# "Hey Players, there is a problem...": On Attribute Inference Attacks against Videogamers

Linus Eisele, Giovanni Apruzzese
Liechtenstein Business School, University of Liechtenstein
{name.surname}@uni.li

***Abstract*—We focus on a subtle privacy issue that affects (potentially hundreds of) millions of videogamers: attribute inference attacks (AIA). Through AIA, evildoers can infer gamers' *private* attributes (e.g., age, gender, occupation) by leveraging in-game statistics that are *publicly available*. Despite some previous research efforts highlighting the practicality of AIA in DOTA2, the overarching gaming community is not yet aware of this threat. We argue that AIA can only be mitigated through the collaboration of the entire videogaming community, and hence all stakeholders should be cognizant of the potential threat of AIA.**

**In this work, we first assess the risk of AIA in a broad range of online video games through a set of (original) criteria that make a game prone to AIA. We further examine some practical ways in which attackers can collect personal user data in order to subsequently correlate it with their publicly available in-game data. Finally, we confirm in a representative user study (n=460) that the gamers are hardly aware of subtle issues related to AIA. In particular, 24% of our participants revealed that they would publicly share their personal data. Clearly, such data can be leveraged by evildoers to launch AIA *against other players*.**

## I. INTRODUCTION

Video games represent the world's leading entertainment industry [1], totaling over $250 billion in 2023 [2]. These numbers are driven by the immense popularity of videogames across all age groups, genders, cultures, and income levels [1]: today, over 40% of the World's population play videogames [3]. Online multi-player videogames, in particular, are preferred [4] to single-player videogames—predominantly due to their intrinsic trait of enabling social interactivity [5–7].

Unfortunately, players of such videogames are exposed to various privacy threats [1]. The gaming ecosystem generates large amounts of data which may "leak" information about the players themselves. For instance, in 2014 [8], it was argued that—by using data accessible only to game-developers—it could be possible to infer certain personal attributes of a given player (e.g., their gender, or age). Such a possibility was confirmed by subsequent studies (e.g., [9]), showing correlations between players' *(a)* in-game activities and their *(b)* off-game personal attributes—which could be used by developers to improve their products [10]. By themselves, such "profiling" activities are not necessarily malicious. Yet, in 2023, a new study demonstrated that such correlations between in- and off-game data can be maliciously exploited to launch a subtle form of privacy violation: "attribute inference attacks," or AIA [7].

The fundamental (and dangerous) aspect of AIA is that they can be carried out by leveraging *publicly available data*. Indeed, in AIA (see Fig. 1), the attacker relies on in-game statistics of players retrieved by so-called "tracking websites" (e.g., stratz.com), which are openly accessible. Such statistics (e.g., win/loss ratio, favourite weapon/hero) can be used to infer a given player's personal attributes—which are *private information*. Through a user study of 484 DOTA2 players, Tricomi et al. [7] demonstrated that it is possible to, e.g., identify underage players in a population with ≈100% precision.

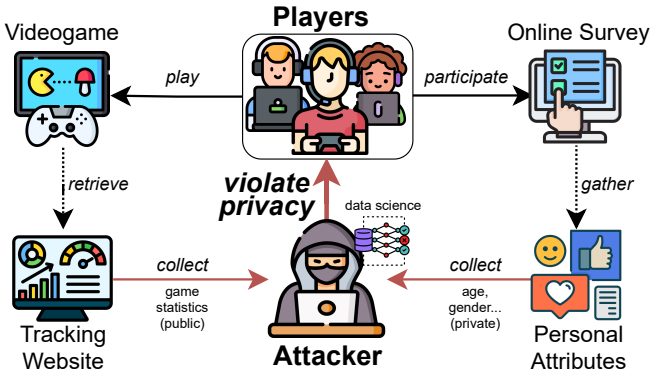


Fig. 1: **Attribute Inference Attacks in Videogames** – The in-game statistics of players are retrieved by "tracking websites", and are *publicly* accessible. An attacker can collect such data, and then collect *private* personal attributes (e.g., age, gender, occupation...) of players by, e.g., distributing "online surveys". Finally, by exploiting data science techniques, it is possible to infer personal information about *other* players—violating their privacy.

According to Tricomi et al. [7], countermeasures against AIA are highly nontrivial and require a joint effort between *players* and *developers*: the former should be more "mindful of privacy," whereas the latter should provide more "privacy options" that the players can use to protect themselves. In this vision paper, we reinforce the message of Tricomi et al. [7] by showing the applicability of AIA across a range of popular multiplayer titles. Hence, we argue that AIA are a threat that should not be underestimated in the gaming landscape.

**CONTRIBUTIONS.** We seek to raise awareness of AIA in videogames. After summarizing the quintessential properties of AIA and providing factual evidence that such a privacy threat has been overlooked in game-related research (§II), we:

- systematically analyse the gaming landscape and identify 11 titles which bear high risk of being targeted by AIA(§III).
- we examine practical ways in which an attacker can collect data to setup an AIA, our particular attention being on surveys distributed in gaming communities (§IV).
- through an ethical user survey (n=460) encompassing gamers of 22 communities, we gauge the extent to which players

979-8-3503-5067-8/24/$31.00 ©2024 IEEE

may "unconsciously contribute" to AIA (§V).
Based on these contributions, we present our vision for mitigation of AIA and discuss directions for future work (§VI).

## II. Background and Motivation

We first describe the major characteristics of AIA (§II-A), and highlight the research gap (§II-B).

### A. Gentle introduction to AIA in the Videogame Ecosystem

To elucidate why AIA are a concerning issue, we emphasize the necessary steps to enact AIA, and highlight the ways in which the gaming community can be affected by AIA.

*1)* ***Requirements****:* AIA are possible thanks to the capabilities of machine learning models to find hidden patterns in data after undergoing a training phase [11, 12]. In the scenario envisioned by Tricomi et al. [7], a model is trained to associate *(a)* a player's in-game statistics with *(b)* their personal attributes—the latter being the ground truth used to guide the learning phase of the model. Hence, to enact AIA an attacker needs to collect a certain amount of such "associations" which will enable the model to infer the personal attributes of other players (i.e., those not included in the training dataset).

