IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 2 December 2025, accepted 21 January 2026, date of publication 23 January 2026, date of current version 29 January 2026.

*Digital Object Identifier 10.1109/ACCESS.2026.3657676*

RESEARCH ARTICLE

# Anchored Log-Based User Personas: A Typology-Guided Framework for Clustering Gameplay Behavior

**JÁN SKALKA, MARTIN DRLÍK, DOMINIK HALVONÍK, PETER KUNA, AND MATÚŠ VALKO**
Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, 949 01 Nitra, Slovakia

Corresponding author: Ján Skalka (jskalka@ukf.sk)

This work was supported by the project 'Nitra Cybersecurity Competence Center' under Project No. 17R05-04-V01-00003, funded by the Recovery and Resilience Plan of the Slovak Republic, under call 17R05-04-V01 'Establishment of Cybersecurity Competence Centers.'

**ABSTRACT** Modeling players in digital games increasingly relies on rich game telemetry, but many studies still provide one-off, game-specific procedures that are difficult to reuse. This paper presents a typologically anchored, game event log-based framework for inferring interpretable and stable player personas from in-game behavior without relying on questionnaires. We use telemetry from a first-person shooter game designed explicitly as a research testbed, and reconstruct player sessions into a set of player-level metrics that capture pacing, exploration, combat performance, risk exposure, and resource management. These metrics are organized into blocks inspired by Bartle's typology (Achiever, Explorer, Killer) to maintain a clear conceptual link between low-level events and higher-level motivational tendencies. After correlation screening and event log transformation, we apply principal component analysis to obtain an orthogonal latent space and perform k-means clustering. The stability of the clusters is assessed using a core-clustering procedure based on pruning and bootstrap resampling. The resulting personas exhibit distinct and interpretable behavioral profiles, including aggressive, high-risk players, cautious survivors, and exploration-oriented tacticians. We discuss how this methodology can be transferred to other games and how the identified personas can influence adaptive design and training scenarios for serious games.

**INDEX TERMS** Player modelling, gameplay telemetry, behavior clustering, principal component analysis, Bartle typology, games.

## I. INTRODUCTION

In modern digital games, understanding the player base is crucial for game developers and has become a central topic in game analytics [1], [2]. Player modeling, the creation of an abstract description of how a player behaves in a game environment, is essential for optimizing game design, ensuring a good user experience, and enabling the personalization or adaptation of the game itself [3], [4]. Such models are particularly important for adaptive games, where content or difficulty is generated or adjusted to match the preferences, skills, or play styles of individual players [5], [6].

The associate editor coordinating the review of this manuscript and approving it for publication was Li He.

Traditional approaches to player segmentation and typology often rely on top-down methods such as questionnaires or designer-driven taxonomies. Models such as Bartle's Player Types [7] or HEXAD [8], [9] are typically derived from self-report measures collected outside of a game session and can suffer from bias, limited ecological validity, and limited applicability across genres [10], [11]. In contrast, game event logs provide the most common quantitative source of information about player behavior, collecting thousands of actions per session. These logs have the potential to support objective analyses of how players interact with game systems [12]. Although unsupervised, bottom-up clustering of gameplay telemetry has been explored in several studies, log-based behavioral profiling remains a relatively young and fragmented field, with limited method standardization and open challenges in producing interpretable, stable clusters. Recent reviews of game analytics point out that, despite widespread industrial adoption, research still lacks systematic examination and classification of analytic approaches, including data-driven player segmentation [13], [14], [15].

Log-based player models face several persistent problems that limit their practical value and reproducibility.

- Advanced machine learning approaches can capture complex patterns in game event logs, but they are often perceived as closed boxes: their internal representations are complex to connect to intuitive concepts such as tactics, strategy, or motivation. Explicit features derived directly from logs are understandable, but latent factors produced by complex models (such as embeddings or other high-dimensional codes) are challenging to translate into practical design insights, hindering their adoption by game designers. The trade-off between predictive performance and interpretability is explicitly identified as a barrier to the adoption of closed-box models in applied areas, including user and behavior modeling [16], [17], [18].
- Different clustering algorithms, parameter settings, or even random initializations can lead to significantly different segmentations of the same population of players [2], [19]. Methods such as k-means summarize average behavior, while archetypal analysis focuses on extremes at the edges of the data space, leading to different interpretations of "typical" and "extreme" play. Most studies report only a static snapshot and rely on internal indices, providing little information about how robust the clusters discovered are to resampling or over time [2], [14].
- There is no generally accepted, clearly documented analytical procedure that others can easily replicate. Although the general process is well known, applying dimensionality reduction and clustering to game logs still requires considerable domain expertise and many ad-hoc decisions [15]. Game event logs are heterogeneous and poorly standardized, so preprocessing and feature engineering often dominate the work, and published studies rarely describe these steps in sufficient detail to be reusable. As a result, it is challenging to derive generalizable, detailed procedures for modeling players from event logs, particularly when more advanced analytical methods are used [2], [11].

Consequently, there is a gap for a reusable pipeline that derives interpretable, typology-anchored behavioral metrics directly from event logs and explicitly validates the stability of the resulting personas. This work addresses the need by presenting and demonstrating a stability-oriented approach to modeling players from gameplay logs of a single-player FPS scenario. We construct a fully specified approach that:

- Infers behavioral features from game event logs that are inspired by existing player type frameworks (notably Bartle's *Achiever* / *Killer* / *Explorer* distinction).
- Reduces redundancy and reveals latent structure using correlation filtering and principal component analysis (PCA).
- Applies stability-oriented clustering in latent space, combining k-means with bootstrap resampling and silhouette-based pruning to separate stable core personas from ambiguous edge cases.
- Produces interpretable personas that can be described in specific game metrics and are suitable for informing game design, analytics, and potential personalization.

Building on log-derived behavioral features from a single-player FPS game, we address the following research questions:

- RQ1: What latent behavioral dimensions emerge from the derived features (via PCA), and to what extent are these dimensions interpretable in terms of meaningful aspects of play?
- RQ2: When players are clustered based on their behavioral characteristics, what distinct play styles (personas) can be identified, and how can these personas be described in terms of concrete in-game metrics?
- RQ3: How stable is the identified persona structure, and how does it differ between the full sample and the high-silhouette core of players?
- RQ4: How do the extracted behavioral dimensions and personas relate to existing player-type frameworks, in particular Bartle's Achiever/Killer/Explorer categories?

Instead of claiming a universal set of player types, we aim to show that a transparent, log-driven pipeline can yield psychologically interpretable, empirically stable personas that bridge data-driven game analytics and theory-driven player modeling.

This paper is structured as follows. Section II discusses related work in player modeling, protocol-based analytics, and behavioral segmentation stability. Section III describes the proposed research methodology and feature construction. Section IV presents the data analysis steps, including correlation screening, dimensionality reduction, and clustering. Section V discusses the results with respect to the research questions, and Section VI concludes the paper and outlines limitations and future work.

## II. RELATED WORK

### A. PLAYER MODELLING AND FRAMEWORKS

Player modeling has a long tradition in game research and human-computer interaction, especially in the form of player typologies based on motivations or preferences [20], [21], [22]. Classic top-down approaches, such as Bartle's Player Types for MUD and MMO environments [7] or later models like Yee's motivations [23] or HEXAD [8], define a few ideal types (e.g., Achievers, Killers, Explorers, Socializers) and map players to these types using questionnaires or qualitative methods.

These frameworks provide a common language for designers and researchers, are used in the design of content, monetization strategies, and gamification systems, and often serve as ''mental models'' for thinking about player segments [11], [24]. However, more recent systematic reviews point to several limitations of such typologies [11], [25], [26]. Empirical studies point to dimensional overlap between different models, poor replicability of factor structures, problems with scale reliability, and unclear transferability across genres and platforms [9], [27], [28]. It often turns out that different typologies capture similar latent factors (e.g., performance, social interaction, exploration) but use different names and categories, which makes it difficult to compare results [11], [29], [30]. Moreover, these are mostly self-report instruments that measure stated preferences rather than actual behavior in specific game situations.

These measurements create a gap between how players describe themselves in questionnaires and how they actually behave, as reflected in the game event logs [8], [31], [32]. For this reason, it makes sense to combine typologies with game event log-based modeling: existing frameworks provide psychologically anchored interpretations, but not specific algorithmic procedures. Game event logs, on the other hand, provide rich behavioral data, but do not in themselves provide clear ''names'' for segments [33]. Current research thus opens the question of how to design features derived from logs so that they are both interpretable in terms of known typologies and suitable for further processing [21], [31], [34], [35].

### B. LOG-BASED GAME ANALYTICS

Event logs and telemetry have become a standard tool in commercial game development: many studies track retention, progression, monetization, and detailed game actions in real time [36], [37], [38]. The field of game analytics has grown out of this infrastructure, covering basic descriptive statistics, A/B testing, and advanced models [2], [39]. Since around 2010, many papers have appeared that use clustering to infer ''game styles'' directly from logs – from MMO games (e.g., World of Warcraft [31]) and free-to-play games [40] to mobile casual titles [1], [41]. Methods used include k-means clustering, hierarchical and model-based clustering, self-organizing maps, non-negative matrix factorization, and archetypal analysis [10], [38], [42].

Research based on gameplay logs shows that player behavior often forms transparent and interpretable patterns. For example, players who prefer PvP (player-versus-player) content tend to focus on combat, ranking, and competition, while PvE (player-versus-environment) players prefer fighting AI-controlled enemies and completing narrative or cooperative missions. Another common contrast is between speed runners, who try to complete the game as quickly as possible by optimizing every action, and completionists, whose goal is to explore all the content, complete every side quest, and collect all the items or achievements [1], [43], [44]. In games with microtransactions, logs can also classify ''whales'', a small group of players who spend disproportionately large amounts of money, from regular spenders who spend infrequently or only in small amounts [45].

However, they also show that the results are very sensitive to feature choice and game context. Many works construct ad hoc features (kill counts, zones visited, login counts) and rarely systematically distinguish between session-level, player-level, and time-dependent characteristics [1], [46]. There is also often a lack of a clear distinction between features that capture abilities, motivations, and structural conditions (e.g., the length of play and the available content). This lack complicates the interpretation of clusters in psychological terms [35], [47].

Another limitation is that existing studies provide a one-time snapshot of a particular game: the pipeline from logs to clusters is tailored to one dataset and is rarely described in a way that can be applied elsewhere [48]. Although there are significant methodological contributions (e.g., comparisons of multiple clustering methods on the same logs), comprehensive, step-by-step guides that define reusable procedures for feature selection, dimensionality reduction, and cluster validation across different games are still lacking [49], [50].

Two main methodological issues in log-based player modeling are interpretability and stability [36], [51]. Complex models (autoencoders, deep networks, complex mixture models) can capture subtle patterns in behavior, but their latent representations are difficult to translate into terms that a designer typically works with (pace of combat, riskiness, exploration rate) [52], [53]. Several papers therefore recommend using more easily interpretable techniques such as PCA or rotated factors, which provide orthogonal dimensions with readable loadings, even at the cost of a slight loss of predictive power [1], [54].

Stability concerns whether clusters and personas appear consistently across different methods, parameters, and data-sampling choices. Methodological studies show that k-means, fuzzy c-means, hierarchical clustering, and archetypal analysis can identify different patterns over the same features [15], [36], [54]. Different initializations or k values often lead to different persona profiles [20]. Stability is usually assessed using internal indices, which provide a number but say little about which players are reliably classified and which are ''borderline''. Few works use bootstrap, cross-dataset validation, or core-vs-periphery approaches to address these questions directly [55], [56].

Closely related to interpretability and stability is again the question of the pipeline. Even though the general processing is known, in practice, there is a lot of room for implicit decisions: which events to include, how to aggregate sessions per player, how to transform distributions, which correlations to consider acceptable, and which to test. Many articles describe these steps only briefly or not at all, so they cannot be exactly replicated. Some authors, therefore, explicitly present ''reusable methodologies'' for player clustering and call for more precise documentation of the entire workflow – from logs to validated personas – as a key prerequisite for cumulative research [2], [10], [11], [15], [57].

Representative studies of log-based segmentation, comparing methods and general clustering based on game telemetry [31], [36], as well as Bayesian-style clustering and time-series pattern discovery in large-scale multiplayer logs [40], [41], [56], typically treat segmentation as a one-off result and report limited evidence of robustness under resampling or parameter variation, suggesting that stability validation is largely neglected in practice.

### C. TOWARDS ADAPTIVE AND SERIOUS GAMES

Player models play a role not only in entertainment titles but also in adaptive and serious games, where they serve as the basis for dynamically adapting tasks, difficulty, and feedback. In educational and training scenarios, they help identify learning styles, skill levels, and typical errors, enabling the system to select appropriate situations and interventions [58], [59]. Reviews of adaptive serious games show that most existing systems use relatively simple models (e.g., score-based or time-on-task rules), and only a minority use rich log data for more sophisticated user segmentation [3], [6], [60], [61].

In many fields (e.g., cybersecurity), serious games and simulations are increasingly used to train professionals and end users. Here, differences in behavior – such as risk-taking, propensity to experiment, and speed of incident response – are directly linked to security consequences. However, many existing assessments are primarily based on outcome metrics (success, completion time) rather than on detailed modeling of problem-solving styles [62], [63]. Log-based personas could help identify, for example, risky strategies, overly cautious behavior, or inattention in routine tasks [64].

These connections suggest that methodological advances in log-based player modeling – interpretable dimensions, stable clusters, replicable pipelines – have relevance beyond commercial games. The same principles of analyzing game logs can be applied to better understand behavior in security training, support competency development, and ultimately design adaptive systems that respond to different "persona styles" of users when working on security-sensitive tasks [2], [37], [65].

## III. RESEARCH METHODOLOGY

This section describes the data, preprocessing steps, and analytical procedure used to derive interpretable player personas from game recordings. We work with telemetry from *Another Useless Zombie Game*, a first-person shooter game designed explicitly as a research testbed. Each playthrough generates a time-stamped event stream that records movement, shooting, resource acquisition, zombie interactions, and charging the exit portal (as a goal) with acceptable temporal resolution.

