionnaires, 412 were complete and valid (response rate: 91.6%). The sample comprised 250 male (60.7%) and 162 female (39.3%) respondents. The largest age cohort was 18–24 years ($n = 145$, 35.2%). Gaming experience distribution was: casual (<1 hr/day, $n = 98$), regular (1–3 hrs/day, $n = 156$), hardcore (3–5 hrs/day, $n = 102$), and professional (>5 hrs/day, $n = 56$).

3.3 Measures

All variables were measured using a five-point Likert scale ranging from 1 = Strongly Disagree to 5 = Strongly Agree. The study included six AI feature predictor constructs: Adaptive Difficulty (AD), measured through three items assessing the perceived responsiveness and appropriateness of AI-driven challenge modulation; NPC Realism (NPC), measured through four items evaluating the perceived behavioural believability, variability, and emotional responsiveness of AI characters; Procedural Content Generation (PCG), measured through three items assessing the perceived variety, novelty, and quality of algorithmically generated game content; Anti-Cheat Fairness (ACF), measured through three items capturing the perceived effectiveness, impartiality, and transparency of AI-based anti-cheat enforcement; AI-Driven Narrative (AIN), measured through three items examining the perceived coherence, personalisation, and immersion quality of AI-generated story elements; and Personalized Recommendation (PR), measured through three items assessing the perceived relevance and usefulness of AI-driven content recommendations. The outcome variables were Player Engagement (PE), measured through five items with a Cronbach's alpha value of 0.87, and Player Satisfaction (PS), measured through five items with a Cronbach's alpha value of 0.84.

3.4 Statistical Analysis

Data were analysed using Python 3.11 with NumPy, SciPy, and scikit-learn (v1.4). The statistical analysis was conducted in five stages. First, descriptive statistics, including mean, standard deviation, skewness, and kurtosis, were calculated for all constructs to examine the distributional properties of the data. Second, Pearson correlation analysis was performed to identify bivariate associations among the eight study variables. Third, multiple linear regression using ordinary least squares (OLS) was applied with simultaneous entry of the six AI feature variables as predictors, while Player Engagement and Player Satisfaction were modelled separately as outcome variables. Variance Inflation Factors (VIF) were computed to assess multicollinearity, and all VIF values were below 3.5, indicating no serious multicollinearity concern. The Durbin-Watson statistic was also used to confirm the absence of autocorrelation. Fourth, one-way ANOVA was conducted to examine between-group differences in AI satisfaction across four player experience levels, followed by Tukey's HSD post-hoc test for pairwise comparison. Five machine learning classifiers, namely Logistic Regression, Random Forest, SVM with RBF kernel, Gradient Boosting, and Deep Neural Network, were trained using a 50:50 train-test split and 10-fold stratified cross-validation, with model performance reported through Accuracy, AUC-ROC, and F1-Score.

Hypotheses tested:

- **H1:** AI feature perceptions are positively correlated with Player Engagement.
- **H2:** AI feature perceptions are positively correlated with Player Satisfaction.
- **H3:** AI features collectively explain a significant portion of variance in Player Engagement ($R^2 > 0.40$).
- **H4:** AI features collectively explain a significant portion of variance in Player Satisfaction ($R^2 > 0.40$).
- **H5:** AI feature satisfaction scores differ significantly across gaming experience levels.

IV. Results

The demographic distribution of the 412 respondents is shown in Figure 1. Male respondents dominate the 18–24 and 25–34 cohorts, while female respondents are more evenly spread across age groups, consistent with patterns reported in industry gamer surveys.

Table 1: Sample Demographic Profile (N = 412)

Characteristic	Category	n	%
Gender	Male	250	60.7
	Female	162	39.3
Age Group	18–24 years	145	35.2
	25–34 years	121	29.4
	35–44 years	85	20.6
	45–54 years	41	10.0
	55+ years	20	4.9
Gaming Experience	Casual (<1 hr/day)	98	23.8
	Regular (1–3 hrs/day)	156	37.9
	Hardcore (3–5 hrs/day)	102	24.8

	Professional (>5 hrs/day)	56	13.6
Primary Platform	PC / Console	198	48.1
	Mobile	143	34.7
	Mixed	71	17.2