*2)* ***Collection****:* There are many ways an attacker can use to gather the associations required to develop an AIA-ready model [7]. However, real attackers operate with a cost/benefit mindset [13], favoring tactics that are cheap to stage. The ecosystem of multiplayer videogames makes it easy for evildoers to prepare an AIA: *tracking websites are a reservoir of in-game statistics*, and players publicly share their profiles [6]; whereas *personal attributes are obtainable by exploiting the social nature of videogamers*. In particular, *online surveys* are a convenient means of doing so: thousands of players provided their "personal attributes" in two Dota2 surveys shared on reddit [14]; Tricomi et al. [7] distributed their questionnaire on other social networks, collecting hundreds of responses with minimal effort. It is even possible to "recruit" participants by offering a small compensation as an incentive (e.g., [9, 15, 16]). Of course—contrarily to all these research works—a real attacker would do so "unethically."

*3)* ***Consequences****:* With an AIA-ready model, the attacker can hence exploit its predictive capabilities to infer the personal attributes of "unknown" players. Such a weapon can be used in three exemplary ways [7]: *(i)* Given the in-game statistics of a player, infer its personal attributes. For instance, a mischievous player may want to see if their opponents can be verbally harassed during a match [17, 18]. *(ii)* Given the in-game statistics of multiple players, find associations "in bulk." For instance, an attacker can scrape the public profiles of many players, infer those who are more willing to purchase in-game content, and sell such information to advertising companies for targeted ads [19]. *(iii)* Identify specific individuals among a set of players. This is a variant of the previous way: an attacker may want to pinpoint the underage players among all the users of a given tracking website—and then bully them [20].[1]

[1]For low-level technical details about machine learning for AIA, see [7, 11].

**⚠ A subtle threat.** The gaming ecosystem is exposed to AIA due to the intrinsic "social" nature of its players. However, the elusive side of AIA is that even those players who are aware of privacy issues can fall victim to AIA.[a]

[a]If "unaware" players participate in (unethical) surveys, the attacker can use such data for AIA and infer the personal attributes of "aware" players.

### B. AIA and Privacy in Gaming Research (Related Work)

To further motivate the need for this vision paper, we show that the themes of AIA and *privacy* in general tend to be overlooked in related literature—starting from IEEE CoG.

*1)* ***IEEE CoG****:* This venue changed its name from "CIG" to the current one in 2019. Hence, we take all papers that have appeared in the proceedings of IEEE CoG from 2019–2023, obtaining 682 papers. Next, we perform a keyword search for the terms "AIA", "attribute inference attack" and "privacy". As expected, we find no match for the first two—despite the term having gained visibility in the security community since 2016 [12], but in the social network context. For "privacy", we have 20 papers (3%) that mention the term at least once: a quick investigation reveals that for 10 of these papers the term appears only in references/appendices, whereas in 6 papers it is mentioned only for "data collection procedures" (e.g., to indicate that user studies have been done respectfully of participants' privacy); in 3 cases it is mentioned out of context.[2] The only relevant match (out of 682) is [22], which hints at potential privacy issues in the ethical statement.[3]

*2)* ***Google Scholar****:* We refine our analysis by querying Google Scholar (in Dec. 2023) with the term "attribute inference attack" and "game". We specifically look for papers that deal with this issue in the gaming ecosystem (some works [23] mention "game theory", which is outside our scope). We could not find any match besides the work by Tricomi et al. [7]. We then expanded our search by looking for game-related studies that look for associations between players' in-/off-game data, scrutinizing whether such papers hint at potential "attacks" or "privacy violations" that can be exploited by leveraging such associations with publicly available resources. We could not find any match: all such papers do not stress the privacy implications (towards players) of their own findings.[4] Some papers carry out privacy-centered user studies (e.g., [27]), but do not focus on in-/off-game associations. Others raise the attention [1] on the data collection policies in the gaming ecosystem, but without carrying out any original investigation.

**Problem Statement.** AIA are overlooked by game-related research.[5] We challenge the status quo, and: examine the current gaming landscape, revealing *where* and *why* AIA can be staged; gauge the players' *awareness and perception*

[2]Intriguingly, it occurs 33 times in [21] because it considers "how *bathrooms* in games convey the sense of 'privacy' expected by real-world ones."

[3]Even [9], despite collecting 2.5k responses pertaining to LoL, has no occurrences of either "privacy" or "ethic" (and, of course, "AIA").

[4]E.g., there are 6.7k players considered in [24], over 4k in [25], and 1k in [26], but the term "privacy" never occurs in either (and [26] is from PETS).

[5]AIA are known in the security/privacy domain, but we want to make the *gaming* community aware of this threat—and CoG is the best venue for this.

*of AIA*; and outline *what can be done* by all stakeholders to collectively mitigate the threat of AIA in videogames.

## III. Videogames Prone to AIA (in theory)

As our first major scientific contribution, we pinpoint the games that "theoretically" enable the enactment of AIA, and explain how to do so systematically (§III-A).[6] We will use these findings (§III-B) as a basis for the rest of our research.

### A. Criteria (how to determine if a game is prone to AIA?)

To reach our first goal, we examine the landscape of multi-player videogames under the lens of an attacker willing to carry out AIA. Specifically, we ask ourselves: "what games present the characteristics that would make an AIA *economically attractive* and *practically viable*?" We answer this question by reflecting on our previous explanation (§II-A2). Hence, we derive an **original list of assessment criteria**, centered on the attacker's cost/benefit mindset [13], that must be scrutinized to determine whether a videogame is "AIA-prone." We express our list through four high-level questions:

- "*Is the game popular?*" Setting up an AIA requires expertise and a resource investment (for data collection, filtering, and model training), hence games with a small playerbase may not be attractive for real attackers.
- "*Do tracking websites exist for the game?*" Tracking websites are essential to ensure that the AIA is feasible (collecting in-game statistics from the game itself is doable, but much less practical [7]).
- "*Does the playerbase contribute to online surveys?*" If true, then it would be a signal that harvesting the ground truth (via "unethical" surveys) will yield practical results (i.e., many and heterogeneous responses) for the attacker.
- "*Have correlations between players' in-/off-game data been found for this game?*" If a prior study proved the existence of such correlations (e.g., [9]), then an attacker can use it as a scaffold for preparing the AIA.[7]

In summary,[8] a videogame for which the answer to all the abovementioned questions is "yes" is an AIA-prone title.

