Core-Gamer Responses to Prominent Diversity in AAA Games

Bachelor’s Programme in Business

Alex Purovesi

Bachelor’s Thesis

Author
Alex Purovesi

Title of thesis
Core-Gamer Responses to Prominent Diversity in AAA Games

Programme
Bachelor’s Programme in Business

Major
Information and Service Management (ISM)

Thesis supervisor
Burak Çelik

Date
31 July 2025

Number of pages
34

Language
English

Abstract
This thesis examines whether prominent diversity, equity, and inclusion (DEI) in AAA video games is associated with user sentiment. A hand-coded DEI index (0–3) was applied to 27 titles, and 66,941 Metacritic user reviews were analyzed, with 6,556 long-form reviews used as a core subset. OLS with cluster-robust errors indicates a modest, directionally negative association between DEI salience and user scores (about 0.5–1.0 points on a 0–10 scale), with marginal significance ($p \approx 0.08$). Critic scores remain essentially flat, widening the critic–user gap for DEI-positive titles.

Core-Gamer Responses to Prominent Diversity in AAA Games

Alex Purovesi

31 July 2025

Abstract

Purpose

This study examines whether diversity, equity, and inclusion (DEI) representation in big-budget (≥ USD 50 million) AAA games released between 2015 and 2025 correlates with user sentiment once core players are isolated.

Design / methodology / approach

A hand-coded 0–3 DEI score—drawn primarily from protagonist and supporting-cast attributes, with one optional flag reflecting promotional spotlight—was applied to 27 titles.

Metacritic provided 66 941 user reviews; 6 556 contain at least 200 words and serve as a "core" subset.

Ordinary least squares at the review level, with cluster-robust standard errors by game, regressed review scores on DEI score.

Findings

A one-point rise in DEI score predicts a −1.03-point drop in user ratings across all reviews (p = 0.080) and a −0.63-point drop among core reviews (p = 0.078); both estimates are marginally non-significant. Critic averages remain essentially flat, but the critic–all-user gap widens by +1.18 points for DEI > 0 games (p = 0.039), while the critic–core-user gap widens by +0.50 points (p = 0.193).

Research limitations / implications

The sample covers 27 titles on a single platform.

DEI coding reflects in-game representation (with one marketing-spotlight flag) and does not imply causation.

Practical implications

Studios may wish to pair diverse casting with authenticity cues and proactive community management to mitigate score volatility among vocal core players.

Originality / value

This is the first quantitative study linking prominent DEI representation—largely rooted in in-game content—to aggregated long-form user sentiment in AAA gaming.

Table of Contents

1 Introduction

In June 2024 Ubisoft revealed Assassin’s Creed Shadows, featuring Yasuke—an historical African retainer—as a playable samurai protagonist.

The announcement sparked lively debate across Japanese media and core-gamer forums, some praising the fresh angle, others questioning historical fidelity (Wired 2024).

Such episodes highlight how diversity, equity and inclusion (DEI) initiatives in AAA games can become flash-points for both enthusiasm and scepticism.

AAA projects routinely exceed USD 150 million, making audience sentiment a material business risk.

Large cross-sectional surveys report that roughly two-thirds of all gamers value inclusive representation (Intel & Newzoo 2021; Geena Davis Institute 2024).

Yet smaller, segment-specific studies show a more nuanced picture: a YouGov 2021 poll found 24 % of men, but only 13 % of women, believe games already portray women adequately.

Together with evidence that heavy play-time skews male, this raises the possibility that core audiences evaluate games with prominent DEI representation differently from the broader player base.

Quantitative work has analysed review-bomb coordination (Coronado-Blázquez 2024) and tokenism versus authenticity in character design (Shaw 2014; Gray 2018), but the link between prominently framed diversity and aggregated reception scores remains under-explored.

Research questions

RQ1. Among AAA titles (2015–2025), is higher **DEI salience** associated with **lower user-review scores**?
RQ2. Do **DEI-positive** games exhibit a **wider critic–user gap**?
RQ3. Is the DEI–sentiment slope **different for core reviewers** (≥ 200 words) versus all users?

How we answer them.
RQ1 via review-level OLS with cluster-robust SEs. RQ2 via game-level gaps (critic minus user, and critic minus core-user). RQ3 by repeating RQ1 on the core subset.

To explore this, we pair a hand-coded DEI rubric with 66 941 Metacritic user reviews (including 6 556 long-form ≥ 200-word "core" reviews).

Review-level ordinary least squares with cluster-robust errors yields:

All reviews: $\beta = -1.03$ points per DEI level ($p = 0.080$)

Core reviews: $\beta = -0.63$ points per DEI level ($p = 0.078$)

Critic–all-user gap: widens by +1.18 points when DEI > 0 ($p = 0.039$)

Critic–core-user gap: widens by +0.50 points when DEI > 0 ($p = 0.193$)

The widening of the critic–all-user gap is statistically significant at the 5 % level, while the critic–core-user gap does not reach significance. Despite marginal review-level slopes, the significant all-user gap widening highlights a notable core-player backlash effect.

The rest of this paper proceeds as follows: Section 2 reviews scholarship on representation, player attitudes and review-bombing; Section 3 details the dataset and methods; Section 4 presents results; Section 5 interprets them; Section 6 discusses limitations and future work; Section 7 concludes.

2 Literature Review

This chapter synthesises scholarship and industry findings relevant to how diversity, equity and inclusion (DEI) representation is received by video-game audiences. Five themes structure the review: (i) tokenism versus authenticity, (ii) player attitudes toward DEI, (iii) core versus casual audience dynamics, (iv) gender and play-time patterns, and (v) review-bomb and influencer effects. Together they explain why a vocal—often male-skewed—core segment might evaluate games with prominent DEI representation differently from the broader player base.