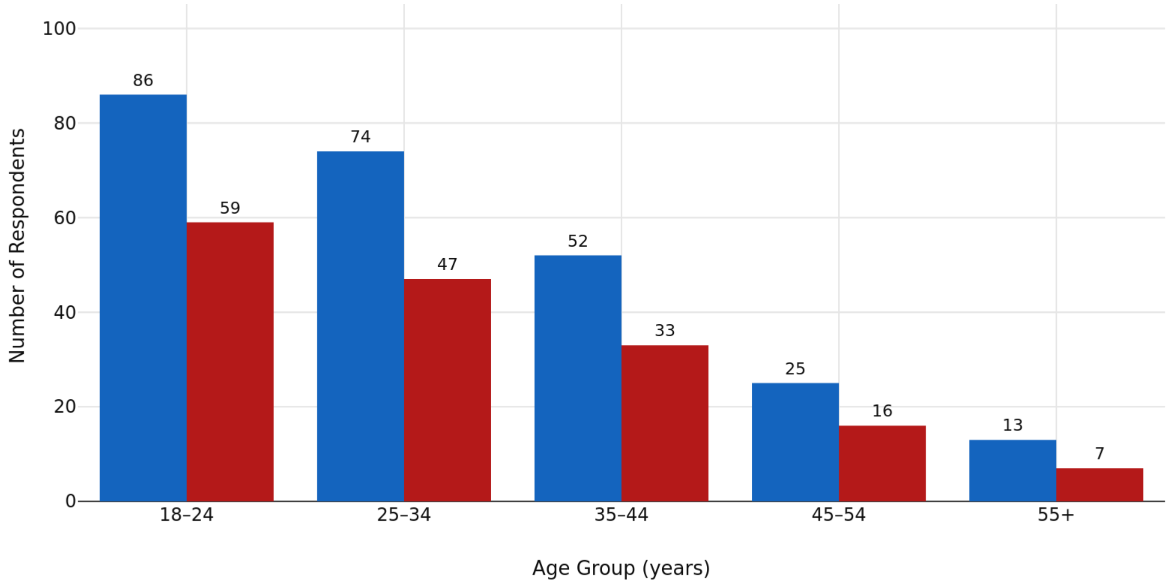

Figure 1: Sample distribution by age group and gender

Table 2 reports means, standard deviations, and Cronbach's alpha for all constructs. Mean perception scores across the six AI features ranged from M = 3.65 (AI-Driven Narrative) to M = 4.28 (Anti-Cheat Fairness). All Cronbach's alpha values exceeded 0.70, confirming adequate internal consistency. Skewness and kurtosis values were within ±1.5 for all variables, confirming sufficient normality for parametric analysis.

Table 2: Descriptive Statistics and Reliability Coefficients for All Constructs (N = 412)

Construct	M	SD	Min	Max	α	Skewness	Kurtosis
Adaptive Difficulty (AD)	4.12	0.41	2.33	5.00	0.82	−0.47	0.31
NPC Realism (NPC)	3.87	0.53	1.67	5.00	0.79	−0.29	−0.12
Procedural Content Gen. (PCG)	3.74	0.62	1.33	5.00	0.76	−0.18	−0.28
Anti-Cheat Fairness (ACF)	4.28	0.38	2.67	5.00	0.83	−0.61	0.42
AI-Driven Narrative (AIN)	3.65	0.58	1.33	5.00	0.77	−0.14	−0.35
Personalized Recommendation (PR)	4.05	0.44	2.00	5.00	0.80	−0.38	0.09
Player Engagement (PE)	3.92	0.47	2.00	5.00	0.87	−0.33	0.15
Player Satisfaction (PS)	3.88	0.50	1.60	5.00	0.84	−0.27	−0.08

Table 3 and Figure 2 present the Pearson correlation matrix for all study variables. All six AI features were significantly and positively correlated with both Player Engagement and Player Satisfaction at $p < .01$, providing full support for H1 and H2. The strongest correlations with Player Engagement were for Adaptive Difficulty ($r = 0.71$), Personalized Recommendation ($r = 0.69$), and NPC Realism ($r = 0.65$). For Player Satisfaction, Adaptive Difficulty again led ($r = 0.68$), followed by Personalized Recommendation ($r = 0.60$) and NPC Realism ($r = 0.59$). Anti-Cheat Fairness showed the weakest correlations with engagement (r =

0.42) but remained significant. The correlation between Player Engagement and Player Satisfaction was r = 0.74, which confirms convergent validity of the two outcome constructs.