⚠ **An unsettling scenario.** Our criteria rely on the assumption that an attacker develops an AIA-ready model *from scratch.* However, it is entirely possible to "share" an AIA-ready model, which can be used at no cost by any entity with malicious intentions. We hope this is not already happening.[a]

[a] Albeit darkweb marketplaces do deal with similar "merchandise" [28].

### B. Findings (what games are prone to AIA in 2024?)

Our investigation follows the criteria presented in §III-A.

**Method.** We begin by looking for those (multi-player) games having a large playerbase. Hence, we rely on popular websites (e.g., activeplayer, steamcharts, playercounter) to derive a list of the 20 games having the largest number of concurrent players. Then, for each game, we

- search for a *tracking website*, i.e., a platform which must be *(i)* publicly available, and which *(ii)* allows to retrieve extensive in-game statistics of a large subset of the game's population, which could *(iii)* potentially be useful for AIA (for instance, we exclude "achievement trackers");
- search for *prior surveys*, through Google queries with the title and two terms among "player,survey,results,reddit". We consider only surveys with >200 responses, and disseminated by individuals unrelated to the game-devs;
- search for *literature-found correlations* between in-/off-game data, through queries on Google Scholar with the title of the game and any combination of "correlation,analysis,profiling,inference,privacy,personality,prediction". We used the snowball method to further investigate the references of relevant papers.

We perform these operations in Oct. 2023, and we repeat them another time in Feb. 2024 for validation purposes.[9] The results of our analysis are shown in Table I (described in the caption).

**Ethical Statement:** We acknowledge that our findings may be helpful for evildoers. However, we follow *ethical disclosure* [29, 30] thereby respecting the right of gamers to be informed. Nonetheless, we emphasize that the authors of prior work that found correlations or carried out surveys were acting in good faith (and, likely, they were not aware of AIA) and their findings are beneficial for science.

**Results.** We found tracking websites for 15 games, online surveys for 17, and also that 12 have correlations discussed in related literature. Given our criteria (§III-A), we consider those games (which are "popular" by definition) having all three of these elements (11 in total) to be AIA-prone, and are marked with a red cell in Table I. In contrast, games for which we could not find a tracking website (which is crucial) are marked with a light-blue cell, and we consider them to have a low likelihood to be involved in AIA (e.g., for Minecraft we could not find a tracking website that could enable AIA). As a side note, all papers discussing these correlations *never* mention the word "privacy" (aside from Tricomi et al. [7]).

⚠ **Attackers appreciate these efforts.** Works that announce the existence of correlations between in-/off-game data make the attackers' job easier (§III-A). The attacker can also use results from prior surveys for validation purposes after (unethically) collecting their own dataset (see [7]).[a]

[a] Prior surveys may even induce attackers to steal such data for AIA [31].

## IV. Collecting Personal Data in Gaming Communities (and survey setup)

After outlining some practical ways to setup an AIA (§IV-A), we will examine some communities of all games in boldface in Table I from the perspective of online surveys (§IV-B).

[6] Tricomi et al. [7] listed some games wherein AIA could be conceived, but such a list included only eSport (which are a subset of multiplayer games), and was not derived with a systematic approach (which we adopt and propose).

[7] AIA require the existence of correlations that allow the model to associate public with private data [7]. The attacker can find the correlations autonomously (as done in [7]) but this increases the cost of the campaign.

[8] We present high-level criteria. However, depending on the attacker's objective, there may be additional fine-grained ones, e.g., "*do underage people play the game?*" or "*is this game popular among people of a certain gender?*"

[9] The operations are done by two authors, who resolved issues via discussions. For some games, we consider also their previous installments, since they tend to be similar (e.g., for BF2042, we consider [24] which is on BF3).

TABLE I: **Games prone to AIA –** We check the 20 games having the highest number of concurrent players. We report the number of active and concurrent players (over the last 30 days), the link to an exemplary tracking website, the number of participants of a representative survey (and its link), and a paper which showed correlations/predictions between the in-/off-game data. Red (blue) cells denote games being prone (not prone) to AIA (according to §III-A); games in boldface are those considered in our following analyses.

| Game | Popularity Active – Concurr | Tracking Website? | Prior Survey? | Correl. Found? |
|---|---|---|---|---|
| **LoL** | 142M – 900K | ✓ | 3.7k | [9] |
| **WoW** | 32M – 250K | ✓ | 500 | [32] |
| **CSGO** | 31M – 900K | ✓ | 13k | [33] |
| **Fortnite** | 237M – 1.15M | ✓ | 1k | [34] |
| **PUBG** | 320M – 200K | ✓ | 4.4k | [35] |
| **OW2** | 25M – 350K | ✓ | 3.2k | [33] |
| **Valorant** | 24M – 600K | ✓ | 1.4k | [36] |
| **CoD:WZ** | 71M – 300K | ✓ | 751 | [37] |
| **RS:S** | 10M – 120K | ✓ | 4.8k | [33] |
| **Destiny2** | 14M – 50K | ✓ | 450 | [38] |
| DOTA2 | 14M – 430K | ✓ | 7.3k | [7] |
| **Apex** | 52M – 250K | ✓ | 296 | |
| **RktLg** | 85M – 220K | ✓ | 6k | |
| **GTA:O** | 24M – 110K | ✓ | 1.9k | |
| **BF2042** | 300K – 15K | ✓ | | [24] |
| FIFA | 5M – 50K | | | |
| Minecraft | 169M – 900K | | 4k | |
| Roblox | 213M – 1.5M | | 1.5k | |
| HeartStone | 6M – 370K | | 21k | |
| Wildlands | 370K – 5K | | | |

### A. *Reflection: how to collect data for AIA?*

Even if a game has a tracking website, the attacker must still collect the data (and, particularly, ground truth for personal attributes) necessary to train an AIA-ready model (§II-A).