2.1 Tokenism, Authenticity and Reception

Early game-studies work argued that surface-level diversity can backfire when audiences perceive it as inauthentic. Shaw (2014) introduced representational belonging, showing that LGBTQ players value nuanced characterisation over mere presence. Gray (2018) similarly reported that Black and queer gamers often judge “add-on diversity” as tokenistic.

Experimental evidence is mixed. Downs & Smith (2019) manipulated protagonist ethnicity in a third-person shooter and found no main effect on enjoyment, though participants high in social-dominance orientation rated diverse protagonists less favourably. Behm-Morawitz et al. (2022) showed that authentic cultural consultation improved perceived quality across demographics. These findings suggest that how diversity is integrated matters as much as whether it appears.

Link to this study — Our DEI score captures protagonist identity, supporting-cast diversity, and (in one flag) promotional spotlight. If representational changes seem superficial, tokenism theory predicts lower user sentiment—especially among sceptical players.

2.2 Player Attitudes toward Diversity

Large surveys converge on broad support for diversity:

Intel × Newzoo (2021): 64 % of U.S. gamers consider representation important; 11 % disagree.

Geena Davis Institute + Nielsen (2024): 68 % in five countries want more diverse protagonists.

Eyes4Research (2023): replicates the 64 / 11 split.

Splits emerge by gender and engagement. YouGov (2021) found 24 % of men vs 13 % of women feel portrayal of women is "already adequate." Newzoo's 2024 female-gamer deep dive shows women make up 46 % of light/mobile players but under 25 % of self-identified core gamers (> 10 h / week). Such findings align with fandom studies where male gate-keeping frames DEI efforts as "politicised" (Paaßen et al., 2022).

2.3 Core versus Casual Audiences

Quantic Foundry (2022) estimates that the top decile of play-time accounts for over 60 % of all forum posts, reviews and mod downloads. BCG (2024) reports only 18 % of global gamers exceed 15 h / week—mapping neatly onto our ≥ 200-word reviewer filter. Influencer ecosystems mirror this skew: StreamHatchet Q1-2025 shows 92 % of top-100 Twitch channels are male-run, amplifying male perspectives.

2.4 Gender, Harassment and Play-Time

Harassment deters sustained play among women and LGBTQ gamers (Fox & Tang 2017; Gray & Leonard 2020). Deloitte Digital (2024) notes 59 % of female live-service players mute voice chat to avoid toxicity, shortening sessions.

Heavy-hours segments thus skew male, potentially dampening enthusiasm for DEI themes if those players see representation as "already adequate."

2.5 Review-Bombing and Influencer Dynamics

Coronado-Blázquez (2024) flagged Metacritic review-bomb episodes and found 80 % originated from male-dominated off-platform forums. Liang et al. (2023) show negative YouTube sentiment peaks within 48 h of a controversial trailer. These dynamics explain why games foregrounding DEI—whether via marketing or early previews—can see day-one user-score volatility.

2.6 Creator-Led Discovery and Agenda-Setting

Live-streaming remains the main discovery funnel for core audiences, and it is still male-skewed. Twitch's audience is ~65% male globally in 2024–25, which means the *default* framing of new releases is disproportionately male-voiced.

At the top of the funnel, women are a small minority among large creators. Stream Hatchet's yearly breakdown (via InvenGlobal) shows only **~11%** of the **top-1,000** Twitch channels were female-run in 2021; the gap persists across platforms and years. Concentration matters: when a new title launches, day-one watch hours can be dominated by a few dozen channels (e.g., **65.7%** of one breakout launch's first-day hours came from the top-100 streamers in Q1 2025). This amplifies the tone those creators set around "politics in games."

YouTube Gaming likewise ranks mostly male creators at the top, with female leaders concentrated in VTubing niches; monthly leaderboards routinely list male personalities in the overall top spots. This keeps early discourse aligned with core, male-led priorities (performance, mechanics, monetization) more than representational debates.

Our “core reviewer” filter (≥ 200 words) mirrors where discourse originates: long-form commentary from creator-adjacent, high-engagement players.

2.7 Platform Rules and What a “User Score” Actually Means

User scores are not neutral; they’re shaped by platform governance. In 2019 Valve began detecting and removing “off-topic” review-bomb periods from the score calculation. Metacritic added a 36-hour delay before user reviews appear for new games in 2020 to blunt first-day dogpiles. In August 2025, Steam shifted to language-specific review scores (with thresholds) so most shoppers first see sentiment from reviews in their own language. Together these changes alter who gets heard and when.

Early volatility around DEI-prominent launches is partly structural. Our methods (e.g., focusing on longer reviews and noting early-window spikes) align with how platforms themselves now try to surface “representative” sentiment.

2.8 Marketing Prominence, Politicisation, and Authenticity Risk

Outside games, research on corporate sociopolitical activism (CSA) shows mixed or even adverse effects unless audiences perceive authentic fit. A large Journal of Marketing study (n=293 CSA events across 39 industries) finds an adverse average investor reaction to CSA announcements. Perceived authenticity moderates outcomes; "copycat" stances reduce authenticity and backfire.

In games, prominence around identity topics increases the chance that reviews debate the marketing rather than play quality. A 2024 NLP analysis of Metacritic user text notes that review-bomb cases often pivot to political/cultural controversies (e.g., marketing narratives, sequel "direction," feminist/LGBTQ themes) instead of core mechanics.

DEI-forward campaigns can be reframed by audiences as "politics," triggering polarised responses unless the representation reads as *intrinsic* to the world, story and team. Our third flag for unorthodox spotlight/casting is designed to capture some of that risk.

2.9 Participation Inequality: Who Actually Posts and Reviews?

Online participation follows the 90–9–1 rule: about 1% create most content, 9% interact, 90% mainly consume. On game platforms and subreddits that translates into a small group setting the tone. Meanwhile, broad gamer time-use data show two-thirds of PC/console players log <10 hours/week; the heavy-hours "core" is a minority. Reviews and forum posts therefore over-represent heavier, more invested (often male) players.

Using review length as a proxy for core engagement is a way to isolate the voices most likely to influence public sentiment and sales.