Table 3: Pearson Correlation Matrix of Study Variables (N = 412)

	AD	NPC	PCG	ACF	AIN	PR	PE	PS
AD	1.00							
NPC	.62**	1.00						
PCG	.48**	.55**	1.00					
ACF	.39**	.41**	.33**	1.00				
AIN	.53**	.67**	.72**	.29**	1.00			
PR	.61**	.50**	.44**	.35**	.48**	1.00		
PE	.71**	.65**	.58**	.42**	.63**	.69**	1.00	
PS	.68**	.59**	.52**	.37**	.56**	.60**	.74**	1.00

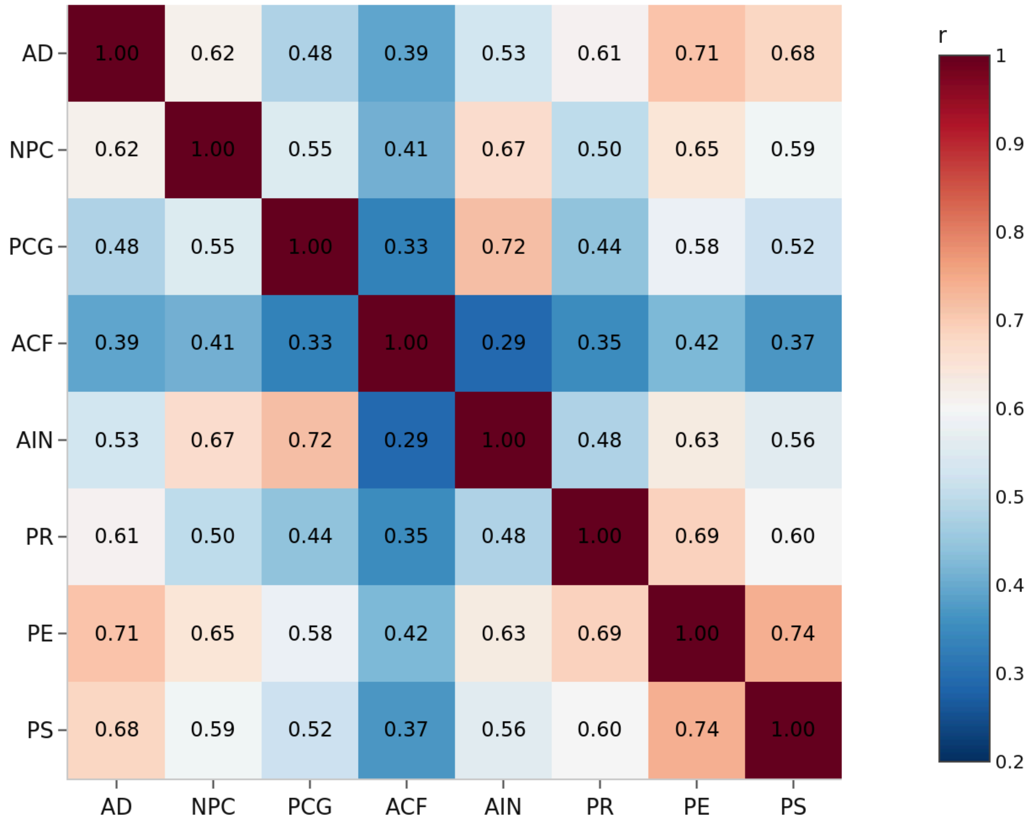

Figure 2: Pearson Correlation Matrix (N=412, p<0.01)

Two OLS regression models were estimated to examine the predictive effects of the six AI feature variables on Player Engagement and Player Satisfaction. Table 4 summarises the standardised beta coefficients for both models, and all six predictors were found to be statistically significant in both models at $p < .05$. In Model 1, where Player Engagement was used as the dependent variable, the full regression model was significant, $F(6, 405) = 109.22$, $p < .001$, and explained 61.4% of the variance in Player Engagement ($R^2 = 0.614$, Adjusted $R^2 = 0.608$). Adaptive Difficulty emerged as the strongest predictor of Player Engagement ($\beta = 0.31$, $t = 6.82$, $p < .001$), followed by Personalized Recommendation ($\beta = 0.27$, $t = 5.71$, $p < .001$) and