**Options.** We put ourselves in the attacker's shoes once more: "what would the attacker do to gather in-/off-game–data associations of players?" We identify two main possibilities:

- *Crawling Social Networks.* Nowadays, many gamers post clips or highlights of their games on social media, sometimes public ones (e.g., Twitter, Twitch or YouTube). These highlights might include information usable to setup an AIA, such as gamertags and favorite characters (as could be seen by aggregating across clips). Moreover, the authors of such clips may have their real identity known or easily discernible: for instance, one could infer that they are a female, or their age, or their preferences [11]. These personal attributes can then be associated to the in-game data obtainable from the clips, and used to setup AIA.
- *Deceitful Surveys in Communities.* An alternative, as demonstrated by Tricomi et al. [7], is to find ways to post/distribute surveys in various "communities", i.e., (online) platforms wherein users interested in a common subject (ideally a videogame) tend to meet and socially interact with each other. An attacker can hence post ill-purposed surveys (potentially by deceiving the mods via social engineering [39]) and use these to collect exact matches of in-game/off-game associations, and use these to setup an AIA.

We now delve deeper into the "survey" option.

**Challenges of Surveys.** While making a survey is trivial, ensuring that such a survey achieves the intended *malicious* purpose is not.[10] Online social networks, such as reddit or forums, can have millions of users. To prevent "spam" which would annoy the community, the administrators typically enforce content moderation policies. For instance, new accounts may be prevented from posting in some sections of a given board; or certain content (e.g., links) must be approved before being publicly displayed. Therefore, an attacker attempting an AIA must deal with such obstacles. In what follows, we will ethically explore such challenges by analysing the guidelines of various communities with respect to surveys.

[10] An attacker can also adopt a brute-force strategy, and create (or purchase) fake accounts/bots which continuously post the link to the survey on various social networks, potentially promising rewards to increase the response rate.

### B. *Community Selection and Analysis*

We analyse some communities related to the games in Table I. Our purpose is twofold: *(i)* understand their guidelines in terms of posting surveys, and *(ii)* identify communities in which we can post our survey to gauge the player's attention and awareness on privacy in video games (which we cover in §V).

**Systematic Approach.** Our approach to select communities is illustrated in Fig. 2. Given a game, we first search the Web for relevant communities, excluding those that are clearly inactive or have a small population. Then, we qualitatively inspect their guidelines. Four cases can happen:

3 posting a survey is generally allowed without approval (possibly under some conditions that the survey must fulfill);
7 posting a survey is explicitly not allowed;
**?** policies do not mention anything about surveys at all and give no clues as to whether they are (dis)allowed
✎ mods must approve (or at least be contacted), or the policies are unclear about whether surveys can be posted (e.g., they may mention that certain types of content are not allowed).

If 3 (or 7), we will post the survey (or discard the community). Otherwise, for either **?** (due to uncertainty) and ✎, we either:

- contact the mods (potentially sending reminders). We will either receive a positive response (👍) and we will post the survey; or we will get a negative response (👎). If the response is unclear or no response is received after at least 24h (🕓), we post the survey (in some cases, we waited more than 2 months before receiving a response).
- if there is no explicit entity to contact for such inquiries, or the mod are inactive, we will post the survey right away.

If we are not given explicit approval, we write a shout-out to the mods, inviting them to "reach out in case anything is non-compliant." As a rule, if we receive explicit indication that *our survey* is not compliant, we do not consider the community.

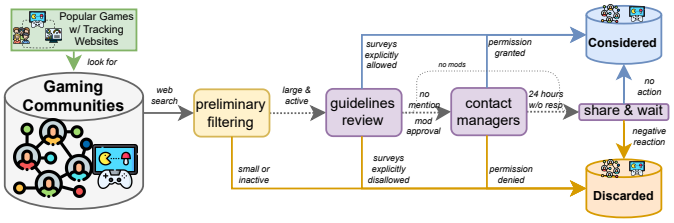


Fig. 2: **Workflow for the Community Analysis** – After identifying communities with a large userbase, we review the guidelines; if necessary, we contact the administrators to get their permission: if we receive no response, we post a survey in a given community, and see what happens.

**Procedure.** For every game among the 14 we consider, we look in well-known forums, on reddit, and also in popular Twitch streams: we set ourselves the target of finding 50 communities (we cannot search the whole Web) that: are "large

and active" (>1k recent users) and whose userbase has interest in our considered 14 games; this requirement can be either explicit (e.g., name of the game in the title of the community, such as /r/wow/) or we derive with a qualitative analysis (e.g., for r/truegaming, there are many posts on our games) Importantly, we focus on international communities, i.e., we do not consider communities of a specific language/region. Hence, for each potential community, we check if it matches our criteria. We eventually found 50 communities (among which we have 70 streamers, which we count as a single "Twitch" community); most of these communities are from Reddit. After reading their guidelines (if available), 3 explicitly allowed surveys, whereas 12 did not (e.g., Steam), 17 did not mention any rule related to surveys (or did not have rules to begin with), and 18 required some contact with the mods. We then contacted the community managers (we could not find any specific contact for 5 communities, and for /r/ubisoft the only mod had been inactive for years) and asked for their approval to post a survey focused on video games. In our exchanges, we always provided the link to the survey. We received permission by 9 communities, so we posted the survey. We were denied approval by 9 communities, which we will not consider. We received no clear denial to post by 17. We posted the survey in 15 of these (we did not post in r/privacy and DynamoGaming because it was explicit that mod approval was required). Among these 15 communities, 9 did not take down our posts; for 6 the posts were taken down (and it was often made explicit that it was because surveys were not accepted, or ours was not allowed). Overall, we received direct (or indirect) approval from 21 communities, which we report in Table II; we reposted the survey up to two times (if the repost is taken down, we will not consider its responses).

**Adversarial Considerations.** Let us analyse our findings in Table II from an attacker's viewpoint.

- Three communities (3) allow surveys by default. This is a double edged sword: these communities facilitate research, but attackers can leverage their openness. However, surveys may be taken down if they do not comply with rules.
- Many communities (✎) have vetting policies. For instance, we were required to provide our identities, institutional email address, and questionnaire before gaining approval. This implies that an attacker would incur a substantial effort to exploit such communities for AIA.

Finally, some communities are very strict, and explicitly turned down our survey because it did not meet some requirements, or simply do not accept surveys at all.[11]

**Helpful mods.** In our interactions with the mods, we observed a positive attitude. Even when turning down our request, they provided us suggestions for alternative communities. This shows that communities listen to researchers.