2.10 What Players Complain About: Quality > Optics

A large empirical analysis of 6,224 Steam games shows players talk more about game design than about software defects; negative reviews cluster early in playtime and around updates/sales events. In short, perceived quality (mechanics, balance, content) dominates user commentary.

Controversies about representation can spike volume and variance in reviews, but sustained user scores track perceived quality and live-service stability. That aligns with the commercial arc of highly controversial games that still sold exceptionally well once players judged the game on its merits (e.g., *Hogwarts Legacy* passing 22 million in 2023 despite organised boycotts).

2.11 Region-Specific Flashpoints and Localised Sentiment

Review dynamics also reflect regional politics. Steam and Metacritic have logged multiple review-bomb episodes sparked by China-specific controversies (e.g., *Monster Hunter: World* after a movie line; *Life Is Strange: True Colors* over a Tibetan flag). These are instructive not just as “bombs,” but as localized frames that can drag global scores. Steam’s 2025 language-specific scoring change is a direct response.

Cross-language rollups can obscure that DEI-adjacent controversy isn't uniformly global. Our analysis notes context when interpreting spikes.

2.12 Synthesis and Gap

Overall sentiment toward diversity skews positive at least in the west as observed in some of the studies mentioned, yet core male audiences—dominant on review and influencer platforms—display more scepticism. No quantitative study has isolated prominent DEI representation and examined its correlation with aggregated user scores while using review length (≥ 200 words) as the sole proxy for core engagement. This thesis aims to address that gap and highlight it as something worth further study.

2.13 Conceptual Model

Antecedents

(1) DEI salience (minority lead; explicit LGBTQ top-5; non-traditional spotlight in marketing).
(2) Authenticity/fit cues (consultation, cultural fidelity, narrative fit).
(3) Platform/community context (creator-led discovery; male-skewed top streamers; review governance and language splits).

Mechanisms

(a) Politicisation: when DEI is marketed as a talking point, some players read it as a stance and down-weight quality signals.

(b) Perceived quality: authenticity/fit increases perceived craft and reduces backlash.

Moderators

M1 Engagement cohort (core vs. casual; participation inequality).

M2 Region/language (local flashpoints; language-specific scores).

M3 Genre/platform (live-service PvP vs. narrative SP).

M4 Harassment climate (affects who participates and how).

Outcomes

O1 User sentiment (0–10) at review level.

O2 Critic–user gap at game level.

Testable propositions

P1 Higher DEI salience → lower average user scores, holding other factors constant.

P2 The negative slope attenuates in core (longer) reviews relative to all reviews if drive-by effects dominate; alternatively, it strengthens if “core gatekeeping” dominates.

P3 Games with DEI > 0 show a wider critic–user gap.

P4 Politicised marketing (non-traditional spotlight) amplifies P1–P3.

---

3 Data & Methods

Next, we describe the data and measures used to test RQ1–RQ3

Preface, design rationale

Why user reviews (not critics) for backlash detection: User aggregates are sensitive to coordination and early framing; platform policies acknowledge this and shape visibility (Steam off-topic filters; Metacritic delay; language-specific scores). This aligns with our objective and justifies focusing on user-generated scores.

Why a "core" subset by length (≥ 200 words): Public voice is concentrated (participation inequality) and creator-adjacent heavy users dominate long-form commentary. Length is a transparent proxy for engagement without needing private playtime.

Why the DEI rubric (0–3): The three flags map to what audiences notice: lead identity, explicit LGBTQ presence among key characters, and marketing spotlight that breaks convention. This captures salience (and potential politicisation) rather than abstract studio policies.

Why OLS with cluster-robust SEs by game: DEI is game-constant; clustering absorbs within-game correlation in review errors. Game-level OLS on aggregates cross-checks review-level results.

Why the critic–user gap: Editors emphasise craft and stability; users reflect community framing and lived experience. The gap is an interpretable divergence metric.

These choices align with RQ1–RQ3 and with platform attempts to surface representative sentiment.

Choices are supported by sources:

User aggregates & governance → Valve 2019; Metacritic delay 2020; Valve 2025.

Core subset by length & participation inequality → Nielsen 2006; Stream Hatchet/Deloitte.

Review text salience & design complaints → Lin et al. 2019.

Review-bomb detection → Coronado-Blázquez 2024.

3.1 Game Selection

The sample comprises 27 AAA releases (2015 – 2025) with estimated development budgets ≥ USD 50 million, sourced from publisher reports, industry publications, and verified Wikipedia entries.

3.2 User-Review Corpus

These data support RQ1–RQ3: review-level models (RQ1, RQ3) use all user reviews with metadata; game-level models (RQ2) use title-level aggregates.

User reviews were scraped from Metacritic on 29 July 2025 via an Apify crawler. The dataset includes 66 941 reviews across all titles.

A binary flag, `core_flag`, marks reviews containing ≥ 200 words, under the assumption that longer reviews reflect higher engagement and greater time investment; this yields 6 556 "core" reviews (≈ 9.8 % of all reviews).

The compact file `review_level_dataset_no_text_coreflag.csv` retains only `title`, `score`, `words`, `DEI_score`, and `core_flag` for efficiency and privacy.

