The American Journal of Engineering and Technology
ISSN 2689-0984 | Open Access

TYPE Original Research
PAGE NO. 01-07
DOI 10.37547/tajet/Volume07Issue10-01

AI as a Catalyst for Automation in High-Level Game Design for Adaptive Game Structures and Enhancing Player Engagement

Oleh Riazanov

Head of Game Design, Bini Games Kitchener, Canada

OPEN ACCESS

SUBMITED 22 August 2025
ACCEPTED 14 September 2025
PUBLISHED 01 October 2025
VOLUME Vol.07 Issue 10 2025

CITATION
Oleh Riazanov. (2025). AI as a Catalyst for Automation in High-Level Game Design for Adaptive Game Structures and Enhancing Player Engagement. The American Journal of Engineering and Technology, 7(10), 01–07. https://doi.org/10.37547/tajet/Volume07Issue10-01

Abstract: The paper discusses the great change at the high end of game development brought about by generative artificial intelligence — not simply a minor tooling upgrade but rather a machine that can be used in automating the building of adaptive systems and increasing the gravitational pull of engagement for players. At a studio level, adoption has already moved beyond pilot projects to institutional practice-a shift that turns mere efficiency gains into matters of existence for quite many teams. The paper's novel input is a neat merge of three paths: first, the lift of a game's "skeleton" to a formal meta-model that lets big language models set up event graphs; second, a change in how hardship is made real — moving from simple time factors (D1/D7) to group-level keeping and leaving-risk hints so that tuning helps long-term involvement rather than short-term win rates; and third, the use of a feedback loop — watcher-helper sending possible changes to a skilled checker, with checker fixes sent back as top-notch training samples to lessen false guesses methodically. AI-fueled automation substantially speeds up prototyping and can enhance extended player engagement; however, these advances are precarious — dependent on legally verifiable data origin, strict dual-path (algorithmic + human) examination, and transparent intervention mechanisms — if not present, then adaptive systems may stray from customization to hidden control. This paper aims at academics and business professionals who seek disciplined, practical

methods for implementing AI in design while maintaining creative freedom and responsibility.

Keywords: artificial intelligence, automation of game design, adaptive game structures, player engagement, dynamic difficulty.

Introduction

In 2025, the digital entertainment industry reached an inflection point: the State of the Game Industry survey records that more than a third of developers already personally employ generative models in production cycles, and over half work in studios where such solutions are deployed at the corporate level (Elderkin, 2025). This implements the discussion of high-level design automation because reducing costs, speeding sprints, and shortening time-to-motion are the major factors for staying alive in groups when confronting rising costs and working under conditions of uncertain funding.

Google Cloud and The Harris Poll elaborate further on the magnitude of this shift in a joint study, indicating that 87% of developers across five countries who were surveyed had reported integrating AI agents into their workflows. Most cite using automation to increase creative throughput by automating repetitive tasks. At the same time, 94% view such agents as a long-term means of cost reduction; 63% are anxious because the rights to data and resulting content remain unclear, thereby emphasizing the need for both regulatory and methodological guardrails when scaling such technologies (Kachwala, 2025a).

To have a good conversation, ensure you understand the meaning of certain words. High-level design denotes work with the abstract architecture of a game: the architect defines the set of systems, the tempo-rhythmic scaffold, the player's role, and the interrelations among mechanics, without delving into code or final parameters. This level answers what and why, leaving the how exactly to the details of low-level implementation. Such a distinction enables parts of solutions to transfer across projects. It serves as a meta-model for further automation via large language models, which are capable of operating on conceptual representations (Bycer, 2014).

An adaptive structure signifies the aggregate of algorithmic rules that modulate the game's flow or difficulty in response to the current player profile. Dynamic difficulty adjustment studies have found that the most effective approaches are those that optimize for return probability and session length, rather than win rate or a simple count of losses. Thus, modern work shifts its focus from single-indicator measures of skill to churn metrics that comprise multiple factors (Mi & Gao, 2025). In such systems, it is the AI that decides when to intervene in the balancing act. However, it is still the designer who sets boundaries beyond which the algorithm cannot go, thereby preserving both authorial style and, indeed, challenge integrity.

In-game analytics define engagement as a regular, repeated return to the interactive environment. Quantitatively, it is measured by N-day retention. The players who are active on a particular calendar day since their first session defines N-day retention percentage. This criterion mixed cognitive interest and emotional value with the systemic appeal of live service; therefore, the adaptation strategy would be aligned with real business indicators and would formalize the effect produced by AI-driven automation.