[11]Excluded communities (29): r/xboxone, r/gamingsuggestions, r/gamingnews, r/esports, r/summonerschool, r/xboxseriesx, r/gaminglaptops, r/fortnitebr, r/pcgaming, r/gaming, r/games, SteamCommHub, r/destiny2, r/gtaonline, r/codwarzone, r/apexlegends, r/Steam, r/DestinyTheGame, r/gamecollecting, Twitch, DynamoGaming, r/GlobalOffensive, r/privacy, r/battlefield2042, r/pubattlegrounds, ChillZone, TheTechGame, r/mmorpg, r/playstation

TABLE II: **Communities which participated in our survey –** We report the 21 communities (having players of the 14 bolded games in Table I) that either: allowed surveys by default (red), or after messaging the mod (blue), or which did not remove our survey after we posted it (yellow cells have mods, but did not respond to our messages; gray cells do not have specific/active mods. We also report the messages we sent (overall, we sent over 120 messages to 50 communities, including those not shown in this table).

| Community<br>Large and Active | Survey Allowed? | Admin Response? | Msgs Sent |
|---|---|---|---|
| truegaming | 3 | | |
| SampleSize | 3 | | |
| JoyFreak | 3 | | |
| Rainbow6 | ✎ | 👍 | 1 |
| wow | ✎ | 👍 | 1 |
| leagueoflegends | ✎ | 👍 | 6 |
| VALORANT | ✎ | 👍 | 3 |
| youtubegaming | ✎ | 👍 | 1 |
| Overwatch | ✎ | 🕒 | 5 |
| GameTheorists | ? | 🕒 | 1 |
| videogames | ? | 🕒 | 1 |
| consoles | ? | 🕒 | 1 |
| AskGames | ? | 👍 | 1 |
| Instant Gaming | ? | 👍 | 1 |
| RocketLeague | ? | 👍 | 3 |
| gamers | ? | 👍 | 1 |
| ubisoft | ? | – | 2 |
| PC Gamer | ? | – | |
| COD Forums | ? | – | |
| Valorant Forums | ? | – | |
| GTA Forums | ? | – | |

## V. Assessing Players' Privacy Awareness (on AIA)

As our final contribution, we now focus on the user survey, which we mentioned in the previous section. The purpose is to assess the overall privacy awareness of players w.r.t. their in-game data and habits, with a focus on those potentially usable for AIA—inspiring reflections about these subtle issues.

### A. Survey Design

Our questionnaire is publicly observable in our repository [40].

**Overview.** We distribute our survey in the communities in Table II. Our survey has a similar structure to the one by Tricomi et al. [7] with two major differences. **(1)** Our survey is shorter. Specifically, we will ask less questions about personality and the game than [7]. This is because *our intent is not to enact an AIA*. **(2)** Our surveys mostly assume a "multi-game" setting. The survey in [7] focused only on DOTA2, whereas ours focuses on the 14 games in boldface in Table I.

**Organization.** The questionnaire has five sections:

1) *Demographics.* We ask exactly the same questions as [7].
2) *Personality.* We ask six personality-related questions.
3) *Gaming.* We first inquire for the most played game (among the 14 we consider; plus an "other" category).[12] Then, depending on the choice, we ask three questions; one for the gamertag (necessary for AIA), one for validation purposes (e.g., "what is your most played hero?", which we can verify with the gamertag), and one for generic knowledge about the game (an attention check).
4) *Privacy,* We ask eight privacy-related questions, and inquire for concerns about privacy issues in videogames.
5) *Extensions.* We ask three questions, inquiring if the participant regularly plays any other game among our considered 14, and which communities they follow.

[12]For some communities, we designed an ad-hoc questionnaire that focused on the specific game, and for which this question was omitted.

We distribute our questionnaire for every community (Table II) in Dec.2023/Jan.2024, and collect responses over 3 weeks.[13]

**Ethical Statement:** We follow ethical guidelines [30]. Our institutions are aware of our research but have no formal IRB. We informed the participants that: *(i)* our study was for research; *(ii)* the questionnaire was anonymous; *(iii)* their data would not be released; *(iv)* questions were about "various aspects" of video games. After submission, the participant is informed about AIA (linking to [7]) and invited not to mention AIA or privacy in the discussion to avoid priming other users, thereby potentially causing bias in future responses.We also provided our institutional contacts, stressing our availability for inquiries. Participation was voluntary and we offered no compensation. Our surveys do not ask for sensitive data and no harm is done to our participants. Our questionnaire is shorter w.r.t. the one by [7] because *we do not want to "simulate" an AIA* and *we do not want to assist attackers by providing "novel" information* about potential correlations obtainable through our survey, which is also why we *(i)* do not set any target number of responses and *(ii)* will not provide details on demographics or personality.

### B. Risk Assessment (are these surveys useful to an attacker?)

Here, we focus on three aspects that are most relevant for the sake of our paper—gauging how useful our surveys are from the perspective of an attacker who wants to carry out an AIA.

**Which responses are valid?** By aggregating the results of all our surveys, we obtain 579 responses. Of these, we remove 28 because they specified an "other" game which did not exist or failed the attention check. We then analyse the remaining 551 responses, scrutinizing which ones are "useful" for an attacker. Specifically, we focus on those answers that *provided a "valid" gamertag*. We find that 91 (16%) answers included an incorrect gamertag, or one which did not match the validation. Intriguingly, we find that in many instances the string provided in the gamertag was *criticizing our survey*. For instance, some participants wrote "NotGivingThatInformation" or "invasive, not answering". We find this intriguing: when we posted our surveys, we clearly specified that the questionnaire required users to provide their gamertag—hence, users were aware of this request.[14] We believe that these "skeptical users" are a *positive result* from the perspective of AIA, since it shows that not all users "blindly" trust requests to fill online questionnaires. However, the remaining 460 (79%) responses can be used for AIA: we will now analyse these.

**Mindful players.** Some participants of our survey refused to provide their in-game handle. We find it *positive*: there is no true reason for providing such information (which we use for validation), which is vital for attackers to setup AIA.[a]

[a]The username is necessary to associate the ground truth (i.e., personal attributes) to the in-game statistics retrievable through tracking websites [7].