A key feature of the presented approach is the connection of the entire process to an explicit typology of players. In this study, we use Bartle's four-part taxonomy (Achiever, Explorer, Socializer, Killer) as a design vocabulary, but not as a measured construct. We do not estimate Bartle scores for individual players from questionnaires, but we use taxonomy to guide:

- construction of feature blocks,
- interpretation of principal components,
- labeling of the resulting personas.

However, the same methodology could be implemented with other typologies (e.g., Hexad, Yee's motivations) by redefining these mappings.

### A. RESEARCH FRAMEWORK

The analysis follows a framework that links raw gameplay records to interpretable and stable player personas.

- **Game data preprocessing** – at the beginning, we collect and preprocess event records from raw format into individual game actions and sessions, and standardize the representation of basic actions (movement, shooting, resource gathering, portal interaction, and zombie events). This step creates a session-level dataset that captures the whole temporal structure of each play session in a form suitable for feature extraction.
- **Behavioral metrics derivation –** we derive behavioral metrics that translate low-level events into interpretable characteristics of play. These features are grouped into conceptual blocks (e.g., engagement, exploration, performance) and annotated with their primary correspondence to Bartle's tendencies: Achiever, Killer, and Explorer. The characteristics of a Socializer in a single-player game are scattered among other traits, and it is almost impossible to isolate them responsibly. The goal of this step is to create a transparent bridge between specific game mechanics and higher-level constructs used in player type research.
- **Metrics aggregation (computation)** – covers the process of computing defined metrics at the player level and examining their underlying distributional properties. We compute descriptive statistics and examine skewness and redundancy to identify variables that require transformation or removal. This step yields a cleaned and normalized feature set that represents each player as a vector of stable behavioral characteristics (not a vector of raw events).
- **Reducing the dimensionality** of the feature space by revealing latent structure using Pearson correlations to remove highly redundant variables, then applying principal component analysis (PCA) to obtain orthogonal components that summarize broader patterns such as combat intensity, exploratory movement, and risk-taking. The retained components provide a compact (mean-based), numerically robust, and still interpretable in terms of game behavior.
- **Candidate player persona identification by clustering in the PCA space – we identify potential clustering options, verify multiple clustering**

**techniques, and select** the most suitable variant. This step aims to reveal a plausible segmentation of the player population, while the cluster boundaries remain preliminary and subject to further validation and refinement.
- **Validation of persona stability** using a pruning-based kernel clustering procedure. For each candidate solution, we rank players by their individual Silhouette values and retain only a highly reliable kernel subset, removing borderline cases with ambiguous membership. We perform bootstrap resampling and clustering refitting on this pruned sample, quantifying stability using the ARI (Adjusted Rand Index) between the resampled and reference partitions; only solutions with consistently high stability are retained as credible persona structures.
- **Back-mapping stable personas** onto the original behavioral traits and onto blocks inspired by a typology (in our case, Bartle) to assess their interpretability and plausibility. For each persona, we examine the average values and profiles across the original metrics (e.g., risk exposure, pace, exploration, resource utilization) and check whether these patterns align with the defined initial tendencies. This final step not only supports the segmentation validity but also provides a basis for analyzing differences in behavior and outcomes between the personas relevant to the design.

The methodology pipeline structure is shown in Figure 1.

### B. PARTICIPANTS AND GAME ENVIRONMENT

The empirical basis for this study is game telemetry from *Another Useless Zombie Game*, a single-player first-person shooter developed as a research prototype.

The game takes place in a small urban environment with several zombie spawners and a central factory area. The player moves freely in 3D space, shooting approaching zombies, collecting ammo and health packs, and supplying energy to charge the final exit portal, which is the main objective of each run. A preview of the game is in Figure 2.

The dataset contains records from 1,962 game sessions played by 156 unique, anonymized players. Each player is identified only by a pseudonymous identifier generated by the game; no personal or demographic information was collected for this article. Players were allowed to play multiple sessions, and there were no scripted scenarios or forced difficulty changes: all observations come from naturalistic play at the same level and with the same set of rules.

The study is based solely on secondary analysis of game telemetry and does not include any linkage to external personal data.

### C. LOGGED EVENTS

The game engine records player activity as a sequence of time-stamped events. Each log entry contains at least a player identifier, a session tag, an event type, a timestamp, and event-specific parameters. The basic types of events include:

- **Motion** events, which record the player's position and orientation over time.
- Shot events, including whether a shot hit a zombie, whether it was a head or body shot, and which zombie spawner the zombie came from.
- **Zombie interaction** events, such as zombie bites and zombie deaths.
- **Resource** events, including picking up ammo, batteries, and health packs.
- **Portal** and **spawner** events, including shots at spawners, spawner destruction, power supplies to an exit portal, and the final state of a portal.
- **Session boundary** events, indicating the start of a new game and its end (win, death, or user abort).

### D. DATA PREPROCESSING

As a first step of data preprocessing, we reconstructed individual game sessions from the raw logs. Events were sorted by timestamp and grouped by player identifier. A new session was started whenever a session start marker was recorded in the log, even if the same player identifier had appeared earlier, to ensure that repeated play by the same player is represented as multiple distinct sessions.

We filtered out incomplete or erroneous sequences, trivially short or inactive sessions that do not reflect meaningful play. Specifically, we removed games shorter than 40 seconds; this threshold corresponds to approximately half the fastest successful run and effectively excludes games in which players quit almost immediately after starting. After this filtering step, 855 sessions remained.

For each of these valid sessions, we then computed a set of session-level behavioral metrics (e.g., duration, movement, combat, risk, and resource indicators).

#### 1) AGGREGATION TO PLAYER-LEVEL FEATURES

To capture the model initial parameters, we transform preprocessed game event logs into player-level behavioral characteristics that can be meaningfully interpreted in terms of known player types. We build on Bartle's taxonomy and define four broad styles that are compatible with the mechanics of *Another Useless Zombie Game*. For each style, we identify which aspects of the identified behavior (movement, shooting, resource gathering, damage dealt, and portal interaction) best reflect its underlying tendencies, and aggregate these into numerical characteristics calculated per player and per session.

The definition of the parameters is based on the following considerations about the individual motivational groups of players:

- *Achiever* is a player focused on performance and goal fulfillment – trying to charge the final portal as quickly and efficiently as possible, optimizing battery collection, destroying spawners, and maximizing kill rate.
- *Explorer* is a player who prioritises exploring the space – moving a lot, covering a large area of the map, trying

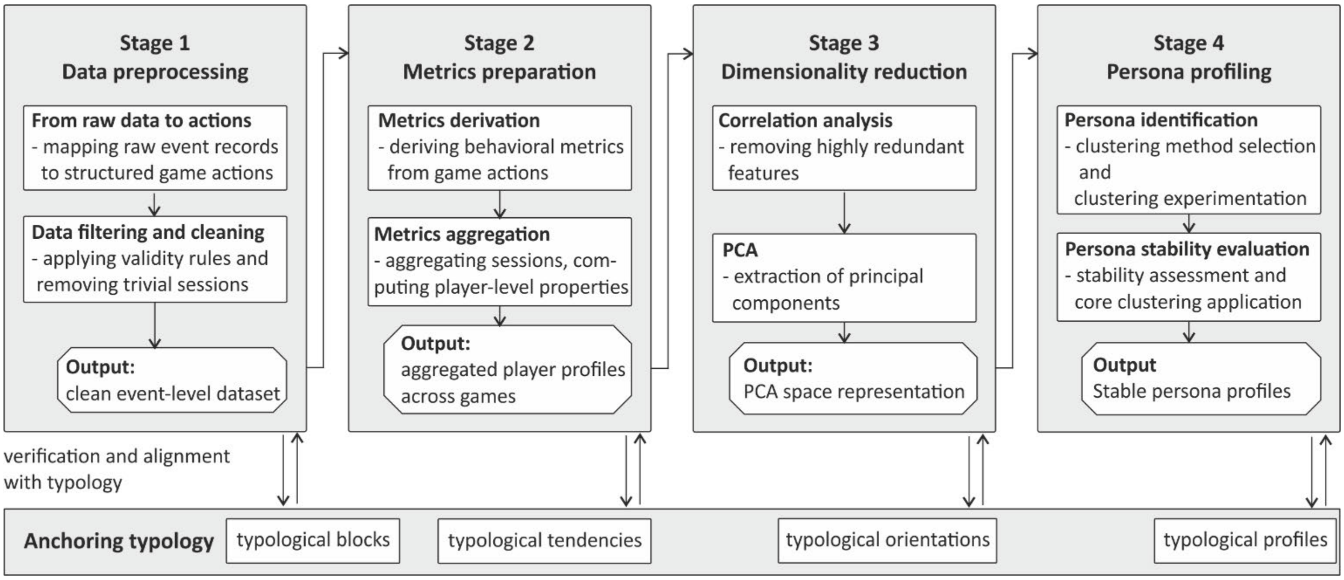



**FIGURE 1.** Visualization of a log-derived pipeline for processing game event logs, with connections to typology.

**FIGURE 2.** Game environment and visualization of a typical zombie attack scene.

different routes and zones, less focused on time-optimal completion.

- *Socializer* in the single-player version is best understood as a relaxed, experience-oriented style. What appears to be social interaction in the classic Bartle typology manifests here as a leisurely, low-competitive game: more walking without shooting and gathering resources primarily "for a sense of safety" than for maximizing performance. However, in this type of game (single-player shooter), we do not consider this profile a separate player type in our typology; it would be a forced inclusion based solely on speculative conclusions about the player's fundamental characteristics.
- *Killer* has an aggressive combat style: the player seeks out fights, often shoots on the move, gets close to zombies, risks bites, but achieves a high kill rate. He is interested in dominance in combat, not necessarily efficient progress to the portal.

The result of the conceptual mapping is Table 1, which summarizes all extracted player-level characteristics. For each feature, the table lists its name, the sources used to calculate it, a brief description of its significance in the game, the primary Bartle style(s) it reflects, and a conceptual description for the calculation.

We focus primarily on per-minute (speed-based) rates because raw event counts are heavily influenced by game duration and overall game time. A player who survives longer will almost inevitably accumulate more kills, items, bites, or shots, even if their actual playstyle or intensity is similar to that of a shorter game. By normalizing these quantities with respect to game time (e.g., kills per minute, path length per minute, bites per minute, item collects per minute), we obtain characteristics that more directly reflect stable behavioral tendencies such as combat pace, movement intensity, risk-taking, and resource seeking, rather than just exposure time. In contrast, inherently temporal indicators (e.g., time to first energy delivery, time to first bite, average time between bites) and bounded ratios (e.g., headshot ratio, shooting accuracy, bite-to-kill ratio) remain at the original scale because they already express behavior relative to time or success and would lose interpretability if transformed to minute measures.

From the preprocessed game event logs, we first computed a set of session-level behavioral characteristics for each game played. In total, we obtained 1,962 game sessions. To remove sessions that are too short to reflect stable behavior, we excluded all games with a duration of less than 40 seconds. This threshold is approximately half of the fastest successful run (80 seconds) and therefore discards interrupted or exploratory test runs while retaining all meaningful playthroughs. After this filtering step, 855 sessions remained.

For subsequent modeling, we aggregated the data to the player level. We retained all players who had at least one valid

**TABLE 1.** Behavioral metrics overview.

| Metric name | Description | Log source(s) | Bartle | Computation from game event logs |
|---|---|---|---|---|
| Time to first energy delivery | How quickly the player starts contributing to the main objective (charging the exit portal). | collect, portal, game start | A | `t(first portal visit) - t(game_start)` |
| Battery collection rate | Tempo of collecting energy batteries. | collect | A | `# batteries/playtime` |
| Time to destroy the first spawner | How fast the player neutralises the first zombie source. | shoot, spawner, game start | A | For first destroyed spawner: `t(destroy) - t(game_start)` |
| Mean spawner destruction time | Typical speed for destroying spawners. | shoot, spawner | A | For each spawner: `t(destroy) - t(first shot on that spawner);` then mean. |
| Spawner focus fire rate | Intensity of shooting at spawners while engaged with them. | shoot, spawner | A | For each spawner: `#shots_on_spawner / (t(destroy) - t(first shot));` mean |
| Headshot hits per minute | Number of accurate zombie hits per minute. | shoot | A/K | Count `shoot` events with `hit = true` and `hit_type = head` |
| Non-head hits per minute | Amount of less efficient hits on zombies (body) per minute. | shoot | A/K | Count `shoot` events with `hit = true` and `hit_type ≠ head` |
| Headshot ratio | Normalised precision measure: how often hits are headshots. | shoot | A/K | `headshot_hits / (headshot_hits + non_head_zombie_hits)` |
| Shooting accuracy | Shooting efficiency, ignoring hit location. | shoot | A/K | `#hits_total / #shots_total` |
| Number of shots at spawners | How strongly the player prioritises killing sources over individual zombies. | shoot | A/K | Count `shoot` events with `target_type = spawner` |
| Portal energy per minute | Speed of charging the exit portal over time. | portal | A | `total_energy_delivered_to_exit_portal / playtime` |
| Kills per minute | Tempo of eliminating zombies (combat performance over time). | shoot/kill | A/K | `#kills / playtime` |
| Total number of kills | Overall combat productivity within the match. | shoot/kill | A/K | Count of zombie kill events. |
| Path length per minute | How actively the player moves. | move | E | `total_path_length / playtime` |
| Total path length | Absolute amount of movement in the level. | move | E | Sum the distance between `move` positions |
| Session duration | How long the player stays in the game before the game ends. | all events | A/E | `t(last event) - t(first event)` |
| Area coverage | How much of the level the player explores. | move | E | Count unique cells visited in discretised map. |
| Health pickups per minute | How often the player actively restores health (self-preservation). | collect | E | Count `collect` with `item_type = health/heart` |
| Ammo pickups per minute | Tendency to manage ammunition and search resources. | collect | A/E | Count `collect` with `item_type = ammo` |
| Minutes ratio without shooting | Ratio spent moving/observing without combat (cautious or exploratory play). | shoot, time | E | Count of 1-minute bins without `shoot` events |
| Mean path length without shooting | Typical length of "peaceful walks" between combat episodes. | move, shoot | E | The mean of path lengths between shots. |
| Number of bites taken | How often do zombies get close enough to hit. | bite | K/E | Count of `bite` events. |
| Mean time between bites | Escape behavior after being hit (do they stay in danger?). | bite | K/E | Sorting `bite` events by time, computing deltas `t(i+1) - t(i),` then the mean. |
| Camping vs roaming index | Distinguishes players staying in a small area vs moving around the whole map. | move | E/K | `(#unique_grid_cells_visited)/ total_path_length` |
| Total games/sessions per user | Overall engagement: how many times the player comes back to play. | session metadata | All | Count unique `session_id` for `user_id` |
| Game ending type | Whether the player usually finishes the objective, dies, or quits. | final, death, exit | A/K/E | Categorical label from last event in session: `{final_portal_open, death, leave}` |
| Shots while moving per minute | Aggressive kiting style (shooting without stopping). | move, shoot | K | For `shoot` event, player speed from nearby `move` events; shots with `speed >threshold` |
| Backward movement per minute | How much the player retreats while facing enemies. | move (with rotation) | K | Information is stored in the final event. |
| Mean shooting distance | Preferred combat distance (close-range vs safe long-range). | shoot, target positions | K/E | For each hit: distance between player position at shot time and target position, then mean. |
| Time to first bite | How long the player avoids getting hit; indicator of risk or caution. | bite, game start | K/E | `t(first bite) - t(game_start);` if no bite, result is full playtime |
| Risk ratio (bites per kill) | Balance between taking damage and being effective in combat. | bite, kill | K | `#bites / max(#kills, 1);` higher = more damage per achieved kill |