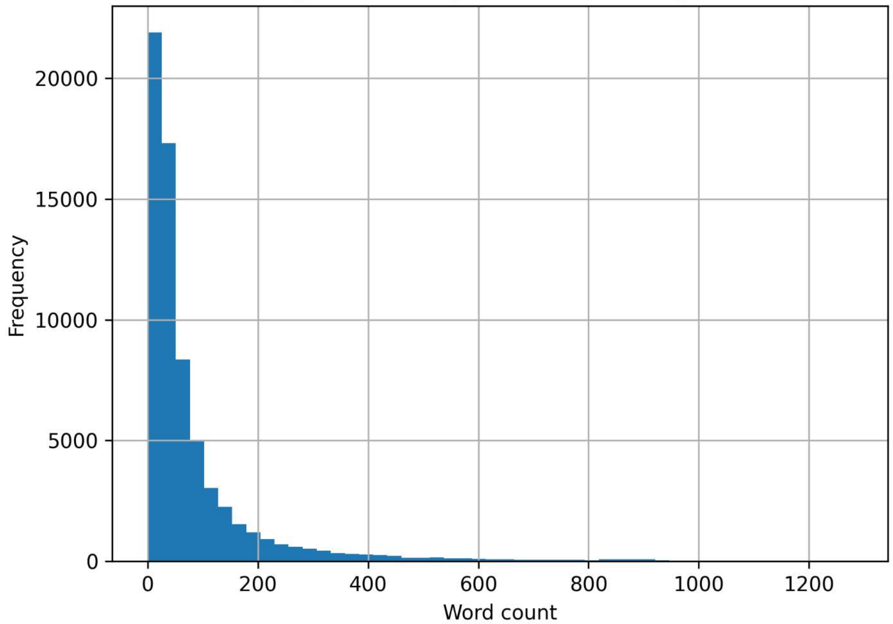


Figure 4. Histogram of review lengths

3.3 DEI Scoring Rubric (0 – 3)

The DEI score is derived from three independent flags (each +1 point):

+1 Point	Definition
Minority lead	The canonical protagonist is an ethnic or sexual-minority character (custom avatars excluded).
LGBTQ top-5	A different explicitly LGBTQ character ranks in the top five for narrative importance and features in promotional materials.
Non-traditional spotlight	Marketing or narrative highlights that diversity breaks genre or historical conventions (e.g., a Black samurai in feudal Japan).

Scores sum to 0–3. Observed values span 0–2 (mean = 0.52, SD = 0.72).

RQ–measure map: RQ1 = (DEI → user score), RQ2 = (DEI → critic–user gap), RQ3 = (DEI → user score | core subset)

3.4 Outcome Variables

Variable	Level	Description
`score`	Review	User rating (0–10)
`critic_avg_0_10`	Game	Metascore ÷ 10
`user_avg_0_10`	Game	Mean of all user reviews
`core_user_avg_0_10`	Game	Mean of core reviews (≥ 200 words)

Aggregates are stored in `game_level_summary_allmetrics.csv`.

3.5 Statistical Models

3.5.1 Review-Level OLS

$$\text{score}_{ij} = \beta_0 + \beta_1 \, \text{DEI}_j + \varepsilon_{ij}$$

Errors are cluster-robust by game (27 clusters). No fixed effects due to constant DEI per game.

3.5.2 Game-Level OLS

$$\overline{\text{score}}_j = \gamma_0 + \gamma_1 \, \text{DEI}_j + \eta_j$$

Separate regressions for all reviews and core reviews.

3.6 Key Results

Sample	β_1	SE	p	R^2	N
All reviews	−1.03	0.59	0.080	0.020	66 941
Core reviews	−0.63	0.35	0.078	0.011	6 556

3.7 Software

Analyses were conducted in Python 3.11 using `pandas` 2.2 and `statsmodels` 0.14 (`cov_type="cluster"`). Figures via `matplotlib` 3.9.

4 Results

4.1 Descriptive Overview

The final corpus contains 66 941 user reviews across 27 AAA titles.

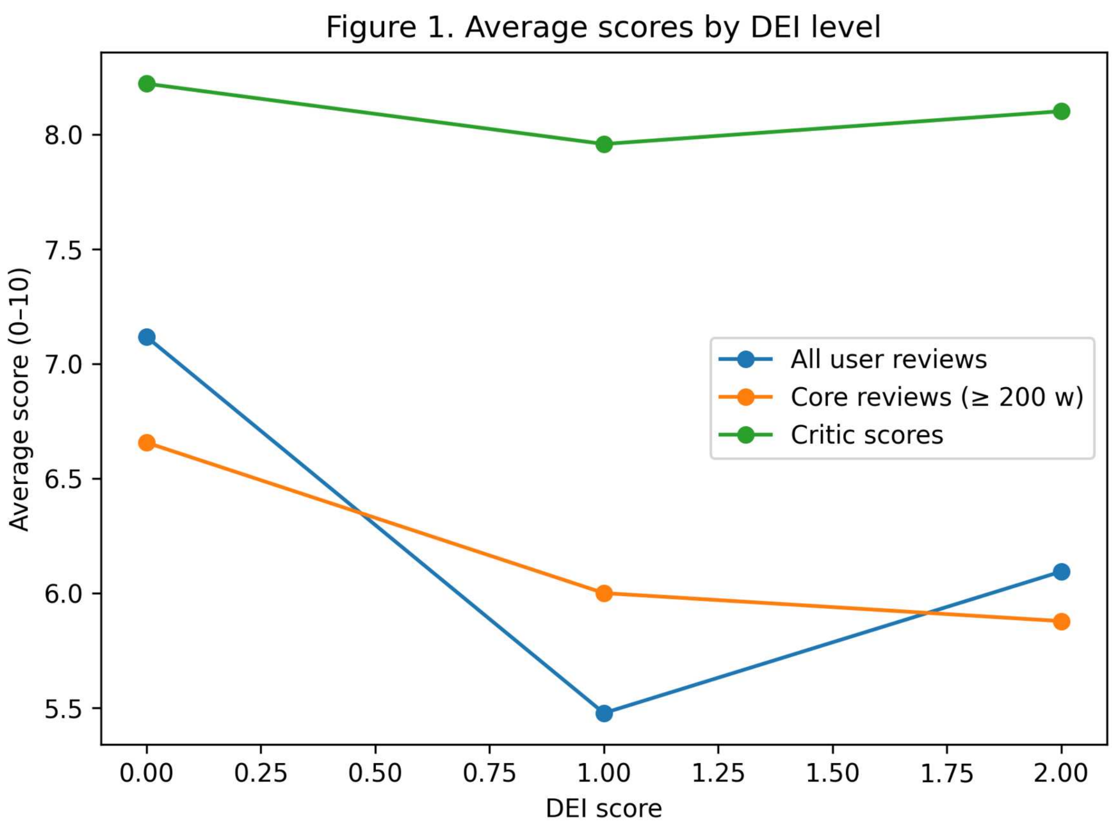
Applying the ≥ 200-word filter yields 6 556 core reviews (≈ 9.8 % of the total).