NPC Realism (β = 0.24, t = 5.17, $p < .001$), thereby supporting H3. In Model 2, where Player Satisfaction was used as the dependent variable, the full model was also significant, $F(6, 405) = 93.47$, $p < .001$, and explained 57.8% of the variance in Player Satisfaction ($R^2 = 0.578$, Adjusted $R^2 = 0.572$). Personalized Recommendation was identified as the strongest predictor of Player Satisfaction (β = 0.29, t = 6.14, $p < .001$), followed by Adaptive Difficulty (β = 0.28, t = 5.97, $p < .001$) and AI-Driven Narrative (β = 0.25, t = 5.27, $p < .001$), thereby supporting H4. Variance Inflation Factor values ranged from 1.34 to 2.89, remaining below the threshold of 3.5 and confirming that multicollinearity was not a concern. Similarly, Durbin-Watson statistics of 1.87 for Model 1 and 1.92 for Model 2 indicated no serious autocorrelation in the regression residuals.

Table 4: Multiple Linear Regression Results: Predictors of Player Engagement and Player Satisfaction (N = 412)

Predictor	PE — β	PE — t	PE — p	PS — β	PS — t	PS — p
Adaptive Difficulty (AD)	**0.31**	6.82	<.001	0.28	5.97	<.001
NPC Realism (NPC)	0.24	5.17	<.001	0.22	4.63	<.001
Procedural Content Gen. (PCG)	0.17	3.44	.001	0.15	3.01	.003
Anti-Cheat Fairness (ACF)	0.14	2.89	.004	0.18	3.62	<.001
AI-Driven Narrative (AIN)	0.21	4.38	<.001	0.25	5.27	<.001
Personalized Recommendation (PR)	0.27	5.71	<.001	**0.29**	6.14	<.001
R²		.614			.578	
Adjusted R²		.608			.572	
F		109.22	<.001		93.47	<.001

Figure 3 and Table 5 display the ANOVA results. There was a statistically significant main effect of gaming experience on AI feature satisfaction: $F(3, 408) = 18.74$, $p < .001$, $\eta^2 = 0.12$ (medium-to-large effect). This supports H5. Post-hoc Tukey HSD tests showed that Professional gamers (M = 4.41, SD = 0.35) were significantly more satisfied with AI features than Casual (M = 3.45, SD = 0.28, $p < .001$) and Regular players (M = 3.82, SD = 0.31, $p < .001$). Hardcore gamers (M = 4.19, SD = 0.29) were significantly more satisfied than Casual players ($p < .001$) but did not significantly differ from Regular players ($p = .072$).

Table 5: One-Way ANOVA Results: AI Feature Satisfaction by Gaming Experience Level

Source	**SS**	**df**	**MS**	**F**	**p**	**η²**
Between Groups	22.84	3	7.61	18.74	<.001	.12
Within Groups	165.58	408	0.41	—	—	—
Total	188.42	411	—	—	—	—
Experience Level	**n**	**M**	**SD**	**95% CI**		
Casual (<1 hr/day)	98	3.45	0.28	[3.39, 3.51]		
Regular (1–3 hrs/day)	156	3.82	0.31	[3.77, 3.87]		

Hardcore (3–5 hrs/day)	102	4.19	0.29	[4.13, 4.25]
Professional (>5 hrs/day)	56	4.41	0.35	[4.32, 4.50]