($\geq$40 sec) session, resulting in a final sample of 156 players. For each player, we computed the mean of all session-level numerical characteristics across their retained sessions and also recorded the number of games included as an additional engagement descriptor. The resulting player-level dataset contains one row per player and was used as input for clustering and further analyses.

### 2) DESCRIPTIVE STATISTICS AND VISUALISATION OF PLAYER-LEVEL FEATURES

To gain an overall picture of the resulting player-level dataset, we calculated basic descriptive statistics (mean, standard deviation, minimum, maximum, median) for all characteristics. We grouped them into four conceptual categories: Engagement (number and duration of game sessions), Exploration (movement and spatial coverage), Performance (combat and tactical effectiveness), and Resources (gathering and use of batteries, ammunition, and health). Table 2 presents these descriptive statistics and shows considerable variability within each category, with extensive ranges for indicators related to game time, movement, and combat performance.

Several features exhibited strongly right-skewed distributions, with significant differences between the median and maximum values (e.g., total path length, portal energy per minute, time-based spawner metrics, and various count-based measures such as kills or bites). To reduce the influence of extreme outliers and stabilise variance, we applied a $\log(1 + x)$ transformation to all strictly non-negative count and duration features in the Engagement, Exploration, Performance, Risk, and Resources categories, while leaving bounded ratios (e.g., headshot ratio, shooting accuracy) on their original scale.

Figure 3 illustrates the raw distributions of four representative features (portal energy per minute, total path length, total number of kills, and risk ratio bites per kill), highlighting the strong right skew that motivates the use of a logarithmic transformation. All subsequent analyses are based on this transformed feature set.

**TABLE 2. Descriptive statistics of player characteristics.**

| | mean | std | min | max | median |
|---|---|---|---|---|---|
| **Engagement** | | | | | |
| Total games/sessions per user | 12.6 | 53.4 | 1 | 560 | 3 |
| Session duration | 409 | 306.4 | 45.8 | 2261 | 347.3 |
| Game-length>40 s | 5.5 | 12.2 | 1 | 92 | 2 |
| Game - won | 1.2 | 2.6 | 0 | 20 | 1 |
| Game - dead | 2.6 | 6.5 | 0 | 55 | 1 |
| Game - interrupt | 1.7 | 4.4 | 0 | 43 | 1 |
| **Exploration** | | | | | |
| Path length per minute | 472.5 | 221.2 | 172.6 | 2736.8 | 444.4 |
| Total path length | 3118 | 2372.2 | 231.8 | 18276 | 2630.4 |
| Area coverage | 213.1 | 144 | 21.8 | 1044 | 187.3 |
| Minutes ratio without shooting | 0.1 | 0.3 | 0 | 3.1 | 0.1 |
| Mean path length without shooting | 35.3 | 169.4 | 2.1 | 1956.6 | 10.6 |
| Camping vs roaming index | 0.1 | 0.3 | 0 | 2.6 | 0.1 |
| **Performance** | | | | | |
| Time to destroy the first spawner | 252.6 | 241.5 | 11.5 | 1308.8 | 177.3 |
| Mean spawner destruction time | 195.5 | 217.7 | 0.6 | 1224 | 122.3 |
| Spawner focus fire rate | 1.7 | 2.2 | 0 | 14.2 | 0.8 |
| Headshot hits per minute | 6.8 | 5.4 | 0 | 30.1 | 5.8 |
| Non-head hits per minute | 14.9 | 10.9 | 0 | 53.5 | 11.9 |
| Headshot ratio | 0.3 | 0.2 | 0 | 1 | 0.3 |
| Shooting accuracy | 0.5 | 0.2 | 0 | 1 | 0.6 |
| Number of shots at spawners | 12.4 | 12.6 | 0 | 56 | 8.3 |
| Portal energy per minute | 76.1 | 169 | 0 | 1310 | 40.8 |
| Kills per minute | 9.2 | 6.6 | 0 | 46 | 8.1 |
| Total number of kills | 58.5 | 48.8 | 0 | 399 | 48.5 |
| Number of bites taken | 14.5 | 10.4 | 0 | 63 | 13.7 |
| Mean time between bites | 32.6 | 63.5 | 0 | 692 | 18.9 |
| Shots while moving per min. | 34.6 | 19.3 | 0 | 106.5 | 32.8 |
| Backward movement per minute | 35.2 | 18.3 | 0 | 101.5 | 34.9 |
| Mean shooting distance | 8.1 | 2.8 | 3.7 | 18.4 | 7.8 |
| Time to first bite | 102.9 | 63.8 | 13.9 | 390.2 | 85.5 |
| Risk ratio (bites per kill) | 0.6 | 1 | 0 | 5.9 | 0.3 |
| **Resources** | | | | | |
| Time to first energy delivery | 184.4 | 173.2 | 11.5 | 1308.8 | 142 |
| Battery collection rate | 0.1 | 0.1 | 0 | 0.4 | 0.1 |
| Health pickups per minute | 6.4 | 6.9 | 0 | 42 | 4.7 |
| Ammo pickups per minute | 14.6 | 15.3 | 0 | 102 | 10.5 |

## IV. DATA ANALYSIS

Based on the feature engineering and aggregation described above, we now analyze the transformed player-level characteristics to identify systematic patterns of behavior in the game and prepare a representation suitable for clustering.

### A. DIMENSIONALITY REDUCTION

To obtain a compact, numerically stable representation of player behavior, we first reduce the feature space dimensionality and then proceed to clustering.

### 1) FEATURES CORRELATION

As a first step towards dimensionality reduction, we examined the correlation structure of the transformed player-level characteristics. Using the $\log(1 + x)$-transformed attributes, we computed a Pearson correlation matrix and visualized it as a heat map (Figure 4) to identify groups of strongly correlated variables and detect redundancy between the indicators. In line with standard practice, we consider absolute correlations with $|r| \, | > 0.7$ to indicate substantial redundancy.

The correlation matrix revealed several coherent clusters of strongly correlated traits ($|r| > 0.7$), which we can interpret as higher-level behavioral factors.

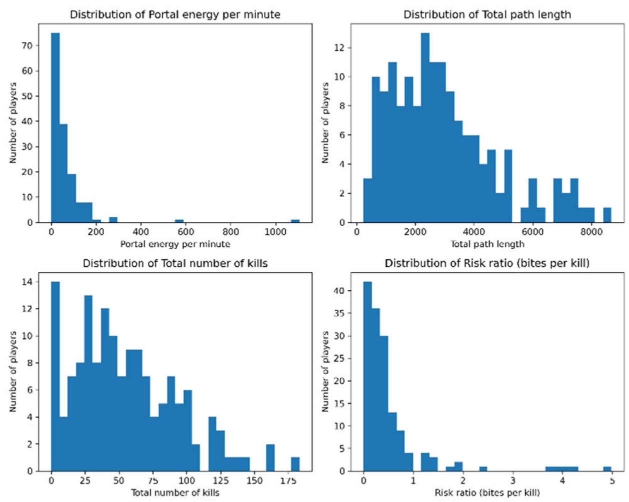



**FIGURE 3. Raw distributions (histograms) of four representative behavioral features at the player level: (a) portal energy per minute, (b) total path length, (c) total number of kills, and (d) risk ratio (bites per kill). All four measures are strongly right-skewed, with a small subset of players accounting for the extreme values.**

The **first group** captures the contrast between safe movement and aggressive combat. It contains the correlated variables *Non-head hits per minute* and *Shots while moving per min,* and a negative correlation with *Mean path length without shooting*. Players who spend much time in long ''walking segments'' without shooting tend to have fewer hits and fewer shots while moving, while players who often shoot while moving accumulate many non-head hits and only short walking segments. In the following analyses, we keep *Shots while moving per min* and *Mean path length without shooting* as complementary poles of this factor, while *Non-head hits per minute* is omitted because it is largely redundant in the ratio of kills per minute to head hits.

The **second group** combines time to first spawner destruction with time to first energy delivery. Both variables describe how quickly a player starts interacting with primary targets and are highly correlated. To avoid duplication of information, we retain *Time to first energy delivery*, dropping the *Time to destroy first spawner* variable.

The **third group** consists of the *Mean spawner destruction time*, *Number of shots at spawners*, and *Spawner focus fire rate*. Together, these measures describe how effectively and intensely players attack a spawner: longer destruction times are associated with fewer shots per second and more dispersed fire, while high-concentration fire reflects bursts of fire until the spawner is destroyed. For modeling purposes, we retain *Mean spawner destruction time* as a compact and interpretable summary of the effectiveness of engaging a spawner and omit the two highly correlated variants based on number and duration.

The **larger cluster** groups variables related to overall activity and game volume, including *Session duration*, *Total path length*, *Area coverage*, *Headshot hits per minute*, *Kills per minute*, *Total number of kills*, *Number of bites taken*, *Health pickups per minute,* and *Ammo pickups per minute*. The high correlations here primarily reflect the fact that players who survive longer and move more also kill more zombies, collect more resources, and receive more bites. Rather than summarizing this cluster into a single index, we selected a small set of conceptually distinct proxy measures: exploration (*Area coverage*), combat pace (*Kills per minute*), and risk-taking (*Number of bites taken*). Highly overlapping measures of ''volume'', such as total kills and headshots per minute, are excluded from the final set of features because the selected proxy measures already capture their information.

Finally, the **meta-level group** combines *Total games per user* (and *Games above 40 seconds*) with game results (*won*, *dead*, *interrupted*), reflecting long-term engagement and typical session outcomes for each player. Instead of using raw counts for each outcome type, we retain a single measure of overall engagement (*Total games per user*) and derive the proportions of sessions ending in death or abandonment (*death rate* and *abandonment rate*) as more interpretable indicators of how players typically end their runs.

These transformed engagement features, along with behavioral proxies from the other clusters, form a reduced set of features used for subsequent principal component analysis and clustering.

### 2) PCA

We included all remaining behavioral features (21 dimensions) in the PCA, as strongly redundant variables had already been removed based on the correlation analysis. PCA was therefore used not to filter out trivial duplicates, but to capture broader patterns of co-variation among movement, combat, risk, and resource-related indicators in fewer orthogonal components.

Table 3 summarizes the proportion of variance explained by each component. The first principal component (PC1) accounted for 19.9% of the total variance, the second component (PC2) for 17.1%, and the third (PC3) for 10.4%. The first three components together accounted for 47.4% of the variance, and the first five components for 63.5%. The curve then flattened, with each additional component contributing approximately 2–4% of the variance.

To balance parsimony and information loss, we retained the first twelve principal components for subsequent analyses, which together explained 91.0% of the total variance. This selection reduced the original 21 features to 12 orthogonal components, while preserving most of the variability in player behavior.

To support the interpretation of the PCA solution, we inspected the principal component loadings, and only those with absolute values $\geq$0.30 are reported in Table 4. Below, we also summarise the first five components.