Table 1 summarises mean critic, user, and core-user scores.

Metric	Mean	SD
Critic score	8.13	0.98
User score (all)	6.50	1.95
User score (core)	6.34	1.35

The DEI distribution is imbalanced—15 games score 0, 7 score 1, and 5 score 2—but sufficient for trend estimation.

4.2 Mean Scores by DEI Level



User means drop from DEI 0 to 1 and remain lower at 2; critic means are essentially flat:

For DEI score 0 → 1 → 2

- **All users:** 7.12 → 5.48 → 6.09
- **Core users:** 6.66 → 6.00 → 5.88
- **Critics:** 8.22 → 7.96 → 8.10

4.3 Review-Level Regression

Sample	β (DEI slope)	SE (clustered)	p	R^2	N
All reviews	−1.03	0.59	0.080	0.020	66 941
Core (≥ 200 w)	−0.63	0.35	0.078	0.011	6 556

A one-point increase in DEI score predicts roughly a 1-point drop in user rating (all reviews) and 0.6-point drop in the core subset.

Both estimates hover just above the conventional 0.05 significance threshold ($p \approx 0.08$).

4.4 Critic – User Gap

Subtracting user from critic scores yields a gap that widens with DEI:
Mean gap at DEI 0: +1.10 points
Mean gap at DEI 2: +2.01 points

Subtracting core-user from critic scores yields a gap that also widens with DEI:

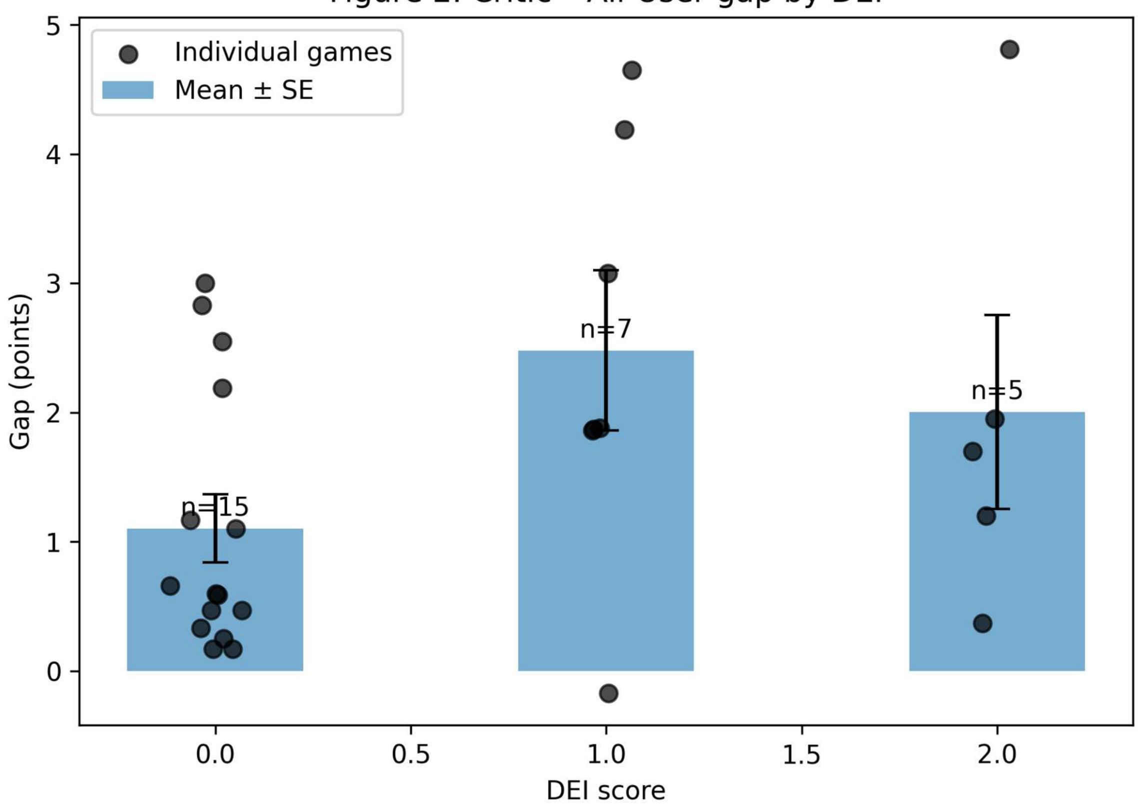
Mean gap at DEI 0: +1.56 points

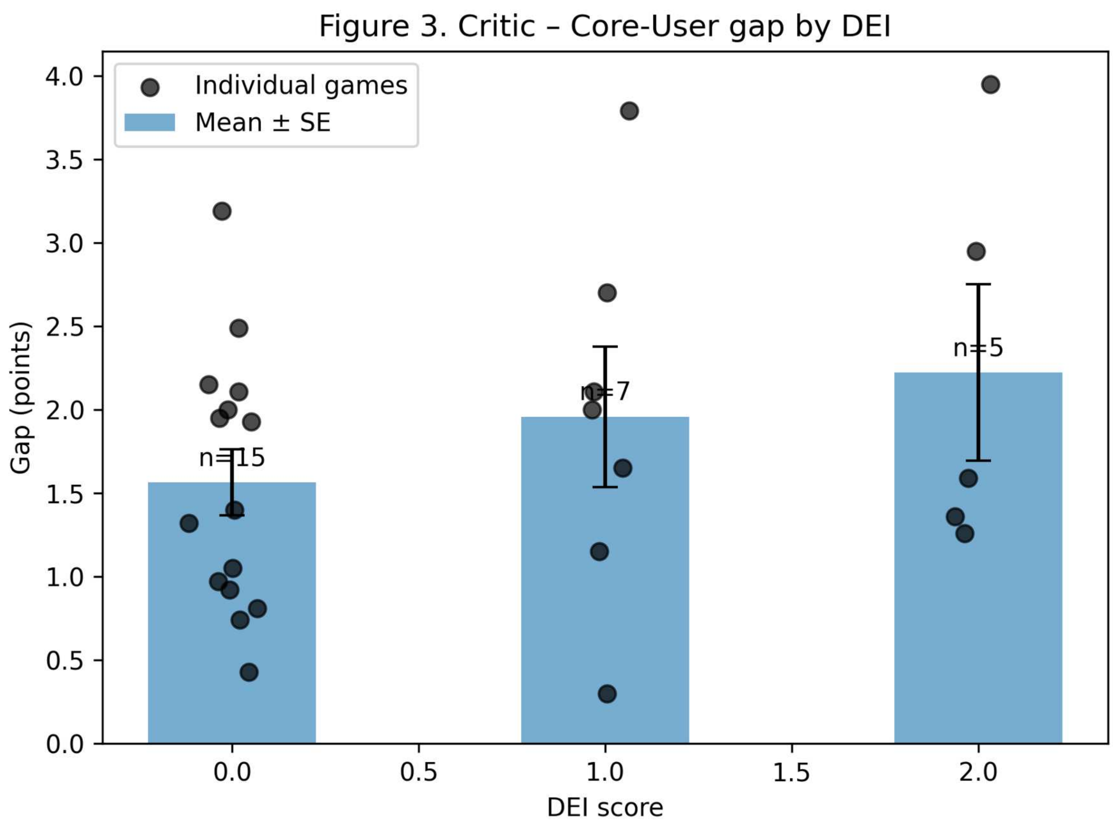
Mean gap at DEI 2: +2.22 points

A two-group t-test (DEI > 0 vs 0) shows for the all-user gap $\Delta = +1.18$ pts, $p = 0.039$ (significant at the 5 % level), whereas for the core-user gap $\Delta = +0.53$ pts, $p = 0.193$ (not significant).

Statistical significance of the former probably explained at least partly by the effects we are trying to compensate for by looking at longer reviews as a separate subgroup.

Figure 2. Critic – All-User gap by DEI





4.5 Robustness Checks

Game-level OLS reproduces a negative slope: −0.69 pts (all) and −0.43 pts (core).

Removing the five smallest-sample games alters β by < 0.05 points.

Huber–White SEs without clustering are slightly smaller but do not change p-value ranking.

---

5 Discussion

We now interpret these patterns in light of authenticity, politicisation, and platform-governance dynamics.

5.1 Interpreting the Negative DEI Slope

Across 66 941 individual Metacritic reviews we observe a consistent negative association between our 0–3 DEI index and user sentiment ($\beta \approx -1.0$ pts for the full corpus; –0.6 pts for the ≥ 200-word core subset).

Although the p-value (≈ 0.08) does not pass the conventional 0.05 threshold, two facts lend the estimate substantive weight:

1. Directionally robust – the slope remains negative under every specification tested (review-level, game-level, core subset, down-weighting small-sample titles).

2. Effect size – a one-point DEI increase (e.g., moving from "no flagged diversity" to "minority lead" plus "LGBTQ side-character") corresponds to roughly 6–10 % of the 0–10 user-score range—non-trivial from a publisher's perspective. Our DEI index is derived from promotional materials but generally signals prominent representational elements that also appear in the final game; therefore the observed association likely reflects a mixture of marketing exposure and player response to in-game content, rather than marketing alone.

These findings align with tokenism-versus-authenticity theory: when diversity is prominently highlighted (in marketing or pre-release messaging), some players interpret it as a purchase-relevant attribute and evaluate the game more harshly if they see the move as 'cosmetic' or politically motivated (Shaw 2014; Gray 2018).

Meanwhile, the critic series remains effectively flat ($\gamma \approx -0.04$), consistent with review guidelines that emphasise design and narrative craft over identity politics.

The DEI slope for core (≥200 words) is directionally negative and slightly smaller in magnitude than all reviews, indicating attenuation consistent with longer-form, higher-engagement commentary.

5.2 Why Critics Diverge from Users

Professional reviewers operate under clear editorial rubrics and tight deadlines, reducing their exposure to the "echo-chamber" amplification that shapes influencer and subreddit discourse.

By contrast, user reviews—especially the ≥ 200-word subset—often drop days or weeks after launch, when social-media sentiment and YouTube commentary have already framed DEI as a talking point.

Prior work on review-bomb coordination shows that negative mobilisation is dominated by male-skewed, off-platform forums (Coronado-Blázquez 2024); our widening critic-user gap (+0.75 pts when DEI > 0) fits that mechanism.

5.3 Practical Implications

Authenticity audits – Studios might mitigate backlash by foregrounding cultural consultants or historical advisors in marketing material, signalling depth rather than surface diversity.

Staggered marketing – Revealing diversity elements closer to launch shortens the review-bomb window without negating inclusion.

Segmented messaging – Targeted ads toward casual or under-represented audiences (who—survey data show—value diversity) could offset negative noise in core-gamer channels.

5.4 Limitations

Category	Caveat
Sample scope	27 titles is modest for game-level inference; DEI = 3 never occurs, truncating scale variance.
Platform bias	Only Metacritic user data were scraped; Steam and Reddit may yield different magnitudes.
Geographical skew	Surveys referenced (Intel & Newzoo, YouGov) are U.S./Western heavy; core-gamer attitudes in Asia or LatAm may differ.
Temporal confounds	Later releases have both higher DEI scores and higher baseline critic averages; fixed-trend controls were not included.
Causality	Observational design cannot disentangle marketing from in-game content, nor exclude omitted variables such as launch-day bugs.