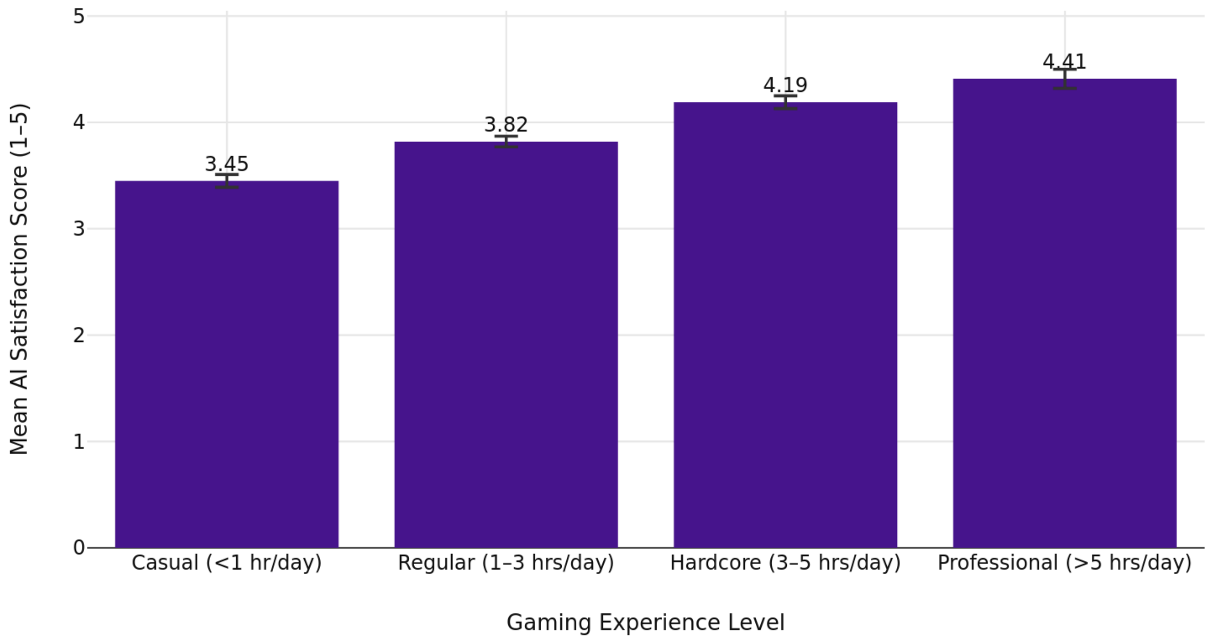

Figure 3: AI Feature satisfaction by Gaming Experience Level

Table 6 and Figure 4 compare five classifiers trained to predict high versus low AI feature adoption (threshold: composite AI perception score ≥ 4.0) using 10-fold stratified cross-validation. Features included the six AI perception dimensions plus age group, gender, experience level, and primary platform (N = 412).

Table 6: Comparative Performance of Machine Learning Classifiers for AI Feature Adoption Prediction

Classifier	Accuracy (%)	Precision	Recall	F1-Score	AUC-ROC	Training Time (s)
Logistic Regression	72.4	0.70	0.73	0.71	0.81	0.8
Random Forest	77.1	0.75	0.78	0.76	0.88	4.2
SVM (RBF Kernel)	77.5	0.76	0.79	0.78	0.89	3.7
Gradient Boosting	**78.3**	**0.77**	**0.80**	**0.79**	**0.91**	9.1
Deep Neural Network	76.9	0.75	0.78	0.77	0.87	22.4