Materials and Methodology

The research is based on a multilevel synthesis of theoretical and empirical analysis, aiming to identify the role of artificial intelligence in automating high-level game design. As a foundational base, we use results from the State of the Game Industry survey, which captures the scale of actual deployment of generative models in production pipelines, where more than a third of developers already apply them personally and over half are engaged in studios that have institutionalized such practices (Elderkin, 2025; GDC, 2025). This snapshot is complemented by data from the joint Google Cloud and Harris Poll study, showing that 87% of professionals have integrated AI agents to automate repetitive tasks, with 94% associating their use with long-term cost reductions, yet 63% voicing concern over legal uncertainty regarding access to data and outputs (Kachwala, 2025a). This context shapes the methodological approach, wherein empirical findings are interlinked with legal and organizational constraints.

The theoretical frame rests on the concept of separating design levels: high-level design is construed as working with the game's abstract scaffold—system sets, rhythmic structure, and objective functions—which permits formalization as a meta-model for integrating generative solutions (Bycer, 2014). In parallel, we consider research on adaptive structures where emphasis shifts from linear win/loss indicators to multifactor retention and engagement metrics (Mi &

Gao, 2025). Thus, the methodological base combines analytical work on conceptual design, behavioral game analytics, and normative sources that define the bounds of permissible AI use.

Practically, the study draws on a comparative case analysis of automation, including projects where large language models produce coherent event graphs and level layouts (Sudhakaran et al., 2023), as well as industry initiatives that implement generative solutions into sandboxes and service platforms (McGuire, 2024; NVIDIA Developer, n.d.). A separate line examines A/B testing practices and cohort retention models offered by analytics platforms Amplitude (Amplitude, 2025), Unity Gaming Services (Unity, n.d.), and PlayFab (Microsoft, 2025), which allows us to view automation not only as a prototyping tool but also as a basis for a hypothesis-driven experimental paradigm.

Results and Discussion

By mid-2025, the role of generative AI had moved beyond experimental plug-ins: the GDC survey records that every third developer practices such models personally, and more than half work in studios where generative solutions are already codified in procedures—accordingly, discussing the automation of upper-level design without AI becomes methodologically untenable (GDC, 2025).

The macroeconomic backdrop reinforces this trend: according to the latest Newzoo report, the global video game market will add only 3,4% to reach 188,9 billion dollars in 2025, meaning the industry is entering a phase of moderate growth where new revenues are extracted not by audience expansion but by extending the life cycle of already acquired users. This inertia, compounded by delays of marquee releases and rising hardware prices, objectively shifts production teams' focus from instantaneous CPI to durable retention and time-in-game (Kachwala, 2025b).

Traditional D1/D7 rates were created in the days of one-and-done hits. They do not work anymore. First, quarterly benchmarks now reveal an ever-sharper genre and regional spread, where casual and mid-core project medians differ by many dozen percentage points; a mechanical comparison to an average temperature misleads. Second, cross-platform services have made session boundaries fuzzy: a player may begin on mobile play, later on console, return through cloud, the classic formula returning exactly on day N loses diagnostic force. Finally, the growth of in-game subscriptions and seasonal battle passes demands accounting not only for the fact of return but also for the modality of content consumption, which compels a switch to cohort return-or-consumption models and dynamic LTV recalculation. This reliance on finer metrics is also supported by Amplitude's analytical guidance, which treats retention as a quasi-continuous function of recoverable demand rather than a discrete binary trait (Amplitude, 2025).

By fall 2025, large language models had moved from experimentation to deployment — learning not how to write lines for conversation, but rather to generate coherent graphs of events with exact pacing and loop structures. MarioGPT should stand as an example, as described in Figure 1, wherein a transformer fine-tuned on data creates Super Mario levels out of text description and gives the designer semantic constraint ability, thereby keeping control over scene objectives; the model removes prototyping obligation at the tile level, allowing analysis at the level of imagination (Sudhakaran et al., 2023).

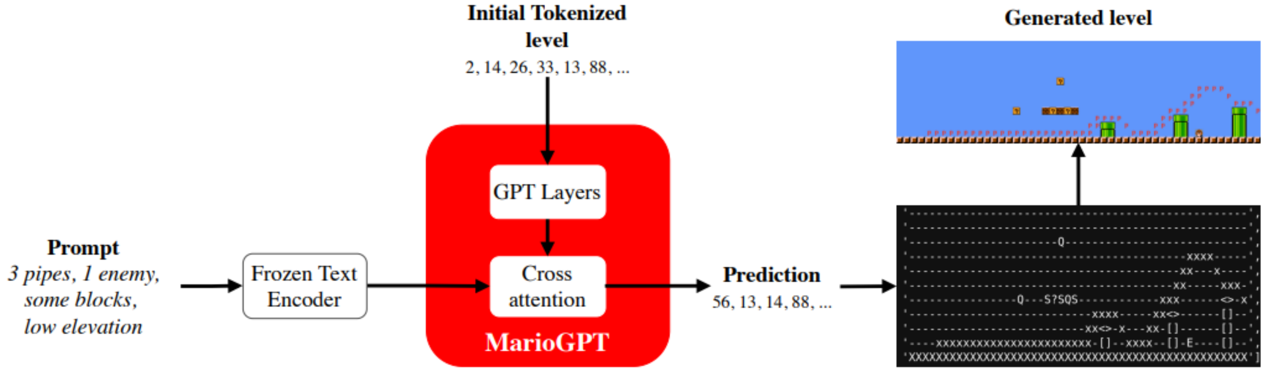

Fig. 1. MarioGPT prediction pipeline (Sudhakaran et al., 2023)