**What games do our participants play?** We analyse the 460 valid responses, investigating the extent to which our surveys enabled to collect data of our considered 14 games (having a tracking website). We report the results of this analysis in Table III, showing the top10 "popular" games among our participants. Specifically, the first row shows the number of participants which marked a game as their "primary" game, and the second row denotes those who specified the game as "another game that they play regularly" (at the end of the survey). We can see that our questionnaires enabled to solicit (valid) responses from ≈100 players for 5 of our games (all of which are AIA-prone), despite our limited efforts (e.g., we offered no incentive/reward). Importantly: *AIA do not need to consider only "professional" players* (as hinted in [7]). Hence, even if a player does not have one of our games as their primary choice, they would still participate in a survey that could fuel an AIA—if the attacker tricked the community.

TABLE III: **Most popular games (top10)** – We aggregate the "primary" with "other" games (among the 14 considered) *often played* by participants. The color refers to Table I. Seven games are often played by >80 participants.

| Game | OW2 | LoL | CSGO | WOW | RS:S | Apex | GTA:O | Vlrnt | Frtnt | Dstny2 |
|---|---|---|---|---|---|---|---|---|---|---|
| **Primary** | 34 | 54 | 12 | 40 | 48 | 8 | 16 | 27 | 8 | 11 |
| **Other** | 102 | 67 | 95 | 52 | 43 | 79 | 66 | 49 | 65 | 44 |
| **Total** | **136** | **121** | **107** | **92** | **91** | **87** | **82** | 76 | 73 | 55 |

**What communities are responsive?** Lastly, we examine the communities that solicited the most responses to our surveys (recall that we distributed one questionnaire to each community in Table II; we ensure there are no duplicate answers). We report the top-10 most "responsive" communities in Table IV. We see that these communities are mostly from reddit. Intriguingly, we received no response at all from three communities entailing forums/boards (JoyFreak, COD Forums, Valorant Forums). More generally, we believe users of such platforms to be unlikely to "contribute" to AIA. We conclude by inspecting the answers to "which gaming communities do you follow?". The top-3 most common responses are: 346 (75%) "Reddit", 177 (38%) "Discord", 79 (17%) "Steam". Reddit being first is expected (most of our respondents *are* from reddit!). The popularity of Discord (which we did not investigate thoroughly) makes such a channel also viable for AIA. Steam being third is encouraging: the current guidelines of Steam prohibit surveys. However, we did find some surveys on the Steam's Community Hub (e.g., [41–44]).

TABLE IV: **Most responsive communities (top10)** – Communities (cf. Table II) from which we received the most (valid) responses to our survey.

| Commu. | /r/truegaming | /r/Rainbow6 | /r/SampleSize | /r/wow/ | /r/leagueoflegends | /r/Overwatch | /r/videogames | /r/VALORANT | /r/GameTheorists | gtaforums.com |
|---|---|---|---|---|---|---|---|---|---|---|
| **Absolute** | 207 | 42 | 35 | 30 | 22 | 21 | 19 | 18 | 18 | 12 |
| **Relative** | 45% | 9.13% | 7.61% | 6.52% | 4.78% | 4.57% | 4.13% | 3.91% | 3.91% | 2.61% |

### C. Awareness (is data-privacy in our participants' mind?)

We now focus on the questions in the fourth section of our survey, inquiring the participants' opinion on privacy-related issues in video games—inspiring a reflection on these topics.

**What do you know?** The first question asks participants to rate their "knowledge about data collection and privacy issues in videogames"; the answer is in a [1–6] Likert scale (1: novice, 6: expert). Across 460 (valid) responses, the average value is 3.13 (std=1.3), which is below the middle point of 3.5 (confirmed with a t-test, $p \ll 0.05$): this denotes that our participants are not very informed about privacy in general.

[13]We carried out pilot tests with colleagues for feedback (avg length=10m).

[14]Users of 6 communities even voiced such a concern in the thread.

**Have you ever worried?** Next, we consider the responses to three (binary) questions: "do you know that your gaming data are being collected by other entities?" and "do you know that it is possible to predict your personal attributes from your gaming data?" (i.e., AIA), and "have you ever worried about your anonymity in games being compromised?". We visualize the responses to these questions with the 3D-plot in Fig. 3, showing the inner relationships between the responses to each questions. At a high-level, 403 (88%) know that their data are being collected, but 112 (24%) *do not know* about AIA, and 267 (58%) *have never worried* about their anonymity being compromised. We find it instructive to further analyse these results: among those (348, 76%) who "know about AIA", 146 have worried about their anonymity, but 202 have not. Given that the overall level of knowledge is below average, this result indicates that even if players "know" (or "suspect") that their personal attributes can be predicted, they may overlook *what can be predicted* (and Tricomi et al. [7] showed that certain attributes are easy to infer, such as age and occupation).

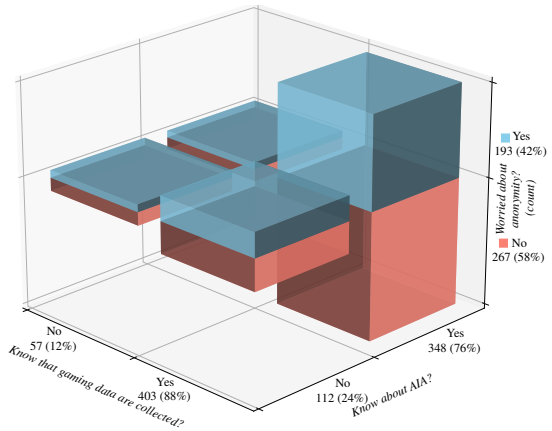


Fig. 3: **Distribution of the responses to three privacy questions** – Each axis denotes a (binary) question; we report the overall number of responses on the axes. The vertical axis shows the "count" of the responses and the question "have you ever worried about your anonymity being compromised?"

**Do you share your data?** Lastly, we inquire about data-sharing. First, we ask "did you explicitly choose to share your statistics?"; four answers were possible. The results are enlightening: 179 (39%) answered "No (there is no option or I do not know about such an option)"; 141 (31%) answered "Yes"; 84 (18%) answered "I am not sure"; and only 56 (12%) answered "No, I explicitly choose not to". These results underscore the *lack of transparent opt-out options*. As for the last question, "Generally, would you choose to publicly share your personal data (age, gender, marital and economic status etc.)?", 348 (76%) answered "No" and 112 (24%) answered "Yes". We find this turnout alarming: first, because *these participants did provide such data* in our survey; second, because such people clearly do not want such information to be known to others, *but they can be targeted by AIA*—which, as shown in [7], can reveal such personal attributes.