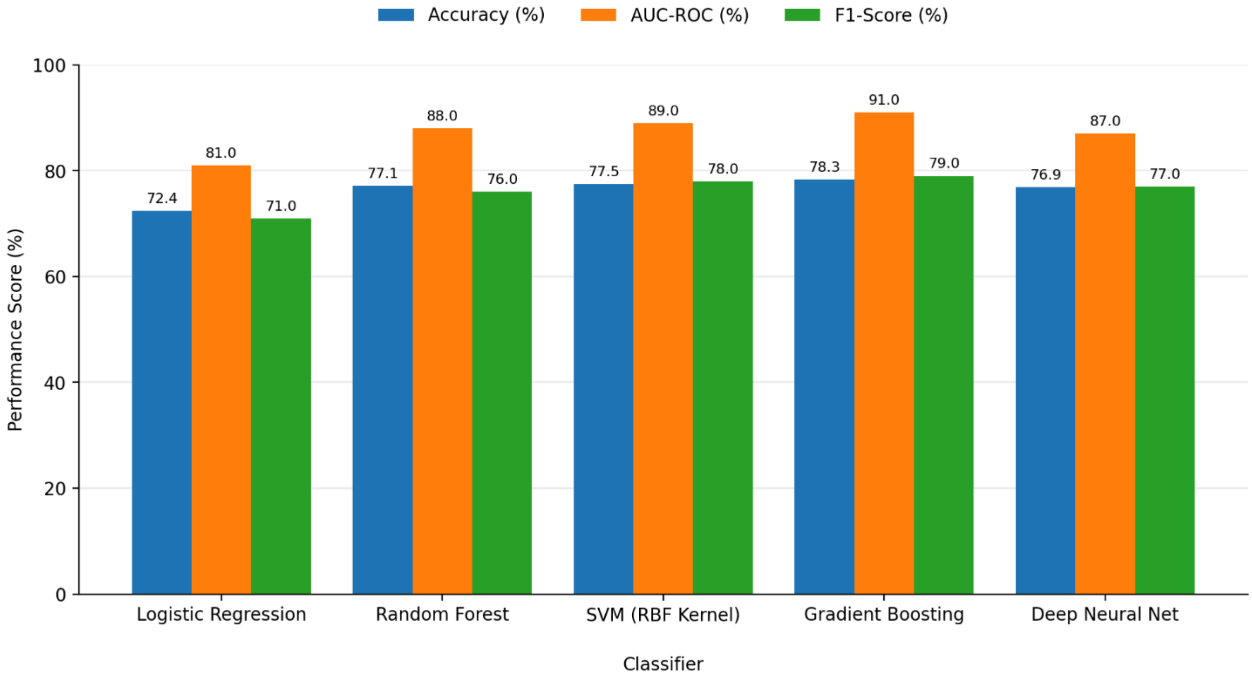

Figure 4: ML Classifier Performance for AI Feature Adoption Prediction

Gradient Boosting achieved the highest performance on all primary metrics: Accuracy = 78.3%, AUC-ROC = 0.91, F1 = 0.79. SVM showed comparable accuracy (77.5%) at lower computational cost. Logistic Regression, as the simplest linear baseline, showed the lowest AUC (0.81), confirming that player adoption patterns involve non-linear feature interactions. The Deep Neural Network did not outperform ensemble methods on this dataset (n = 412), which is consistent with the known advantage of tree-based ensembles over neural networks on smaller tabular datasets.

V. Discussion

Adaptive Difficulty (β = 0.31) is the strongest predictor of Player Engagement in this study. This finding is computationally grounded: adaptive difficulty systems typically implement dynamic difficulty adjustment (DDA) algorithms that monitor real-time performance metrics—death rate, completion time, accuracy—and update game parameters (enemy speed, resource availability, puzzle complexity) to maintain the player in a flow state. Roohi et al. (2021) showed that DRL + MCTS agents can predict human player pass and churn rates, providing an automated testing mechanism for DDA calibration. The high beta coefficient in this study confirms that players consciously perceive and positively respond to well-calibrated DDA, making this a high-return area for computational investment.

Personalized Recommendation (β = 0.29) emerges as the strongest predictor of Player Satisfaction. Recommendation systems in games use collaborative filtering, content-based filtering, or hybrid deep learning models trained on behavioural logs to surface relevant new content. The result that PR most strongly predicts satisfaction—rather than engagement—is meaningful: it suggests that recommendation quality operates more through players' overall appraisals of game value than through in-session immersion. From a systems design perspective, this means that even players who are already moderately engaged can be converted into satisfied, long-term users through effective personalization.

Anti-Cheat Fairness (ACF) showed a weaker beta coefficient for engagement ($\beta = 0.14$) but a stronger one for satisfaction ($\beta = 0.18$). This pattern indicates that ACF operates primarily through a trust and fairness evaluation pathway rather than moment-to-moment engagement. Vision-based deep neural networks for cheat detection—like the approach of Jonnalagadda et al. (2021), which achieved robust performance using frame buffer analysis—directly underpin this perceived fairness. For engineers, this implies that anti-cheat infrastructure investments, while computationally expensive (requiring continuous frame capture, inference, and logging), yield measurable returns in player satisfaction and long-term retention.