PC1 showed strong positive loadings for kills per minute (0.44), shots while moving per minute (0.35), ammo collection per minute (0.33), shooting accuracy (0.31), and number of bites (0.30), and a negative loading for average length of

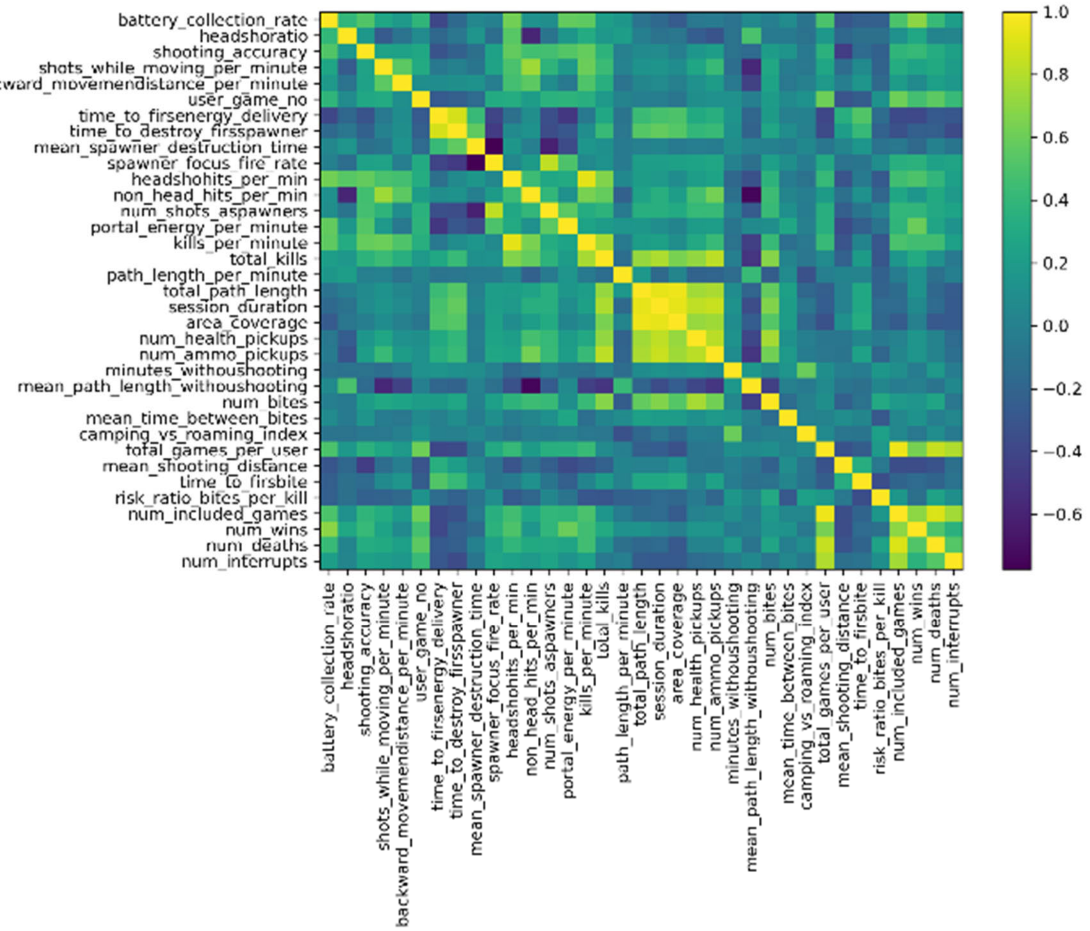



**FIGURE 4.** **Heatmap of Pearson correlations between transformed player-level behavioral features.**

path without shooting (−0.33). Players with high PC1 scores, therefore, engage in frequent and effective combat, shoot a lot while moving, constantly replenish ammo, and get into many close encounters (bites), while spending relatively little time just walking around without shooting. Conversely, low PC1 scores characterize a more passive game with longer stretches of non-combat movement and lower combat throughput.

PC2 compared an efficient, resource-focused approach with a more exploratory, wandering approach. It had positive loadings for time to first energy delivery (0.43), area coverage (0.32), and time to first bite (0.29), and negative loadings for battery collection rate (−0.37), portal energy per minute (−0.32), and proportion of games lasting more than 40 seconds (−0.33). Thus, high PC2 scores correspond to players who move a lot, delay their first energy delivery, and encounter enemies relatively late, while gathering resources less efficiently and achieving longer games less often. Low PC2 scores characterize more experienced, goal-oriented players who quickly collect batteries, generate portal energy, and achieve longer runs with less unnecessary wandering.

PC3 captured a distinct dimension of camping-related risk. It showed strong positive loadings for the camping vs. roaming (0.45), bite risk ratio per kill (0.47), and minutes without shooting ratio (0.34), and a strong negative loading for path length per minute (−0.41) with a smaller negative loading for time to first bite (−0.27). Players with high PC3 scores tend to move very little, spend long periods in static positions without shooting, and allow zombies to get close enough to bite them frequently relative to the number of kills, while often being bitten early in the run. Conversely, low PC3 scores characterize more mobile players who are constantly moving, engage more continuously, and receive relatively few bites per kill.

PC4 groups indicate proactive control over the course of the level. Positively weighted were area coverage (0.41), portal energy per minute (0.39), average time between bites (0.36), and amount of ammo gained per minute (0.34), and negatively weighted were backward movement per minute (−0.36) and average time to destroy spawners (−0.33). High PC4 scores, therefore, correspond to players who move around the map, destroy spawners relatively quickly, collect

**TABLE 3. The proportion of variance explained by each component.**

| PC | explained variance ratio | cumulative variance |
|---|---|---|
| PC1 | 0.199 | 0.199 |
| PC2 | 0.171 | 0.37 |
| PC3 | 0.104 | 0.474 |
| PC4 | 0.085 | 0.559 |
| PC5 | 0.076 | 0.635 |
| PC6 | 0.065 | 0.701 |
| PC7 | 0.058 | 0.758 |
| PC8 | 0.039 | 0.798 |
| PC9 | 0.035 | 0.833 |
| PC10 | 0.03 | 0.863 |
| PC11 | 0.025 | 0.888 |
| PC12 | 0.022 | 0.91 |
| PC13 | 0.021 | 0.931 |
| PC14 | 0.016 | 0.946 |
| PC15 | 0.013 | 0.959 |
| PC16 | 0.011 | 0.97 |
| PC17 | 0.009 | 0.978 |
| PC18 | 0.007 | 0.985 |
| PC19 | 0.006 | 0.991 |
| PC20 | 0.005 | 0.996 |
| PC21 | 0.004 | 1 |

**TABLE 4. The influence of the first five principal components on behavioral traits (only loadings with |loading| ≥ 0.30 are shown).**

| Category / Feature | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| **Engagement** | | | | | |
| Game-length>40 s | | -0.33 | | | |
| **Exploration** | | | | | |
| Path length per minute | | | -0.41 | | |
| Area coverage | | 0.32 | | 0.41 | |
| Minutes ratio without shooting | | | 0.34 | | 0.5 |
| Mean path length without shooting | -0.33 | | | | |
| Camping vs roaming index | | | 0.45 | | 0.36 |
| **Performance** | | | | | |
| Mean spawner destruction time | | | | | 0.47 |
| Headshot hits per minute | | | | | |
| Shooting accuracy | 0.31 | | | | |
| Portal energy per minute | | -0.32 | | 0.39 | |
| Kills per minute | 0.44 | | | | |
| Number of bites taken | 0.3 | | | | |
| Mean time between bites | | | | 0.36 | |
| Shots while moving per min. | 0.35 | | | | |
| Backward move-ment per min. | | | | -0.36 | |
| Mean shooting distance | | | | | |
| Time to first bite | | 0.29 | -0.27 | | |
| Risk ratio (bites per kill) | | | 0.47 | | |
| Resources | | | | | |
| Time to first energy delivery | | 0.43 | | | |
| Battery collection rate | | -0.37 | | | |
| Ammo pickups per minute | 0.33 | | | 0.34 | |

**TABLE 5. Internal validity indices for k = 2..8.**

| k | Silhouette score | Calinski-Harabasz index | Davies-Bouldin index |
|---|---|---|---|
| 2 | 0.267 | 24.874 | 2.148 |
| 3 | 0.174 | 25.242 | 1.793 |
| 4 | 0.191 | 23.511 | 1.528 |
| 5 | 0.148 | 23.401 | 1.696 |
| 6 | 0.168 | 23.761 | 1.484 |
| 7 | 0.161 | 22.939 | 1.484 |
| 8 | 0.152 | 23.468 | 1.422 |

portal energy and ammo at a steady pace, and keep zombies at a safe distance (longer intervals between bites) without backing up excessively. Low PC4 scores reflect a more reactive, defensive style with more backward movement, slower spawner destruction, and less sustained resource collection.

PC5 emphasized an extremely slow and static pace of play. It showed strong positive loadings for minutes without shooting (0.50), average time to destroy spawners (0.47), and camping vs. roaming index (0.36). Players with high PC5 scores spend long periods of inactivity without shooting, stay in camping positions, and destroy spawners very rarely. Conversely, low PC5 scores indicate a fluid pace with fewer long periods of inactivity, faster spawner destruction, and less static gameplay.

Overall, the principal components summarize the original behavioral traits into a compact set of dimensions that can be used as an intermediate representation for player grouping.

## B. CLUSTER ANALYSIS

### 1) CLUSTERS NUMBER IDENTIFICATION

After extracting the first 12 principal components (PC1–PC12, which explain 91% of the variance), we clustered players in this PCA space using k-means with k = 2-8 clusters. For each solution, we calculated three standard internal validity indices: the silhouette coefficient, the Calinski-Harabasz index, and the Davies-Bouldin index. Table 5 summarizes the results.

The two-cluster solution achieved the highest silhouette coefficient (0.27) and a relatively high Calinski-Harabasz index (24.9), but produced a very coarse dichotomy that merged clearly distinct play styles into broad "more vs. less active" groups. Among solutions with $k \geq 3$, the silhouette coefficient ranged from 0.15 to 0.19, with local maxima at k = 3 (0.17), k = 4 (0.19), and k = 6 (0.17). The Calinski-Harabasz index was highest for k = 3 (25.2), but the values for k = 4–8 were very similar (approximately 23–24), suggesting that there is no clear single optimum. The Davies-Bouldin index, which indicates better-separated clusters, increased almost monotonically with more clusters, from 2.15 at k = 2 to 1.42 at k = 8. Together, these indices suggest that several solutions in the range of k = 3–6 provide similarly good trade-offs between compactness and separation, and that increasing k beyond this range yields diminishing returns in internal validity.

To complement centroid-based clustering, we also fit Gaussian mixture models with varying numbers of components and selected the optimal number using the Bayesian Information Criterion (BIC). The BIC was minimized for an eight-component mixture ($k = 8$, BIC = 5766.1, AIC = 3548.9), suggesting that a relatively fine-grained Gaussian approximation of the density in PCA space requires more than a few components. However, the mixture components do not necessarily correspond to psychologically significant individuals: additional Gaussian curves can capture local skewness or heavy tails within broader styles, rather than truly distinct types of play.

We next used DBSCAN for density-based clustering on the exact PCA representation. Over a wide range of parameter settings, DBSCAN assigned all players to noise (labels, counts: $\{-1: 156\}$), i.e., it did not find any stable islands of high density that would qualify as well-separated clusters suggesting that the player behavior in our dataset forms a relatively smooth continuum, rather than a set of sharply defined ‘‘natural’’ clusters in terms of density.

As an additional check on the dimensionality of the behavioral feature space, we performed a Horn parallel analysis of the original and logarithmically derived features (Figure 5).

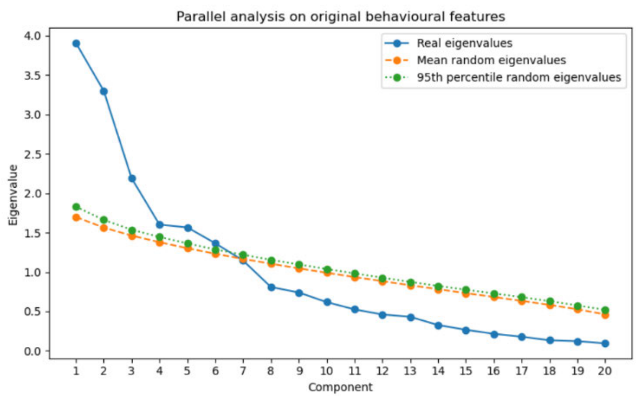



**FIGURE 5.** **Parallel analysis of the original behavioral features.**

Although parallel analysis formally informs the number of latent dimensions rather than the number of clusters, the finding of five to six substantive components suggested that solutions with $k = 5$–6 personas would provide a reasonable compromise between granularity and interpretability, and we therefore examined k-means solutions in this range in more detail.

## 2) CLUSTERS PROFILES

Based on the PCA results, we used the scores of the first twelve principal components (PC1–PC12), which together explained 91% of the total variance, as an orthogonal representation of player behavior for clustering. Since our goal was to obtain a small number of compact and reasonably spherical characters in this continuous behavioral space, we chose k-means clustering with Euclidean distance as the primary method. Based on the indices described above, we focused on solutions with medium granularity ($k = 5$ and $k = 6$).

To characterize the resulting clusters, we examined their average scores on the original behavioral traits and visualized them using radar plots. As part of the analysis, we first examined a five-cluster k-means solution that divided the 156 players into clusters of sizes 77, 40, 31, 4, and 4 (Figure 6).

For each solution, we first standardized the selected game metrics across all players using z-scores. Then we plotted the averages across clusters, with values above zero indicating above-average behavior and values below zero indicating below-average behavior on each axis.

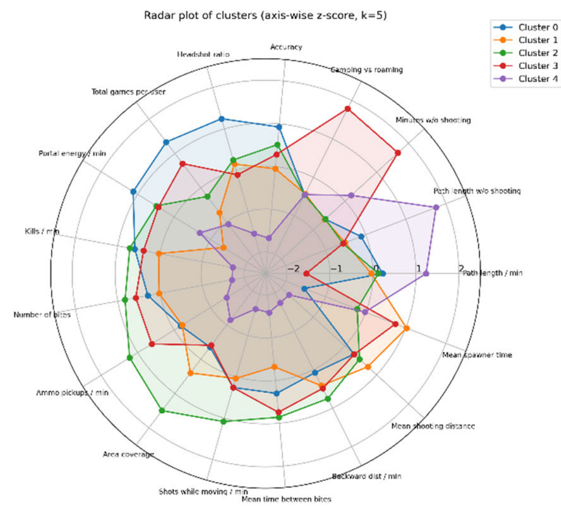



**FIGURE 6.** **Mean behavioral profiles of player clusters ($k = 5$).**

Using z-scores along the axes of the radar plot (each axis normalized across clusters), three clusters appear to be the main characters, while two are tiny edge cases:

Cluster 2 ($n = 77$) – members have the highest area of coverage, kills per minute, shots while moving, and backward movement. They also have above-average accuracy, portal energy, ammo, and battery collection, with low movement without shooting.