## VI. Discussion, Disclaimers, and the Way Forward

We now coalesce and discuss our findings and outline our vision for mitigation of AIA in video games.

Our findings reveal that privacy concerns raised in the previous academic work have broad implications across the gaming landscape. Our analyses, supported by the study of 20 highly popular online video games built on our novel assessment criteria, suggest that more than half of these games may be prone to AIA. Furthermore, we have established that gaming communities, and especially guidelines with respect to conducting surveys among their user base, may play a substantial role for the privacy of individual gamers. Finally, we observed that, although players may be aware of their data being collected, a non-negligible fraction of them would still be willing to share their personal data. The latter finding potentially exposes *entire player bases* to subtle AIA threats.

From the scientific perspective, our results (in §III) can serve as a guide for further empirical studies of gaming communities, especially under the lens of privacy. Specifically, our assessment of the exposure of specific games to AIA reveals which games can be studied with low risk of AIA (e.g., Hearthstone, or those in [45]), and for which games special care should be taken to ensure players' privacy. However, we acknowledge that our user study (§V) with 460 users is preliminary in nature and of limited statistical significance, as there are billions of gamers in the world. Future studies should address their perspective in more detail. Needless to say, the ethical aspects of such studies would require extreme care.

We hope that our findings motivate the video game ecosystem to reflect on its privacy challenges. The relationships between in-/off-game data discovered in previous work have substantial practical implications. Furthermore, we can envisage various possibilities for attackers to leverage or gather such relationships in insidious ways. For example, they can *(i)* trick community moderators with social engineering, *(ii)* incentivize participation in surveys with economical rewards, *(iii)* share or rent AIA-ready models, and *(iv)* exploit the fact that many players are not adequately informed (§V-C). Therefore, we emphasize the need for collaboration of all key stakeholders in the gaming ecosystem, giving rise to:

**Our Vision.** We strive to ensure that the risks of AIA are properly accounted for in the gaming ecosystem. We hence argue that the recommendations by Tricomi et al. [7], who only mentioned players and developers, are *expanded* to incorporate researchers into the quest for privacy in online videogames. Specifically, we endorse that:

- Researchers *adopt a privacy-oriented mindset* in their research, exploring technical privacy instruments as well as social aspects of gaming communities, transferring their findings to developers and educating players.
- Developers *introduce privacy-oriented features* into games as well as game-tracking sites and uptake scientific results.
- Players *develop awareness* to privacy-related issues, voice their concerns when they suspect potential privacy threats, and demand for more privacy-preserving options to be implemented in the gaming ecosystem.

Finally, *security researchers can help*, too—but they may not be aware of problems in gaming.[a]

[a] During some discussions on the Distinguished Paper Award winner [46] at USENIX SEC23 (a top-venue in security), some security researchers commented that "I did not know *aimbots* were a security problem!"

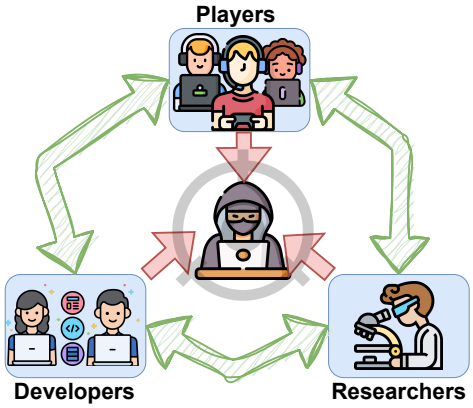


Fig. 4: **Our Vision to counter AIA in Videogames** – AIA can only be mitigated with a collective effort. *Researchers* should inform players and developers of novel privacy threats. *Developers* should account for the findings of research, and devise privacy-friendly initiatives. *Players* should be aware of privacy risks, and engage in social activities with a privacy-oriented mindset.

**CONCLUSIONS.** This paper is a call to action. We have provided further evidence of the threats of AIA in videogames. AIA are a subtle (and overlooked) threat that can target even privacy-aware players, as well as minorities (e.g., children). Every stakeholder (gamers, developers, researchers) should be more mindful of privacy, and collectively contribute to raising the overall awareness of privacy-related issues in videogames. We hope our paper inspires such a change.[15]

**Acknowledgement.** We thank: the gaming communities for their contributions; the IEEE CoG's organizers and Pavel Laskov for many suggestions; and the Hilti Group for funding.

## REFERENCES

[1] J. L. Kröger, P. Raschke, J. Percy Campbell, and S. Ullrich, "Surveilling the gamers: Privacy impacts of the video game industry," *Entertainment Computing*, 2023.
[2] https://statista.com/outlook/dmo/digital-media/video-games/worldwide.
[3] https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2021-free-version.
[4] theesa.com/resource/2022-essential-facts-about-the-video-game-industry.
[5] D. Staat, G. Wallner, and R. Bernhaupt, "Towards a community-based ranking system of overwatch players," in *Int. Conf. Entert. Comp.*, 2022.
[6] E. Kleinman and M. S. El-Nasr, "Using data to" git gud": a push for a player-centric approach to the use of data in esports," in *CHI*, 2021.
[7] P. P. Tricomi, L. Facciolo, G. Apruzzese, and M. Conti, "Attribute inference attacks in online multiplayer video games: A case study on dota2," in *ACM CODASPY*, 2023.
[8] D. Martinovic, V. Ralevich, J. McDougall, and M. Perklin, ""you are what you play": Breaching privacy and identifying users in online gaming," in *IEEE PST*, 2014.
[9] Z. Wang, A. Sapienza, A. Culotta, and E. Ferrara, "Personality and behavior in role-based online games," in *IEEE CoG*, 2019.
[10] R. Sifa, A. Drachen, and C. Bauckhage, "Profiling in games: Understanding behavior from telemetry," in *Social interactions in virtual worlds: An interdisciplinary perspective*, 2018.
[11] N. Z. Gong and B. Liu, "Attribute inference attacks in online social networks," *ACM TOPS*, 2018.
[12] ——, "You are who you know and how you behave: Attribute inference attacks via users' social friends and behaviors," in *USENIX Sec.*, 2016.
[13] G. Apruzzese, H. S. Anderson, S. Dambra, D. Freeman, F. Pierazzi, and K. Roundy, ""Real attackers don't compute gradients": Bridging the gap between adversarial ml research and practice," in *IEEE SaTML*, 2023.