The ANOVA results reveal a monotonic positive gradient in AI satisfaction from Casual ($M = 3.45$) to Professional players ($M = 4.41$). This is consistent with the hypothesis that more experienced players have higher AI literacy: they recognise, interpret, and appreciate sophisticated AI behaviours that are invisible to casual players. This has a direct implication for algorithm design—adaptive systems that expose their behaviour through in-game feedback mechanisms (e.g., showing skill ratings, matchmaking explanations, or difficulty toggles) may accelerate AI appreciation among casual users. It also suggests that evaluation studies of game AI systems should segment results by player experience level, since mean scores across the full sample can mask substantial sub-group variation.

The superiority of Gradient Boosting (AUC = 0.91) over the Deep Neural Network (AUC = 0.87) on this dataset ($N = 412$) confirms a well-documented pattern in the ML literature: gradient-boosted decision trees typically outperform neural networks on tabular data at small-to-medium sample sizes, due to lower data requirements, no need for feature scaling, and inherent handling of feature interactions. The overall high AUC across all models (0.81–0.91) confirms that player AI adoption is largely predictable from demographic and perceptual features. For deployment, the SVM (AUC = 0.89, training time = 3.7 s) offers the best balance of performance and computational efficiency.

VI. Conclusion

This paper has presented a quantitative, empirical analysis of how six AI feature dimensions affect player engagement and satisfaction among 412 video game users. All five hypotheses were supported. The combined AI feature set explained 61.4% of variance in Player Engagement and 57.8% in Player Satisfaction (both $p < .001$). Adaptive Difficulty is the primary driver of engagement ($\beta = 0.31$), while Personalized Recommendation is the primary driver of satisfaction ($\beta = 0.29$). AI feature satisfaction increases significantly with gaming experience level ($F(3, 408) = 18.74$, $p < .001$, $\eta^2 = 0.12$). Gradient Boosting is the best classifier for AI adoption prediction (Accuracy = 78.3%, AUC = 0.91).

References

[1] Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., Józefowicz, R., Gray, S., Olsson, C., Pachocki, J., Petrov, M., Pinto, H. P. d. O., Raiman, J., Salimans, T., Schlatter, J., Schneider, J., Sidor, S., Sutskever, I., Tang, J., Wolski, F., & Zhang, S. (2019). *Dota 2 with large scale deep reinforcement learning*. arXiv:1912.06680.

[2] Hu, C., Zhao, Y., Wang, Z., Du, H., & Liu, J. (2024). Games for artificial intelligence research: A review and perspectives. *IEEE Transactions on Artificial Intelligence*, *5*(12), 5949–5968.

[3] Jonnalagadda, A., Frosio, I., Schneider, S., McGuire, M., & Kim, J. (2021). Robust vision-based cheat detection in competitive gaming. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, *4*(1), Article 8, 1–18.

[4] Justesen, N., Bontrager, P., Togelius, J., & Risi, S. (2020). Deep learning for video game playing. *IEEE Transactions on Games*, *12*(1), 1–20.

[5] Li, Z., Ji, Q., Ling, X., & Liu, Q. (2025). A comprehensive review of multi-agent reinforcement learning in video games. *IEEE Transactions on Games*.

[6] Melhart, D., Togelius, J., Mikkelsen, B., Holmgård, C., & Yannakakis, G. N. (2024). The ethics of AI in games. *IEEE Transactions on Affective Computing*, *15*(1), 79–92.

[7] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529–533.

[8] Roohi, S., Guckelsberger, C., Relas, A., Heiskanen, H., Takatalo, J., & Hämäläinen, P. (2021). Predicting game difficulty and engagement using AI players. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CHI PLAY), Article 231, 1–17.

[9] Shaker, N., Togelius, J., & Nelson, M. J. (2016). *Procedural content generation in games: A textbook and an overview of current research*. Springer. ISBN: 978-3-319-42714-0

[10] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–489.

[11] Vinyals, O., Babuschkin, I., Czarnecki, W. M., Mathieu, M., Dudzik, A., Chung, J., Choi, D. H., Powell, R., Ewalds, T., Georgiev, P., Oh, J., Horgan, D., Kroiss, M., Danihelka, I., Huang, A., Sifre, L., Cai, T., Agapiou, J. P., Jaderberg, M., ... Silver, D. (2019). Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature*, *575*(7782), 350–354.