Cluster 0 ($n = 31$) – contains members with the highest total games played, shooting accuracy, and headshot ratio, fast spawner kills, and early first energy delivery. They also have a decent kill rate, but less extreme roaming than Cluster 2. The group looks like a group of more experienced, efficient players who ‘‘play clean’’ rather than maximizing raw power.

Cluster 1 ($n = 40$) – consists of players with low total games played, average movement, medium area coverage, accuracy, and kills, but very low inactive time. They also show increased death and interruption rates. These players move and shoot, but they convert their activity into objectives poorly and often interrupt runs.

Cluster 3 (n = 4) – represents a small group with a very high proportion of time without shooting, an extremely high camping index, and below average path length.

Despite some portal energy and kills, they have a very long time between bites and a high death rate. They are extreme "stay in one place and minimize risk" players; their profiles differ significantly from those of the others.

Cluster 4 (n = 4) – represents the next small cluster with the highest path length per minute and no-shoot path length, in contrast with low area coverage, and strongly low scores in almost all combat metrics (accuracy, headshots, kills, bites, portal energy, ammo collection). They rarely die and seldom get bitten - they mostly move around without encountering enemies.

Overall, the k = 5 solution yields three large, interpretable characters and two small, extreme clusters that capture rare behavior at the edges of space.

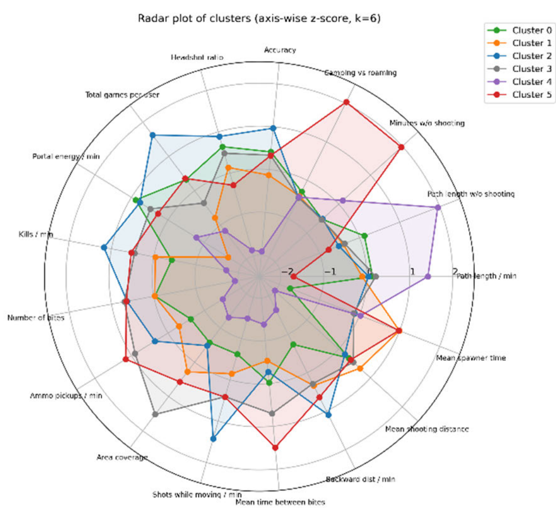



**FIGURE 7.** **Mean behavioral profiles of player clusters (k = 6).**

The six-cluster solution produces a very similar overall structure (Fig. 5). The cluster sizes are 68, 37, 24, 20, 4, and 3. Compared to k = 5, the additional granularity mainly affects the large, high-activity group, while the small edge clusters remain.

Among high-performance players, the previous single-cluster "strong fighter" splits into two subtypes. One cluster (Cluster 2, n = 24) shows the highest kill rate per minute, the highest shooting accuracy and headshot ratio, and a substantial portal energy gain, corresponding to the most efficient killers. The second cluster (Cluster 3, n = 68) has a slightly lower but still good performance, with high area coverage, solid accuracy, and portal energy, and a relatively low number of games per user.

Two small edge clusters persist. Cluster 4 (n = 4) remains essentially the same "low-engagement walker" profile as in the k = 5 solution, and Cluster 5 (n = 3) becomes an even more extreme camping profile with a very high camping index and no-shooting time, short journey length, increased mortality, and many bites per kill.

Moving from five to six clusters does not eliminate microclusters; it retains the rare "walkers" and "campers" and divides the high-performance area into two finer subtypes.

### 3) CLUSTERS STABILITY

Bootstrap stability was evaluated using the Adjusted Rand Index (ARI) with 300 resamples. For k = 5, the mean ARI was 0.51 (median 0.52); for k = 6, the corresponding values were 0.53 and 0.50. Both solutions achieved only moderate replicability. Moreover, the k = 6 solution produced two tiny clusters (n = 3 and 4), whereas the k = 5 solution yielded three sizeable clusters (n = 31, 40, and 77) and two tiny clusters (n = 4 each). Given the negligible ARI difference and the presence of micro-clusters in both solutions, we retained k = 5 as the more parsimonious option, interpreting the three large clusters as core personas and the two tiny clusters as rare edge cases.

To separate the stable cluster cores from the noisy variability typical of human behavior, we used a silhouette-based pruning procedure. For solutions with k = 5 and k = 6, we calculated silhouette scores for all players and ranked players within each cluster. This rank allowed us to retain only the high-silhouette "core" members of each cluster and test whether this core reproduces a clustering structure similar to that of the full sample.

We defined a series of core samples by retaining only the highest q proportion of players per cluster (q = 0.70–0.95), i.e., those with the highest cohesion within the cluster. Each trimmed core sample was reclassified with the same k-means settings, and the reproducibility of the cluster assignment within each core was assessed using bootstrap ARI with 300 resamplings.

For k = 5, keeping the top 70% of players per cluster yielded a core of 110 members, with a mean silhouette score of 0.235 and a median bootstrap ARI of 0.924 (mean ARI = 0.887). Increasing q to 0.80 and 0.90 gradually increased the core (127 and 143), but at the cost of lower coherence (silhouette score: 0.21 and 0.17) and reduced stability (median ARI dropped to 0.83 and 0.72).

The pattern for k = 6 was very similar: at q = 0.70, the core consisted of 111 players with a silhouette score of 0.242 and a median ARI of 0.882, while larger cores again led to monotonically decreasing stability (Table 6).

The most coherent 70–75% of players in each cluster form the core of the sample, in which the k-means solution is highly reproducible (median ARI = 0.9), while the addition of low-silhouette players mainly introduces ambiguity and noise. Since the stability profiles of k = 5 and k = 6 were comparable, and k = 5 already provided a more transparent, more interpretable persona structure (three large clusters and two rare edge cases), we adopt the k = 5 solution with q = 0.70 as our primary model.

**TABLE 6.** The proportion of variance explained by each component.

| q | N core | Silhouette score | ARI median | ARI mean | ARI SD |
|---|---|---|---|---|---|
| k = 5 | | | | | |
| 0.7 | 110 | 0.235 | 0.924 | 0.887 | 0.119 |
| 0.75 | 118 | 0.219 | 0.9 | 0.85 | 0.14 |
| 0.8 | 127 | 0.216 | 0.83 | 0.791 | 0.158 |
| 0.85 | 135 | 0.196 | 0.759 | 0.729 | 0.169 |
| 0.9 | 143 | 0.166 | 0.723 | 0.672 | 0.175 |
| 0.95 | 151 | 0.166 | 0.611 | 0.592 | 0.194 |
| k = 6 | | | | | |
| 0.7 | 111 | 0.242 | 0.882 | 0.851 | 0.12 |
| 0.75 | 118 | 0.227 | 0.866 | 0.833 | 0.129 |
| 0.8 | 128 | 0.212 | 0.819 | 0.771 | 0.167 |
| 0.85 | 135 | 0.2 | 0.773 | 0.737 | 0.172 |
| 0.9 | 144 | 0.184 | 0.685 | 0.673 | 0.172 |
| 0.95 | 151 | 0.165 | 0.559 | 0.58 | 0.163 |

To verify that pruning primarily removes ambiguous cases, rather than altering the underlying persona structure, we compared the solution with k = 5 for the entire sample with the corresponding core with q = 0.70. In all five clusters, the resulting profiles were remarkably similar: differences in individual features were minor, and the arrangement of clusters along key behavioral dimensions remained virtually unchanged. Figure 8 shows a radar plot of core profiles, which can be compared with the nearly identical personas outlined in Figure 6.

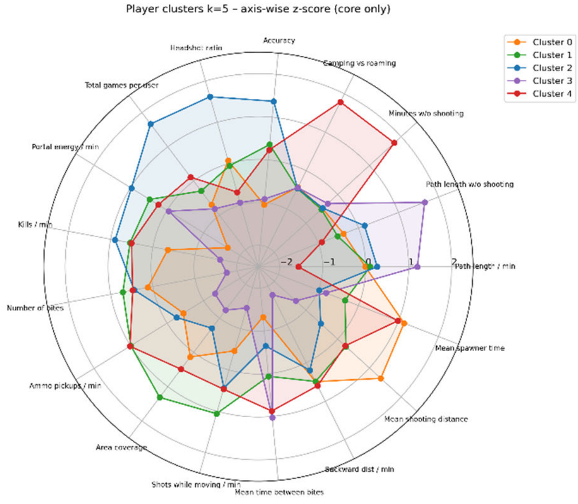



**FIGURE 8.** Mean behavioral profiles of core player clusters (k = 5).

The core analysis serves as a robustness check, and the result is also consistent with the ARI results: after removing low-profile players, the remaining 70–75% of players per cluster form a highly stable core in which the k-means assignment is almost perfectly reproducible, while the added 25–30% of outliers contribute mainly to noise and local instability.

### C. MATCHING PERSONA PROFILES WITH BARTLE TYPES

To connect the five personas to established player types, we summarized their profiles using all 21 behavioral characteristics and grouped them into four categories: Engagement, Exploration, Performance, and Resources. Figure 9 shows the resulting heat map, where green indicates higher values and red indicates lower values within each trait.

| Category | Bartle | Feature | P1 | P2 | P3 | P4 | P5 |
|---|---|---|---|---|---|---|---|
| Engagement | | Game-length>40 s | 2.51 | 1.04 | 1.38 | 2.06 | 0.79 |
| Exploration | E | Path length per minute | 6.15 | 6.07 | 6.12 | 5.59 | 6.47 |
| Exploration | E | Area coverage | 4.69 | 5.02 | 5.51 | 4.66 | 4.34 |
| Exploration | E/S | Minutes ratio without shooting | 0.08 | 0.07 | 0.08 | 0.75 | 0.32 |
| Exploration | E/S | Mean path length without shooting | 3.07 | 2.43 | 2.34 | 2.35 | 6.09 |
| Exploration | E/K | Camping vs roaming index | 0.07 | 0.08 | 0.07 | 1.06 | 0.07 |
| Performance | A | Mean spawner destruction time | 3.66 | 5.14 | 4.43 | 4.98 | 4.54 |
| Performance | A/K | Headshot hits per minute | 0.49 | 0.3 | 0.32 | 0.25 | 0 |
| Performance | A/K | Shooting accuracy | 0.68 | 0.42 | 0.57 | 0.51 | 0 |
| Performance | A | Portal energy per minute | 4.77 | 0.68 | 3.73 | 3.63 | 1.75 |
| Performance | A/K | Kills per minute | 2.22 | 1.67 | 2.33 | 2.02 | 0 |
| Performance | K/E | Number of bites taken | 2.33 | 2.09 | 2.81 | 2.58 | 0.55 |
| Performance | K/E | Mean time between bites | 2.94 | 2.65 | 3.2 | 3.14 | 2.06 |
| Performance | K | Shots while moving per minute | 30.07 | 26.81 | 42.46 | 30.1 | 1.54 |
| Performance | K | Backward movement per minute | 29.54 | 34.55 | 39.53 | 35.61 | 2.64 |
| Performance | K/E | Mean shooting distance | 1.97 | 2.39 | 2.14 | 1.98 | 0 |
| Performance | K/E | Time to first bite | 3.94 | 4.7 | 4.6 | 4.24 | 4.35 |
| Performance | K | Risk ratio (bites per kill) | 0.52 | 0.43 | 0.27 | 0.78 | 0 |
| Resources | A | Time to first energy delivery | 3.8 | 5.41 | 5.08 | 4.92 | 4.54 |
| Resources | A | Battery collection rate | 0.15 | 0.03 | 0.08 | 0.06 | 0.06 |
| Resources | A/E | Ammo pickups per minute | 1.72 | 1.69 | 2.97 | 2.43 | 0.6 |

**FIGURE 9.** Heatmap of behavioral traits across player personas (k = 5).

Three significant clusters represent the core of the population, while two tiny clusters capture rare extremes:

- **P1 – Efficient achievers** (n = 31) show consistently high values of the Achiever-coded performance and resource indicators. These players played an above-average number of games, reached the objective (recharging portal energy) at an above-average rate, had the highest shooting accuracy and headshot frequency, and had solid kill and ammo collection rates. Their exploration metrics are moderate rather than extreme: they move and efficiently traverse levels without excessive wandering or camping. Overall, P1 exhibits a largely Achiever-like profile with only a moderate Killer component.
- **P2 – Learning wanderers** (n = 40) maintain comparable movement and area coverage to the other large clusters, but lag in most Achiever metrics: lower accuracy, headshot rate, portal energy recharge, and kill rate. Despite active movement and shooting, these players are less effective at converting activity into progress and often play shorter games. Their profile combines moderate Explorer and Killer signals with weaker Achiever performance, consistent with an unclassifiable persona or learning situation.
- **P3 – Active warriors** (n = 77) exhibit the most extreme levels of activity and combat. These players have the longest firing range, the highest shots per minute while moving, strong area coverage, high kill and ammo acquisition rates, and low bites per kill. This combination of high throughput and high risk marks P3 as a clear hybrid

of Achiever and Killer, representing the most aggressive combat-oriented playstyle.

- **P4 – Extreme campers** (n = 4) are characterized by a very high camping vs. roaming index and a large proportion of time and distance traveled without shooting. Despite relatively long games and non-trivial progression, these players move predominantly from static positions, leaving them vulnerable to enemy approach, increasing the risk of being bitten. Their pattern corresponds to a rare, extreme camping-oriented one that is not well captured by the classic Achiever/Killer/Explorer labels, but is clearly distinct in the behavioral space.
- **P5 – Low-engagement walkers** (n = 4) exhibit the highest distance traveled per minute, combined with almost zero combat engagement: very low kills, accuracy, bites, portal energy, and resource harvest rates. These players spend most of their time moving around, often without exploring new areas or engaging enemies or objectives. This persona resembles an explorer-like style with minimal Achiever or Killer components.

The cluster profiles in Figure 9 together show that five personas can be meaningfully described using the predefined A/E/K indicators: two large personas are predominantly achievement-oriented (P1, P3), one combines moderate elements of all three tendencies (P2), and two small personas represent extremes focused on exploration or enjoying passivity (P4, P5).