[15] In the supplementary material (also available in our publicly-accessible repository [40]) we include: the questionnaires of our survey (§V-A); a document explaining our literature analysis (§II-B) and some recommendations received by the communities; as well as additional resources.

[14] https://docdroid.net/ZeJTLar/rdota2-demographics-report-2021-pdf.
[15] D. Kao, R. Ratan, C. Mousas, A. Joshi, and E. F. Melcer, "Audio matters too: How audial avatar customization enhances visual avatar customization," in *ACM CHI*, 2022.
[16] A. Eidelberg, C. Jacob, and J. Denzinger, "Using active probing by a game management ai to faster classify players," in *IEEE CoG*, 2019.
[17] A. Canossa, D. Salimov, A. Azadvar, C. Harteveld, and G. Yannakakis, "For honor, for toxicity: Detecting toxic behavior through gameplay," in *ACM CHI PLAY*, 2021.
[18] R. Kowert and C. Cook, "The toxicity of our (virtual) cities: prevalence of dark participation in games and perceived effectiveness of reporting tools," in *HICSS*, 2022.
[19] G. Johnson, J. Runge, and E. Seufert, "Privacy-centric digital advertising: Implications for research," *Customer Needs and Solutions*, 2022.
[20] P. C. Ferreira, A. M. V. Simão, A. Paiva, C. Martinho, R. Prada, A. Ferreira, and F. Santos, "Exploring empathy in cyberbullying with serious games," *Computers & Education*, 2021.
[21] D. Antognoli and J. Fisher, "The purposes and meanings of video game bathrooms," in *IEEE CoG*, 2021.
[22] J. T. Bowey, J. Frommel, B. Piller, and R. L. Mandryk, "Predicting beliefs from npc dialogues," in *IEEE CoG*, 2021.
[23] J. Jia and N. Z. Gong, "Attriguard: A practical defense against attribute inference attacks via adversarial machine learning," in *SEC*, 2018.
[24] S. Tekofsky, P. Spronck, A. Plaat, J. Van den Herik, and J. Broersen, "Psyops: Personality assessment through gaming behavior," in *International Conference on the Foundations of Digital Games*, 2013.
[25] T. Kennedy *et al.*, "Predicting mmo player gender from in-game attributes using machine learning models," in *Predicting real world behaviors from virtual world data*, 2014.
[26] P. Likarish, O. Brdiczka, N. Yee, N. Ducheneaut, and L. Nelson, "Demographic profiling from mmog gameplay," in *PETS*, 2011.
[27] A. C. Tally, Y. R. Kim, K. Boustani, and C. Nippert-Eng, "Protect and project: Names, privacy, and the boundary negotiations of online video game players," *Proc. ACM Human-Comp. Inter.*, 2021.
[28] https://netenrich.com/blog/fraudgpt-the-villain-avatar-of-chatgpt.
[29] T. Kohno, Y. Acar, and W. Loh, "Ethical frameworks and computer security trolley problems: Foundations for conversations," *SEC*, 2023.
[30] M. Bailey, D. Dittrich, E. Kenneally, and D. Maughan, "The Menlo report," *IEEE Security & Privacy*, 2012.
[31] https://bleepingcomputer.com/news/security/hackers-try-to-extort-survey-firm-questionpro-after-alleged-data-theft/.
[32] Z. Halim, M. Atif, A. Rashid, and C. A. Edwin, "Profiling players using real-world datasets: clustering the data and correlating the results with the big-five personality traits," *IEEE TAC*, 2017.
[33] S. Lesmana, O. Ariwana, R. P. Halim, and A. A. Gunawan, "Behavior correlation between games in first-person shooter genre based on personality traits," *Procedia Computer Science*, 2021.
[34] D. L. King, A. M. Russell, P. H. Delfabbro, and D. Polisena, "Fortnite microtransaction spending was associated with peers' purchasing behaviors but not gaming disorder symptoms," *Addictive Behaviors*, 2020.
[35] S. M. F. Gillani, "Evaluation of games monetization approaches: A case study on players unknown's battlegrounds (PUBG)," MSc. Thesis, 2021.
[36] T. Ide and H. Hosobe, "Supporting online game players by the visualization of personalities and skills based on in-game statistics." in *Int. J. Conf. Computer Vision, Imaging and Computer Graphics*, 2023.
[37] M. Krcmar, R. McGloin, K. M. Farrar, and S. Scott Li, "'what is my call of duty?': Exploring the importance of player experience in a first-person shooter video game," *Journal of Gaming & Virtual Worlds*, 2018.
[38] M. Schaekermann *et al.*, "Curiously motivated: profiling curiosity with self-reports and behaviour metrics in the game "Destiny"," in *ACM Ann. Symp. on Computer-Human Interaction in Play (CHI PLAY)*, 2017.
[39] R. Heartfield and G. Loukas, "A taxonomy of attacks and a survey of defence mechanisms for semantic social engineering attacks," *ACM Computing Surveys (CSUR)*, 2015.
[40] Repository of our paper: https://github.com/hihey54/cog24_aia, 2024.
[41] steamcommunity.com/app/1517290/discussions/0/3191360735183020391/.
[42] steamcommunity.com/app/730/discussions/0/2592234299558984436/.
[43] steamcommunity.com/app/570/discussions/0/3818529263636669669.
[44] steamcommunity.com/app/730/discussions/0/6993585599474786899/.
[45] A. J. Bisberg, J. Jiang, Y. Zeng, E. Chen, and E. Ferrara, "The gift that keeps on giving: Generosity is contagious in multiplayer online games," *ACM Conference on Human Factors in Computing Systems (CHI)*, 2022.
[46] M. Choi, G. Ko, and S. K. Cha, "Botscreen: Trust everybody, but cut the aimbots yourself," in *USENIX SEC*, 2023.