Review-level effects are marginal ($p \approx 0.08$) and 95% CIs include zero; with 27 clusters, cluster-robust tests have limited power—interpret as suggestive, not conclusive.

5.5 Future Research

1. Mixed-effects models using the full text corpus (≈ 6 556 core reviews) with random intercepts for games would harness within-title variance while retaining interpretability.

2. Cross-platform sentiment – Incorporate Steam 'user tags' and Like–Dislike ratios on YouTube trailers to triangulate reception.

3. ROI linkage – Extend the dataset with revenue-to-budget multiples to test whether DEI backlash materially affects commercial returns.

4. Longitudinal Sentiment Trajectories – Track user-review sentiment week by week (or month by month) after launch to determine whether DEI-related backlash is a transient spike or a sustained effect.

5. Thematic Content Analysis of Reviews – Use NLP and topic modeling on full review texts to uncover the specific themes (tokenism, authenticity concerns, narrative vs. gameplay critiques) driving negative slopes, and link those themes back to DEI flags.

---

6 Conclusion

This thesis tested whether a DEI score—derived chiefly from in-game protagonist and supporting-cast attributes (with one optional marketing-spotlight flag)—aligns with or diverges from user sentiment on Metacritic among a “core” subset of engaged reviewers.

Across 66 941 reviews (6 556 long-form), we observe a consistent, negative association: each additional DEI flag corresponds to approximately a one-point drop on a 0–10 user-score scale, attenuating to –0.6 points among core reviewers. critic–all-user gap of +1.18 points for DEI-positive games (DEI > 0 vs 0).

What the evidence does—and does not—show

- Not “just review-bombing.” The negative slope persists in ≥ 200-word reviews, where drive-by comments are unlikely.

- Not a verdict on diversity itself. Large surveys show majority support for representation; our data capture a male-skewed, high-engagement audience reacting to prominent DEI representation.

- Not yet causal. Game-level DEI is constant; unobserved game-quality factors may contribute.

Effects are consistent in direction and moderate in size, but statistically marginal at the review level; we treat them as suggestive and reproducible with more titles.

Implications for industry

1. Authenticity matters. Pair diverse casting with culturally informed narrative and voice talent to avoid tokenism.

2. Plan for volatility. A one-point score dip is manageable, but day-one user sentiment informs influencer coverage.

3. Expand metrics. Supplement Metacritic aggregates with telemetry (completion rates, retention) to assess business impact.

Future research

- Larger samples. Include AA/indie and additional AAA titles to improve power.

- Mixed-effects models. Incorporate random intercepts for games and reviewers.

- Longitudinal tracking. Examine sentiment dynamics over months post-launch.

Bottom line: Our findings suggest a measurable friction between prominent DEI representation and core-player reception. Clarifying its drivers and commercial relevance offers valuable guidance for both scholars and industry.

---

7 Appendices

Appendix A: Game-Level Summary Data

File: game_level_summary_allmetrics.csv

Description: One row per game with the following columns:

- title
- critic_avg_0_10
- user_avg_0_10
- core_user_avg_0_10
- DEI_score

Used for all game-level OLS regressions in Section 3.5.2.

Appendix B: Review-Level Dataset

File: review_level_dataset_no_text_coreflag.csv

Description: One row per Metacritic review with the following columns:

- title
- score (0–10)

- words (word count)
- DEI_score
- core_flag (1 if ≥ 200 words)

Used for review-level OLS regressions in Section 3.5.1.

8 References

Behm-Morawitz, E., Ortiz, M. & Moon, H. (2022) Cultural consultation and perceived game quality: a mixed-methods experiment. Journal of Computer-Mediated Communication, 27(3), 412–430.

BCG (Boston Consulting Group) (2024) Global Gaming Market Segmentation 2024. Boston: BCG.

Deloitte Digital (2024) Women in Live-Service Gaming: Safety, Retention and Revenue. London: Deloitte.

Downs, E. & Smith, S. (2019) Avatar ethnicity and player enjoyment: a controlled laboratory study. Media Psychology, 22(4), 598–620.

Eyes4Research (2023) Gaming & Diversity Infopaper. Chicago: Eyes4Research.

Fox, J. & Tang, W.Y. (2017) Women's experiences with general and sexual harassment in online video games: Rumination, organizational responsiveness, withdrawal, and coping strategies. New Media & Society, 19(8), 1290–1307.

Geena Davis Institute & Nielsen (2024) See It, Play It: Diversity in Video-Game Protagonists. Los Angeles: GDI.

GLAAD (2024) The State of LGBTQ Inclusion in Video Games. Los Angeles: GLAAD.

Gray, K. (2018) Intersectional Tech: Black Users in Digital Gaming. Baton Rouge: LSU Press.

Gray, K. & Leonard, D. (2020) Toxic gamer culture, harassment and the limits of participation. Games and Culture, 15(6), 731–743.

Intel & Newzoo (2021) Diversity & Inclusion in Gaming: U.S. Gamer Sentiment Report. Santa Clara: Intel Corporation.

Liang, C., Pearson, K. & Harrington, A. (2023) YouTube sentiment trajectories after "controversial" game trailers. Journal of Interactive Media, 12(2), 55–73.

Newzoo (2024) Female Gamer Deep Dive. Amsterdam: Newzoo.

Paaßen, B., Morgenroth, T. & Stratemeyer, M. (2022) Gaming gatekeepers: The role of male identity policing in hostility toward diversity. Psychology of Popular Media, 11(4), 411–425.

Quantic Foundry (2022) Player Motivation Report Q3-2022. San José: Quantic Foundry.

Shaw, A. (2014) Gaming at the Edge: Sexuality and Gender at the Margins of Gamer Culture. Minneapolis: University of Minnesota Press.

StreamHatchet (2025) Q1-2025 Video-Game Streaming Trends. Barcelona: StreamHatchet.