## V. DISCUSSION

This study aimed to demonstrate and schematize a data-driven approach to modeling players from game event logs. In the following sections, we discuss the results of each phase of this process, organized around our four research questions.

### A. RQ1: BEHAVIORAL DIMENSIONS FROM DERIVED CHARACTERISTICS

RQ1 asked whether the derived characteristics yield latent behavioral dimensions and whether these dimensions are interpretable in terms of meaningful aspects of the game. The PCA results suggest that the dominant components do not form arbitrary numerical mixtures but reflect recognizable aspects of how players approach the game.

For more straightforward interpretation and understanding of the identified dimensions, we will refer to PC1–PC5 as *Risky combat* (PC1), *Wandering* (PC2), *Static high-risk play* (PC3), *Active goal-achievement* (PC4), and *Lazy style* (PC5).

First, PC1 and PC2 together describe a broad space from performance-oriented to exploratory, experience-focused play. Players with high PC1 scores and low PC2 scores correspond to experienced, goal-oriented players who engage in frequent and effective combat, efficiently gather resources, and advance rapidly toward objectives. In contrast, players with lower PC1 scores and higher PC2 scores spend more time wandering the environment and delay the use of initial energy stores, consistent with more exploratory, novice behavior.

PC3 and PC5 distinguish different aspects of risk and pace. PC3 separates mobile, low-risk roamers from more static, high-risk campers with multiple bites per kill, while PC5 highlights players who adopt an extremely slow, static pace with long periods of inactivity and very slow destruction of spawners. These components suggest that camping is not just about position, but also about the level of risk and delay players are willing to accept.

PC4 isolates a pattern of proactive flow and objective control that combines rapid destruction of spawners, broad area coverage, constant resource gathering and use, and fewer bites. This dimension differentiates players who actively shape the flow of the game and focus on objectives from those who react defensively and tend to retreat.

Together, these findings suggest that the PCA constructs a small set of latent dimensions that map to intuitive aspects of the game: combat effectiveness (with strong Achiever/Killer signals), exploratory versus focused progression (Explorer vs. Achiever), risk associated with camping (primarily Killer-related), and pace of action (combining Achiever and Killer tendencies).

Table 7 summarizes the first five principal components, their dominant high-loading features, and their informal labels used throughout the remainder of the paper.

**TABLE 7.** Behavioral trait profiles of player personas, mapped to Bartle's typology.

| PC | Label | Dominant behavioral pattern | Interpretation / Bartle link |
|---|---|---|---|
| PC1 | Risky combat | kills per minute, shots while moving, ammo collection, bites | strong Achiever/ Killer signals with elevated risk-taking |
| PC2 | Wandering | roaming, path length without shooting, first energy delivery | wandering, exploratory behavior (Explorer) |
| PC3 | Static high-risk play | camping vs. roaming index, time, and distance without shooting, bites per kill | static, high-risk campers. mainly the Killer-related risk |
| PC4 | Active goal-achievement | area coverage, portal energy per minute, resource pickups (ammo /health/ batteries) | combines Achiever (efficient progression) with some Explorer (systematic coverage) |
| PC5 | Lazy style | session duration with long idle periods, time without shooting, time to destroy spawners, | Very slow, passive play; weak Achiever/ Killer signals, closer to an experience- /tourist-like style |

Although these axes are specific to the mechanics of our game, they also provide a psychologically meaningful behavioral representation of playstyles, rather than merely a purely technical compression. These references to Achiever, Killer, and Explorer should be understood as interpretive mappings based on our a priori annotations to the features, rather than as direct measurements of player motivation.

A similar procedure, in which semantic features are first extracted from events and then interpretable behavioral dimensions are obtained using PCA, is also used by Güneş et al. [66] in their generic event-trait approach to player modeling, Swanson et al. [67] who uses a similar feature-based pipeline to identify play styles in the game Lakeland, and finally Grosu and Nicola [68], who show that a small number of derived interaction dimensions is sufficient to capture significantly different play styles in a VR environment.

### B. RQ2: GAMING PERSONAS IDENTIFICATION

RQ2 focuses on identifying distinct player play styles (personas) that are clustered in a latent behavioral space. The solution k = 5 suggests that, instead of one dominant type, three large, well-represented personas and two very small, yet structurally consistent, extremes emerge in the data. This result is consistent with the notion that FPS player behavior lies on a continuum, with dense "clusters" around a few preferred styles rather than a set of clearly distinct types [66].

The first large persona, *Efficient achievers*, represents around 20 percent of players who play many games, have above-average accuracy and headshot ratio, quickly eliminate spawners, and efficiently generate portal energy. In radar profiles, their stats are above average across most resource-based performance metrics, while their exploration metrics are moderate rather than extreme. We can describe them as experienced, purposeful players who "read" the game as a task to be mastered effectively, rather than primarily as a space for experimentation or risk-taking.

The second primary persona, *Learning wanderers*, covers about 25% of players and has a similar level of basic activity (movement, shooting) as the other clusters, but is less able to convert it into progress. These players have lower accuracy and lower kill rates, yet still end up dead or have their games interrupted. They are interesting precisely because they separate "active and successful" play from "active but ineffective": the volume of action alone is not enough to indicate a play style - the way in which this activity is structured is key. This group probably includes players who are still learning the controls, are less familiar with the genre, or use a more exploratory approach that does not yet translate into practical goal completion.

The third leading persona, *Active warriors*, accounts for almost 50 percent of players and represents the most aggressive and intense style. These players have high movement speed when shooting, high kill rates, intensive ammo and energy harvesting, but also accept higher risks: they suffer more hits (bites) and have a higher risk ratio than other clusters. From a design perspective, this looks like a "power user" combat style that rewards the system with high progress, but at the cost of increased threat, which can be key when considering difficulty balancing or adaptation. These players know precisely what style is suitable for this type of game and have applied effective tactics learned in previous games of a similar focus.

Two minor personas, *Extreme campers* and *Low-engagement walkers*, each with 4 players (around 2.5 %), are numerically small but behaviorally consistent. *Extreme campers* combine a very high camping index and long periods without shooting with substandard gameplay: the game is played mainly from static positions, often with a higher risk of being killed. *Low-engagement walkers* are at the opposite extreme: players walk a lot, often without expanding their area of coverage and without significant interaction with enemies or game objectives. From the perspective of RQ2, it is crucial that these extremes do not appear as large personas, but rather as rare, behaviorally coherent strategies that are often hidden in overall averages.

In line with RQ2, the results suggest that a small number of interpretable play styles can be extracted from the game event logs [67]: two dominant personas oriented towards successful play (*Efficient achievers* and *Active warriors*), one large transitional persona with mixed outcomes (*Learning wanderers*), and two significant, albeit rare, marginal extremes (*Extreme campers* and almost non-interactive *Low-engagement walkers*).

Importantly, these personas are not independent of the latent dimensions identified in RQ1. Each persona occupies a characteristic region in the space spanned by the five principal components:

- *Efficient achievers* combine moderately high *Risky combat* (PC1) with low *Wandering* (PC2) and elevated *Active goal-achievement* (PC4).
- *Active warriors* are placed at the extreme high end of *Risky combat* (PC1), again with low *Wandering* (PC2) and high *Active goal-achievement* (PC4), forming the most combat-intensive variant of goal-focused play.
- *Learning wanderers* lie between these two personas, with intermediate *Risky combat* (PC1), higher *Wandering* (PC2), and weaker *Active goal-achievement* (PC4), reflecting active but less efficiently structured play.

Two edge personas map onto the remaining components:

- *Extreme campers* show very high *Static high-risk play* (PC3) and often elevated *Lazy style* (PC5),
- *Low-engagement walkers* exhibit high *Wandering* (PC2), low *Risky combat* (PC1), and low *Active goal-achievement* (PC4).

Combined with the generally low silhouette scores, this pattern indicates that the personas should be understood as macro-structures in a continuous behavioral space rather than as sharp categories into which every player can be reliably assigned, an essential consideration for the subsequent analysis of their stability and their relation to theoretical typologies.

### C. RQ3: CLUSTER STABILITY AND CORE PERSONAS

RQ3 focuses on identifying the stability of the identified persona structure. Across the sample, bootstrap stability was only moderate: for both $k = 5$ and $k = 6$, the average ARI across 300 resamplings was around 0.5 with very similar medians. Combined with the low average silhouette bootstrap stability scores, it suggests that slight variations in the data are sufficient to cause a nontrivial subset of players to be assigned to different personas and that the boundaries between personas are relatively blurred within the population.

Silhouette-based pruning analysis shows that this instability is not uniform. When we retain only the top 70% of members per cluster according to their silhouette values, the resulting core exhibits very high bootstrap stability. For $k = 5$, retaining the top 70% results in a core of 110 players with an average silhouette of 0.235 and a median ARI of around 0.92 (mean ARI = 0.89). Increasing the threshold to 0.80 or 0.90 increases the core size but systematically reduces the silhouette and ARI, indicating that additional members are increasingly ambiguous. The solution with $k = 6$ shows the same monotonic pattern: its core is most stable around $q = 0.70$ and becomes less stable as lower-silhouette players are added.

Thus, most of the instability in the full sample is concentrated in the low-silhouette tails of each cluster. Key to RQ3 is that the persona structure itself changes very little when we move from the full sample to the high-silhouette core. When we recalculate the k-means on the core with $q = 0.70$ and compare the resulting centroids with those of the full sample, the profiles of the five persons remain remarkably similar (Fig. 4 and Fig. 6). The differences in individual characteristics are small, the ordering of personas along the main behavioral dimensions is preserved, and the radar plots for the full vs. basic solutions are very similar. Pruning, therefore, does not reveal new structure; it sharpens the existing one by removing ambiguous edge cases.

Conceptually, each persona has a robust core of prototypical players whose assignments are highly reproducible across resampling, and a fuzzy periphery of borderline players who tend to move between neighboring personas. Small borderline groups also form coherent modes in the feature space, but their small sample sizes limit the accuracy with which their stability can be estimated.

From a player modeling perspective, RQ3 supports a stability-oriented interpretation of personas. The basic structure of the persona is stable: the same five behavioral profiles reappear when we focus on the high-silhouette core. The emphasis on persona stability and repeatability is also consistent with infrastructure designs such as Open Game Data, which explicitly support regular, repeatable extraction of game metrics for secondary analysis [69].

### D. RELATIONSHIP TO BARTLE PLAYER TYPES

RQ4 aims to identify relationships between the derived personas and Bartle's motivational categories and to replicate these patterns across the types of Achiever, Killer, and Explorer. In our analysis, this link was implemented in two steps: first, individual traits were defined and annotated a priori as indicators of tendencies similar to Achiever, Killer, or Explorer (or combinations thereof); second, we examined how these annotated traits loaded into the principal components and shaped the behavioral profiles of the five personas.

At the latent dimension level, the correspondence with Bartle's categories is intuitive but not unambiguous.

- PC1 (*Risky combat*) combines high kill rates, shooting while moving, collecting ammo, and exposure to bites, and therefore carries strong mixed signals of *Achiever* and *Killer*: players are both effective in combat and willing to take risks.
- PC2 (*Wandering*) separates focused progression from exploratory wandering and delayed progress towards objectives, which is closely aligned with the *Achiever* and *Explorer* distinction.
- PC3 (*Static high-risk play*) captures the risk associated with camping and can be interpreted as a maladaptive, *Killer*-oriented style where players let enemies get dangerously close.
- PC4 (*Active goal-achievement*) reflects effective control of the map and objectives and is essentially an *Achiever*-like style with some systematic exploration.
- PC5 (*Lazy style*) represents very slow, low-involvement play that falls outside the classic *Achiever* and *Killer* patterns and is better described as an experiential or tourist mode.

Overall, the PCA dimensions resonate with Bartle's axes, but are based on concrete mechanics (movement, shooting, resource use) rather than abstract preference statements.

The personas derived in RQ2 can be read as characteristic combinations of these dimensions, and thus also partially correspond to Bartle's framework.

- *Efficient achievers* exhibit moderate combat risk, low Wandering, and strong active goal pursuit, corresponding mainly to the *Achiever* style with only moderate *Killer* involvement.
- *Active warriors* take this pattern to the extreme, with very high combat risk and still low *Wandering*, creating a hybrid of the *Achiever* and *Killer* types that aggressively optimizes the course of combat.
- *Learning wanderers* occupy a middle ground with moderate combat risk, increased *Wandering*, and weaker active goal pursuit; they can be understood as a mixed or developing profile, where exploratory or hesitant behavior does not yet translate into effective *Achiever*-style progression.
- *Extreme campers* score very high in static, high-risk play and often in an extremely passive style, combining the *Killer's* willingness to accept danger with unusually static, slow tactics.
- *Low-engagement walkers* exhibit high levels of wandering and low risk of combat and goal attainment,

resembling an *Explorer* or hiker style with minimal signals of the *Achiever* or *Killer* type.

RQ4 highlights significant limitations of directly mapping from game records to classical taxonomies of player types. Our data comes from a single-player, combat-focused scenario without social interaction; the real *Socializer* type cannot be directly observed, and any ''experience-oriented'' patterns (such as a lazy style or low-engagement walkers) are only rough analogies. Traits and components systematically intersect multiple motives: in this game, being an effective *Achiever* almost necessarily involves combat, so indicators of *Achiever* and *Killer* are linked at both the trait and component levels.

The mapping from behavior to motive is inherently interpretive: we observe what players do, not why they do it. The A/K/E labels, therefore, provide helpful insight into interpretation but should not be confused with direct measures of underlying motivational types.

The findings for RQ4 suggest that the proposed modeling pipeline can create behavioral dimensions and personas that are compatible with, but not reducible to, Bartle's style categories. PCA reveals axes that reflect differences between *Achiever*, *Killer*, and *Explorer*, while adding more detailed constructs such as camping risk and game pace. Personas then combine these axes into a small set of meaningful playstyles that cover most of the observed behavior.

From a theoretical perspective, this supports the use of Bartle's framework as a conceptual scaffold rather than a rigid ground truth: it helps to label and communicate emerging patterns, but the actual structure of player behavior in a particular game is better captured by data-driven dimensions and personas tested for stability than by assuming a small fixed set of pure types.

Across all four research questions, the findings point in the same direction: player behavior in this FPS scenario is best understood as a continuous space structured by a small number of latent dimensions and a few stable core personas, rather than a fixed set of explicit types. Protocol-derived features inspired by player type frameworks can be compressed into interpretable components that capture combat intensity, goal effectiveness, wandering, camping risk, and overall game pace.

Clustering within this space yields a few robust personas whose cores are highly stable across resampling, while ambiguity is concentrated at their peripheries. These personas are compatible with the Achiever-Killer-Explorer distinction, but they also refine it and reveal mixed or transitional patterns that classical taxonomies largely ignore.

Methodologically, the findings support a stability- and data-driven approach to player modeling: instead of assuming a predefined typology, researchers and designers can create game-specific features, reduce redundancy, derive latent dimensions, and then identify core personas that are both behaviorally interpretable and empirically stable.

## VI. CONCLUSION

This paper presents and evaluates a protocol-based player model that focuses on stability, using Another Useless Zombie Game as a case study. Starting from raw game event logs, we derive behavioral characteristics inspired by player type frameworks, reduce redundancy using correlation filtering and PCA, and apply k-means clustering in latent space combined with bootstrap resampling and silhouette-based pruning. The resulting model identifies a small set of interpretable personas with robust, highly silhouetted cores, demonstrating that even relatively simple protocols can support nuanced, empirically stable representations of play styles.

Our goal was not to propose a new universal taxonomy of player types or to replace existing models such as Bartle or HEXAD. Instead, we argued for a practical, replicable approach that builds on the specific mechanisms of a particular game and creates data-driven personas that are both behaviorally interpretable and statistically validated. The latent dimensions and personas we derive are compatible with the Achiever-Killer-Explorer distinction, but also refine it by revealing mixed, transitional, and extreme patterns that classical typologies poorly capture. We see this as a step toward bridging theory-based approaches and log-derived methods: designers and researchers can use established frameworks as conceptual scaffolding while relying on transparent and robust procedures to derive game-specific models of how players actually behave [70].

Although demonstrated here in an FPS game scenario, the same approach is directly applicable to serious games and cyber-outreach-style simulations, where modeling how different personas explore, ignore, or actively exploit system mechanics can support the design of more effective cybersecurity training and assessment.

### A. LIMITATIONS

This study has several limitations that limit the generalizability of its findings.

All analyses are based on a single game and a single core mechanic (a short, combat-oriented FPS scenario with fixed objectives). The latent dimensions and personas we observe are therefore tailored to this design and may not be directly transferable to other genres or interaction patterns.

We did not use explicit motivational or immersion questionnaires in this article, so the inferred associations between behavioral personas and game-type frameworks remain interpretive rather than validated through self-report.

The game event logs do not include any data on social interaction: cooperative, competitive, or communicative aspects of the game are absent, making it impossible to identify real Socializer-like patterns or study social dynamics.

### B. FUTURE RESEARCH

These limitations point to several avenues for future research. A natural next step is to apply the same modeling approach to other games and genres, varying the mechanics, pacing, and social context to test which dimensions and personas generalize and which are game-specific. Another direction is to combine questionnaire-based personas or related psychometric profiles, allowing for calibration between "why and how deeply players want to play" and "how they actually behave in the game."

Ultimately, a stability-focused approach to personas opens the door to online adaptive systems that estimate a player's position in latent space over time, track their movement between core personas, and use this information to drive dynamic adaptation of difficulty, content selection, or personalized feedback.

In addition to entertainment analytics, a similar modeling approach focused on stability could also be applied to cybersecurity training scenarios and serious games, where understanding stable versus ambiguous behavioral patterns (e.g., cautious versus risky task strategies) is crucial for designing effective interventions and developing competencies.

## DECLARATION OF AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work, the author(s) used Grammarly and ChatGPT 5 to edit and phrase sections for better readability. After using this tool/service, the authors reviewed and edited the content as needed, and they take full responsibility for the content of the publication.

## REFERENCES

[1] X. Fu, X. Chen, Y.-T. Shi, I. Bose, and S. Cai, "User segmentation for retention management in online social games," *Decis. Support Syst.*, vol. 101, pp. 51–68, Sep. 2017, doi: 10.1016/j.dss.2017.05.015.

[2] J. Tan and M. Katchabaw, "A reusable methodology for player clustering using Wasserstein autoencoders," in *Proc. Int. Conf. Entertainment Comput.*, in Lecture Notes in Computer Science, vol. 13477, 2022, pp. 296–308, doi: 10.1007/978-3-031-20212-4_24.

[3] S. C. J. Bakkes, P. H. M. Spronck, and G. van Lankveld, "Player behavioural modelling for video games," *Entertainment Comput.*, vol. 3, no. 3, pp. 71–79, Aug. 2012, doi: 10.1016/j.entcom.2011.12.001.

[4] A. Alvarez, J. Font, and J. Togelius, "Toward designer modeling through design style clustering," *IEEE Trans. Games*, vol. 14, no. 4, pp. 676–686, Dec. 2022, doi: 10.1109/TG.2022.3143800.

[5] P. D. Paraschos and D. E. Koulouriotis, "Game difficulty adaptation and experience personalization: A literature review," *Int. J. Hum.-Comput. Interact.*, vol. 39, no. 1, pp. 1–22, Jan. 2023, doi: 10.1080/10447318.2021.2020008.

[6] F. Mortazavi, H. Moradi, and A.-H. Vahabie, "Dynamic difficulty adjustment approaches in video games: A systematic literature review," *Multimedia Tools Appl.*, vol. 83, no. 35, pp. 83227–83274, Mar. 2024, doi: 10.1007/s11042-024-18768-x.

[7] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit muds," *J. MUD Res.*, vol. 1, no. 1, p. 19, 1996. [Online]. Available: https://www.hayseed.net/MOO/JOVE/bartle.html%0A,https://www.researchgate.net/profile/Richard_Bartle/publication/247190693_Hearts_clubs_diamonds_spades_Players_who_suit_MUDs/links/540058700cf2194bc29ac4f2.pdf

[8] G. F. Tondello, R. R. Wehbe, L. Diamond, M. Busch, A. Marczewski, and L. E. Nacke, "The gamification user types hexad scale," in *Proc. Annu. Symp. Comput.-Human Interact. Play*, Oct. 2016, pp. 229–243, doi: 10.1145/2967934.2968082.

[9] G. F. Tondello, A. Mora, A. Marczewski, and L. E. Nacke, "Empirical validation of the gamification user types hexad scale in English and Spanish," *Int. J. Hum.-Comput. Stud.*, vol. 127, pp. 95–111, Jul. 2019, doi: 10.1016/j.ijhcs.2018.10.002.

[10] L. F. Bicalho, A. Baffa, and B. Feijo, "A game analytics model to identify player profiles in singleplayer games," in *Proc. 18th Brazilian Symp. Comput. Games Digit. Entertainment (SBGames)*, Brazil, Oct. 2019, pp. 11–20, doi: 10.1109/sbgames.2019.00013.

[11] J. Kirchner-Krath, M. Altmeyer, L. Schürmann, B. Kordyaka, B. Morschheuser, A. C. T. Klock, L. Nacke, J. Hamari, and H. F. O. von Korflesch, "Uncovering the theoretical basis of user types: An empirical analysis and critical discussion of user typologies in research on tailored gameful design," *Int. J. Hum.-Comput. Stud.*, vol. 190, Oct. 2024, Art. no. 103314, doi: 10.1016/j.ijhcs.2024.103314.

[12] S. Melo, L. Thurler, A. Paes, and E. Clua, "Game provenance graph-based representation learning vs metrics-based machine learning: An empirical comparison on predictive game analytics tasks," *Entertainment Comput.*, vol. 52, Jan. 2025, Art. no. 100755, doi: 10.1016/j.entcom.2024.100755.

[13] R. Sifa, A. Drachen, and C. Bauckhage, "Profiling in games: Understanding behavior from telemetry," in *Social Interactions in Virtual Worlds: An Interdisciplinary Perspective*, K. Lakkaraju, G. Sukthankar, and R. T. Wigand, Eds., Cambridge, U.K.: Cambridge Univ. Press, 2018, pp. 337–374.

[14] Y. Su, P. Backlund, and H. Engström, "Comprehensive review and classification of game analytics," *Service Oriented Comput. Appl.*, vol. 15, no. 2, pp. 141–156, Jun. 2021, doi: 10.1007/s11761-020-00303-z.

[15] M. Ferguson, S. Devlin, D. Kudenko, and J. A. Walker, "Player style clustering without game variables," in *Proc. Int. Conf. Found. Digit. Games*, Sep. 2020, pp. 1–4, doi: 10.1145/3402942.3402960.

[16] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.

[17] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, and R. Ranjan, "Explainable AI (XAI): Core ideas, techniques, and solutions," *ACM Comput. Surv.*, vol. 55, no. 9, pp. 1–33, Sep. 2023, doi: 10.1145/3561048.

[18] A. Rai, "Explainable AI: From black box to glass box," *J. Acad. Marketing Sci.*, vol. 48, no. 1, pp. 137–141, Jan. 2020, doi: 10.1007/s11747-019-00710-5.

[19] M. S. El-Nasr, T. H. N. Dinh, A. Canossa, and A. Drachen, "Clustering methods in game data science," in *Game Data Science*,vol. 18. Oxford, U.K.: Academic, Nov. 2021.

[20] N. Ballou, A. Denisova, R. Ryan, C. S. Rigby, and S. Deterding, "The basic needs in games scale (BANGS): A new tool for investigating positive and negative video game experiences," *Int. J. Hum.-Comput. Stud.*, vol. 188, Aug. 2024, Art. no. 103289, doi: 10.1016/j.ijhcs.2024.103289.

[21] S. Bunian, A. Canossa, R. Colvin, and M. S. El-Nasr, "Modeling individual differences in game behavior using HMM," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2017, pp. 158–164, doi: 10.1609/aiide.v13i1.12942.

[22] L. E. Nacke, C. Bateman, and R. L. Mandryk, "BrainHex: Preliminary results from a neurobiological gamer typology survey," in *Proc. Int. Conf. Entertainment Comput.*, vol. 6972, 2011, pp. 288–293, doi: 10.1007/978-3-642-24500-8_31.

[23] N. Yee, "Motivations for play in online games," *CyberPsychology Behav.*, vol. 9, no. 6, pp. 772–775, Dec. 2006, doi: 10.1089/cpb.2006.9.772.

[24] L. E. Nacke, C. Bateman, and R. L. Mandryk, "BrainHex: A neurobiological gamer typology survey," *Entertainment Comput.*, vol. 5, no. 1, pp. 55–62, Jan. 2014, doi: 10.1016/j.entcom.2013.06.002.

[25] J. C. Coffee Jr., L. Lowenstein, and S. Rose-Ackerman, *Knights, Raiders, and Targets: The Impact of the Hostile Takeover*. Oxford Univ. Press, 1988, p. 274. [Online]. Available: https://scholarship.law.columbia.edu/books/274

[26] P. Montreal, M. Passalacqua, S. Sénécal, H. Montreal, M. Frédette, H. Montreal, L. E. Nacke, R. Pellerin, P. Montreal, P.-M. Léger, and H. Montreal, "Should gamification be personalized? A self- deterministic approach," *AIS Trans. Hum.-Comput. Interact.*, pp. 265–286, Sep. 2021, doi: 10.17705/1thci.00150.

[27] C. Bateman, R. Lowenhaupt, and L. Nacke, "Player typology in theory and practice," in *Proc. DiGRA Digit. Library*, Jan. 2011, pp. 1–24, doi: 10.26503/dl.v2011i1.562.

[28] M. Busch, E. Mattheiss, R. Orji, P. Fröhlich, M. Lankes, and M. Tscheligi, "Player type models: Towards empirical validation," in *Proc. CHI Conf. Extended Abstr. Human Factors Comput. Syst.*, May 2016, pp. 1835–1841, doi: 10.1145/2851581.2892399.

[29] J. Tuunanen and J. Hamari, "Meta-synthesis of player typologies," in *Proc. DiGRA Nordic*. Tampere, Finland: Univ. of Tampere, 2012, pp. 29–53, doi: 10.26503/dl.v2012i1.604.

[30] C. S. González-González, P. A. Toledo-Delgado, V. Muñoz-Cruz, and J. Arnedo-Moreno, "Gender and age differences in preferences on game elements and platforms," *Sensors*, vol. 22, no. 9, p. 3567, May 2022, doi: 10.3390/s22093567.

[31] A. Drachen, C. Thurau, R. Sifa, and C. Bauckhage, "A comparison of methods for player clustering via behavioral telemetry," 2014, *arXiv:1407.3950*.

[32] S. Colombo, P. Hansson, and M. B. T. Nyström, "Mining players' experience in computer games: Immersion affects flow but not presence," *Comput. Hum. Behav. Rep.*, vol. 12, Jan. 2023, Art. no. 100334, doi: 10.1016/j.chbr.2023.100334.

[33] S. F. Erümit, L. Şılbır, A. K. Erümit, and H. Karal, "Determination of player types according to digital game playing preferences: Scale development and validation study," *Int. J. Hum.-Comput. Interact.*, vol. 37, no. 11, pp. 991–1002, Jul. 2021, doi: 10.1080/10447318.2020.1861765.

[34] J. Konert, M. Gutjahr, S. Göbel, and R. Steinmetz, "Modeling the player: Predictability of the models of bartle and kolb based on NEO-FFI (big5) and the implications for game based learning," *Int. J. Game-Based Learn.*, vol. 4, no. 2, pp. 36–50, 2014, doi: 10.4018/ijgbl.2014040103.