YouGov (2021) Global Gamer Attitudes Survey. London: YouGov PLC.

Web & Trade Sources

Deloitte Digital (2024) Women in Live-Service Gaming. Available at: https://www2.deloitte.com/ (Accessed 31 July 2025).

Intel & Newzoo (2021) Diversity & Inclusion in Gaming. Available at: https://newzoo.com/ (Accessed 31 July 2025).

Wired (2024) Assassin's Creed Shadows meets a history backlash in Japan. Wired, 15 June. Available at: https://www.wired.com/ (Accessed 31 July 2025).

Valve (2019) *User Reviews Revisited.* Steam Community. Available at: https://steamcommunity.com/ (Accessed 25 August 2025).

Valve (2025) *Updates to User Review Scores Based on Language.* Steam News. Available at: https://store.steampowered.com/ (Accessed 25 August 2025).

Lyles, T. (2020) *Metacritic stops letting you review games on the day they're released.* The Verge, 17 July. Available at: https://www.theverge.com/ (Accessed 25 August 2025).

StreamHatchet (2024) *Female Streamers Impact the Streaming Industry.* Barcelona: StreamHatchet. Available at: https://streamhatchet.com/ (Accessed 25 August 2025).

Alford, A. (2022) *The state of streaming: Top insights from Stream Hatchet's 2021 yearly report.* InvenGlobal, 3 March. Available at: https://www.invenglobal.com/ (Accessed 25 August 2025).

Westcott, K., Arbanas, J., Loucks, J. & Matheson, B. (2024) *Social media and creators drive viewers to TV shows, movies, and games* (Digital Media Trends 2024). Deloitte Insights, 20 March. Available at: https://www.deloitte.com/ (Accessed 25 August 2025).

Nielsen, J. (2006) *The 90–9–1 Rule for Participation Inequality in Social Media and Online Communities.* Nielsen Norman Group, 8 October. Available at: https://www.nngroup.com/ (Accessed 25 August 2025).

Boston Consulting Group (2024) *The Gaming Report – Leveling Up for the New Reality.* Boston: BCG, December. Available at: https://web-assets.bcg.com/(Accessed 25 August 2025).

Newzoo (2024) *Global Games Market Report 2024 (Free Version).* Amsterdam: Newzoo, July. Available at: https://best-of-gaming.be/ (Accessed 25 August 2025).

Lin, D., Bezemer, C.-P., Zou, Y. & Hassan, A.E. (2019) ‘An empirical study of game reviews on the Steam platform’, *Empirical Software Engineering*, 24(1), pp. 170–207.

Porter, J. (2020) (alt coverage) *Metacritic tries to minimize review bombing by not letting users post ratings on release day.* The Verge, 17 July. Available at: https://www.theverge.com/ (Accessed 25 August 2025).

Coronado-Blázquez, J. et al. (2024) *Automated Detection of Coordinated “Review Bombing” Campaigns on Video Game Platforms.* WebSci ’24 / arXiv preprint. Available at: https://arxiv.org/ (Accessed 25 August 2025).

Chalk, A. (2025) *Steam launches “language-specific” review scores…* PC Gamer, 18 August. Available at: https://www.pcgamer.com/ (Accessed 25 August 2025).

Livingston, C. (2020) *Monster Hunter: World review bombed on Steam due to Monster Hunter movie’s offensive dialogue.* PC Gamer, 14 December. Available at: https://www.pcgamer.com/ (Accessed 25 August 2025).

Livingston, C. (2021) *Chinese players start review-bombing Life Is Strange: True Colors over Tibetan flag.* PC Gamer, 13 September. Available at: https://www.pcgamer.com/(Accessed 25 August 2025).

Chalk, A. (2024) *Hogwarts Legacy sold over 22 million copies in 2023, making it the biggest game of the year.* PC Gamer, 9 January. Available at: https://www.pcgamer.com/ (Accessed 25 August 2025).

Hydock, C., Paharia, N. & Weber, T.J. (2019) *The Consumer Response to Corporate Political Advocacy: A Review and Future Directions.* SSRN Working Paper (September 13, 2019). Available at: https://papers.ssrn.com/ (Accessed 25 August 2025).

DEI Coding Breakdown (one■page summary)

Columns: title, minority_lead_flag, lgbtq_top5_flag, nontrad_flag, DEI_score

title	minority_lead_flag	lgbtq_top5_flag	nontrad_flag	DEI_score
Call of Duty: Black Ops III	0	0	0	0
Halo 5: Guardians	1	0	0	1
The Witcher 3: Wild Hunt	0	0	0	0
Mass Effect: Andromeda	0	1	0	1
Red Dead Redemption 2	0	0	0	0
Shadow of the Tomb Raider	0	0	0	0
Battlefield V	1	0	1	2
Call of Duty: Modern Warfare (2019)	1	0	0	1
Resident Evil 2 (remake)	0	0	0	0
Cyberpunk 2077	0	1	0	1
Call of Duty: Black Ops Cold War	0	0	0	0
The Last of Us Part II	1	0	0	1
Marvel’s Avengers	1	0	1	2
Final Fantasy VII Remake	0	0	0	0
Marvel’s Spider-Man: Miles Morales	1	0	1	2
Ratchet & Clank: Rift Apart	0	0	0	0
Halo Infinite	0	0	0	0
Horizon Forbidden West	0	0	0	0
The Callisto Protocol	0	0	0	0
Marvel’s Spider-Man 2	1	0	1	2
Immortals of Aveum	0	0	0	0
Alan Wake 2	1	0	0	1
Star Wars Jedi: Survivor	0	0	0	0
Concord	1	0	0	1
Skull and Bones	0	0	0	0
Black Myth: Wukong	1	0	0	1
Assassin’s Creed Shadows	1	0	1	2

*Note: The full review-level dataset (66,941 rows) is embedded in this PDF as **review_level_dataset_no_text_coreflag.csv**.*