[35] L. Żuchowska, K. Kutt, and G. J. Nalepa, "Bartle taxonomy-based game for affective and personality computing research," in *Proc. CEUR Workshop*, vol. 2995, 2021, pp. 51–55.

[36] C. Bauckhage, A. Drachen, and R. Sifa, "Clustering game behavior data," *IEEE Trans. Comput. Intell. AI Games*, vol. 7, no. 3, pp. 266–278, Sep. 2015, doi: 10.1109/TCIAIG.2014.2376982.

[37] A. Perišić and M. Pahor, "Clustering mixed-type player behavior data for churn prediction in mobile games," *Central Eur. J. Operations Res.*, vol. 31, no. 1, pp. 165–190, Mar. 2023, doi: 10.1007/s10100-022-00802-8.

[38] K. Mustač, K. Bačić, L. Skorin-Kapov, and M. Sužnjević, "Predicting player churn of a free-to-play mobile video game using supervised machine learning," *Appl. Sci.*, vol. 12, no. 6, p. 2795, Mar. 2022, doi: 10.3390/app12062795.

[39] D. Melhart, D. Gravina, and G. N. Yannakakis, "Moment-to-moment engagement prediction through the eyes of the observer: PUBG streaming on twitch," in *Proc. Int. Conf. Found. Digit. Games*, Sep. 2020, pp. 1–10, doi: 10.1145/3402942.3402958.

[40] A. Normoyle and S. T. Jensen, "Bayesian clustering of player styles for multiplayer games," in *Proc. 11th AAAI Conf. Artif. Intell.*, 2015, pp. 163–169, doi: 10.1609/aiide.v11i1.12805.

[41] A. Saas, A. Guitart, and A. Perianez, "Discovering playing patterns: Time series clustering of free-to-play game data," in *Proc. IEEE Conf. Comput. Intell. Games (CIG)*, Sep. 2016, pp. 1–8, doi: 10.1109/CIG.2016.7860442.

[42] P. García-Sánchez, A. Fernández-Ares, A. P. Tonda, and A. M. Mora, "Data mining of deck archetypes in hearthstone," in *Proc. CEUR Workshop*, vol. 2719, 2020, pp. 132–144.

[43] A. Drachen, M. S. El-Nasr, and A. Canossa, "Game analytics—The basics," in *Game Analytics: Maximizing the Value of Player Data*, M. Seif El-Nasr, A. Drachen, and A. Canossa, Eds., London, U.K.: Springer, 2013, pp. 13–40. [Online]. Available: https://doi.org/10.1007/978-1-4471-4769-5_2

[44] S. Sipone, V. Abella, M. Rojo, and J. L. Moura, "Gamification and player type: Relationships of the HEXAD model with the learning experience," *J. New Approaches Educ. Res.*, vol. 14, no. 1, p. 1, Jan. 2025, doi: 10.1007/s44322-024-00021-w.

[45] J. Close, S. G. Spicer, L. L. Nicklin, M. Uther, J. Lloyd, and H. Lloyd, "Secondary analysis of loot box data: Are high-spending 'whales' wealthy gamers or problem gamblers?" *Addictive Behaviors*, vol. 117, Jun. 2021, Art. no. 106851, doi: 10.1016/j.addbeh.2021.106851.

[46] M. Milošević, N. Živić, and I. Andjelković, "Early churn prediction with personalized targeting in mobile social games," *Expert Syst. Appl.*, vol. 83, pp. 326–332, Oct. 2017, doi: 10.1016/j.eswa.2017.04.056.

[47] R. Habibi, J. Pfau, and M. S. El-Nasr, "Modeling player personality factors from in-game behavior and affective expression," in *Proc. 11th Int. Conf. Affect. Comput. Intell. Interact. Workshops Demos (ACIIW)*, Sep. 2023, pp. 1–8, doi: 10.1109/ACIIW59127.2023.10388138.

[48] A. Tlili, M. Chang, J. Moon, Z. Liu, D. Burgos, N. Shing Chen, and Kinshuk, "A systematic literature review of empirical studies on learning analytics in educational games," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 7, no. 2, pp. 250–261, Dec. 2021, doi: 10.9781/ijimai.2021.03.003.

[49] S. K. Banihashem, H. Dehghanzadeh, D. Clark, O. Noroozi, and H. J. A. Biemans, "Learning analytics for online game-based learning: A systematic literature review," *Behav. Inf. Technol.*, vol. 43, no. 12, pp. 2689–2716, Sep. 2024, doi: 10.1080/0144929x.2023.2255301.

[50] K. K. Yu, M. Guzdial, and N. R. Sturtevant, "The FarmQuest player telemetry dataset: Playthrough data of a Cozy farming game," in *Proc. AAAI Conf. Artif. Intell.*, 2024, vol. 20, no. 1, pp. 263–268, doi: 10.1609/aiide.v20i1.31889.

[51] M. Busch, E. Mattheiss, W. Hochleitner, C. Hochleitner, M. Lankes, P. Fröhlich, R. Orji, and M. Tscheligi, "Using player type models for personalized game design—An empirical investigation," *Interact. Design Archit.*, no. 28, pp. 145–163, Mar. 2016, doi: 10.55612/s-5002-028-008.

[52] F. Cheng, D. Liu, F. Du, Y. Lin, A. Zytek, H. Li, H. Qu, and K. Veeramachaneni, "VBridge: Connecting the dots between features and data to explain healthcare models," *IEEE Trans. Vis. Comput. Graph.*, vol. 28, no. 1, pp. 378–388, Jan. 2022, doi: 10.1109/TVCG.2021.3114836.

[53] X. Wang, L. Xie, H. Wang, X. Xing, W. Wan, Z. Wu, X. Ma, and Q. Li, "Deciphering explicit and implicit features for reliable, interpretable, and actionable user churn prediction in online video games," *IEEE Trans. Vis. Comput. Graphics*, vol. 31, no. 9, pp. 5990–6007, Sep. 2025, doi: 10.1109/TVCG.2024.3487974.

[54] G. N. Yannakakis and J. Togelius, *Artificial Intelligence and Games*. Cham, Switzerland: Springer, 2018.

[55] S. Agrawal, S. Bech, K. Bærentsen, K. De Moor, and S. Forchhammer, "Method for subjective assessment of immersion in audiovisual experiences," *J. Audio Eng. Soc.*, vol. 69, no. 9, pp. 656–671, Sep. 2021, doi: 10.17743/jaes.2021.0013.

[56] A. Normoyle and S. T. Jensen, "Bayesian learning of play styles in multiplayer video games," 2021, *arXiv:2112.07437*.

[57] P. P. Tricomi, F. Nenna, L. Pajola, M. Conti, and L. Gamberini, "You can't hide behind your headset: User profiling in augmented and virtual reality," *IEEE Access*, vol. 11, pp. 9859–9875, 2023, doi: 10.1109/ACCESS.2023.3240071.

[58] C. Alonso-Fernández, A. Calvo-Morata, M. Freire, I. Martínez-Ortiz, and B. Fernández-Manjón, "Game learning analytics: Blending visual and data mining techniques to improve serious games and to better understand player learning," *J. Learn. Anal.*, vol. 9, no. 3, pp. 32–49, Dec. 2022, doi: 10.18608/jla.2022.7633.

[59] M. Aydin, H. Karal, and V. Nabiyev, "Examination of adaptation components in serious games: A systematic review study," *Educ. Inf. Technol.*, vol. 28, no. 6, pp. 6541–6562, Jun. 2023, doi: 10.1007/s10639-022-11462-1.

[60] A. Rashed, S. Shirmohammadi, I. Amer, and M. Hefeeda, "A review of player engagement estimation in video games: Challenges and opportunities," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 21, no. 7, pp. 1–33, Jul. 2025, doi: 10.1145/3722116.

[61] D. Chiotaki, V. Poulopoulos, and K. Karpouzis, "Adaptive game-based learning in education: A systematic review," *Frontiers Comput. Sci.*, vol. 5, pp. 1–8, Jun. 2023, doi: 10.3389/fcomp.2023.1062350.

[62] C. Y. Ng and M. K. B. Hasan, "Cybersecurity serious games development: A systematic review," *Comput. Secur.*, vol. 150, Mar. 2025, Art. no. 104307, doi: 10.1016/j.cose.2024.104307.

[63] N. Donekal Chandrashekar, A. Lee, M. Azab, and D. Gracanin, "Understanding user behavior for enhancing cybersecurity training with immersive gamified platforms," *Information*, vol. 15, no. 12, p. 814, Dec. 2024, doi: 10.3390/info15120814.

[64] J. Prümmer, T. van Steen, and B. van den Berg, "A systematic review of current cybersecurity training methods," *Comput. Secur.*, vol. 136, Jan. 2024, Art. no. 103585, doi: 10.1016/j.cose.2023.103585.

[65] F.-ul-Ain, D. Akhmetzyanova, I. Matias, and V. Tomberg, "Behaviour models-enriched user personas for digital behaviour change interventions," in *Proc. 17th Int. Conf. Pervasive Technol. Related Assistive Environments*, 2024, pp. 140–146, doi: 10.1145/3652037.3652069.

[66] M. A. Gunes, M. F. Kavum, and S. Sariel, "A generic approach for player modeling using event-trait mapping supported by PCA," in *Advances in Computer Games (ACG 2021)* (Lecture Notes in Computer Science), vol. 13262, C. Browne, A. Kishimoto, and J. Schaeffer, Eds., Cham, Switzerland: Springer, 2022, pp. 197–207, doi: 10.1007/978-3-031-11488-5_18.

[67] L. Swanson, D. Gagnon, J. Scianna, J. McCloskey, N. Spevacek, S. Slater, and E. Harpstead, "Leveraging cluster analysis to understand educational game player experiences and support design," 2022, *arXiv:2210.09911*.

[68] M.-A. Grosu and S. Nicola, "Analyzing player behavior in a VR game for children using gameplay telemetry," *Multimodal Technol. Interact.*, vol. 9, no. 9, p. 96, Sep. 2025, doi: 10.3390/mti9090096.

[69] L. Swanson and D. J. Gagnon, "An architecture for repeatable, large-scale educational game data analysis: Building on open game data BT–serious games," in *Serious Games* (Lecture Notes in Computer Science), vol. 15259, L. Plass and X. Ochoa, Eds., Cham, Switzerland: Springer, 2025. [Online]. Available: https://doi.org/10.1007/978-3-031-74138-8_4

[70] J. Salminen, J. Vahlo, A. Koponen, S. G. Jung, S. A. Chowdhury, and B. J. Jansen, "Designing prototype player personas from a game preference survey," in *Proc. Conf. Human Factors Comput. Syst.*, 2020, pp. 1–8, doi: 10.1145/3334480.3382785.


**JÁN SKALKA** received the master's degree in mathematics and informatics education from the Department of Informatics, Constantine the Philosopher University in Nitra, Slovakia.

He defended the Ph.D. thesis in the research study program of technology of education from the Constantine the Philosopher University in Nitra, in 2004. Since 2000, he has been an assistant Professor and an Associate Professor at the Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra. He is currently a member of the Multimodal User Analytics and Cognitive Behavior Laboratory, where he explores the intersection of artificial intelligence, human behavior, and multimodal data analysis. His work focuses on AI applications, information systems, and user behavior analytics, particularly in virtual and interactive environments, such as VR applications and intelligent systems. His research interests span human–computer interaction, cognitive modeling, AI-driven decision-making, and problem-solving in virtual environments.

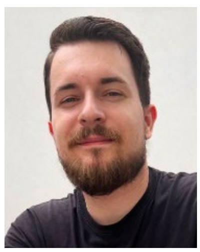
**MARTIN DRLÍK** received the master's degree in biophysics from the Faculty of Mathematics, Physics, and Computer Science, Comenius University in Bratislava, Slovakia, in 2001, and the Ph.D. degree from the Research Study Program (Theory of computer science education), Constantine the Philosopher University in Nitra, Slovakia, in 2009. Since 2019, he has been an Associate Professor at the Department of Informatics, Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra. He is actively involved in the Knowledge Discovery Research Group's projects focused on machine learning, learning analytics, and software and data engineering. He is the author and reviewer of many scientific conferences and journals. His research interests include learning analytics, educational data mining, data and software engineering, and database systems.

**DOMINIK HALVONÍK** received the master's degree in applied informatics from the Department of Informatics, Constantine the Philosopher University in Nitra, Slovakia.

He defended his Ph.D. thesis in the research study program in informatics at the Constantine the Philosopher University in Nitra, in 2020. Since 2020, he has been a Scientific Researcher and an Assistant Professor at the Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra. He is currently a member of the Natural Language Processing Laboratory, where he explores the intersection of artificial intelligence, human language, and dialog systems. His work focuses on AI applications, information systems, and natural language processing, particularly in games and also in interactive environments, such as standard web and mobile applications.

**PETER KUNA** received the master's degree in teaching informatics and technology from the Department of Informatics at Constantine the Philosopher University in Nitra, Slovakia, and the Engineering degree in automation of manufacturing systems from the Faculty of Engineering, Slovak University of Agriculture in Nitra.

He defended his Ph.D. thesis in the didactics of technical vocational subjects with Constantine the Philosopher University in Nitra, in 2013. Since 2013, he has been an Assistant Professor, currently with the Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra. His work focuses on the application of AI platforms and virtual reality systems in industrial practice. He has initiated and managed multiple projects to integrate VR and AI technologies directly into industrial environments, stemming from his long-term efforts to bridge academic research with real-world industrial applications.



**MATÚŠ VALKO** received the bachelor's degree in applied informatics from the Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, and the master's degree in applied informatics from the Faculty of Natural Sciences and Informatics, Constantine the Philosopher University in Nitra, Slovakia.

Since 2023, he has been a full-time Doctoral Student with the Applied Informatics Study Program, Department of Informatics, Faculty of Natural Sciences and Informatics, UKF in Nitra. His research focuses on identifying user behavior patterns in virtual environments, mainly in educational and virtual reality environments, including e-learning, learning analytics, and gamification.