A comparable role in the sandbox segment is played by Roblox's four-dimensional generation initiative, wherein the system distributes interactions across time and space—work that previously fell entirely to scripters

(McGuire, 2024). Even in consolidated projects, the windows between idea and verification have narrowed: the NVIDIA ACE kit provides ready models for speech, animation, and logic that assemble an interactive assistant character within hours, enabling measurement of training effects without a traditional production cycle (NVIDIA Developer, n.d.).

The next layer of automation concerns dynamic difficulty. A 2024 empirical review confirms that no fixed adjustment scheme dominates competitors across all cohorts; performance depends on the signal chosen by the developer and the players' cultural context (Fisher & Kulshreshth, 2024). In response, the EDDA approach was proposed, where the policy of adjusting challenge optimizes churn risk rather than the average win rate, allowing the algorithm to operate in tandem with a retention metric and thus directly stitch the business objective into the adaptation loop (Mi & Gao, 2025). Here, a large language model acts as a watcher, pulling in telemetry. At the same time, mode switches are handed off to proxies set by the designer, which keeps the creative mark intact and cuts down on chances for help to butt in.

The quality-assurance contour is changing. In Human-AI Collaborative Testing research, vision-language agents analyzed 800 scenarios and correctly identified defects 97% of the time. However, for every single misinterpretation made by the agent, all of the human testers then fell into a cascading error (Zhang et al., 2025). An example of such an architecture is provided in Figure 2.

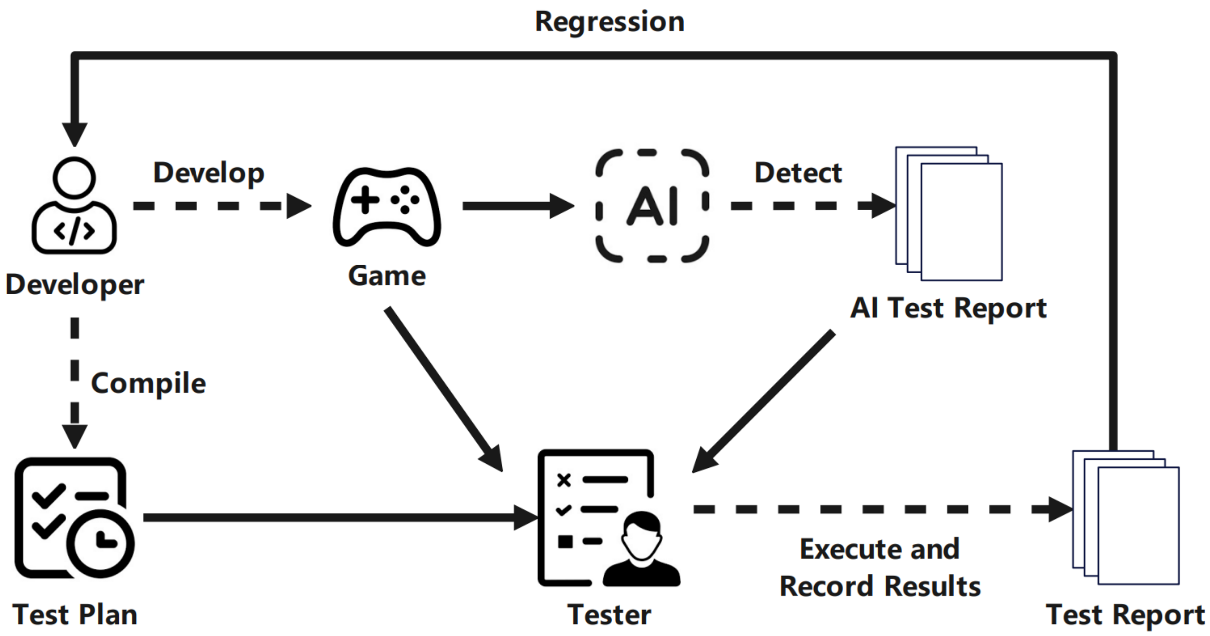

Fig. 2. AI-Assisted Game Testing Workflow (Zhang et al., 2025)

Practice shows that the optimal configuration is agent detects while the human confirms and classifies, after which data on model misses flow back into the training set. This loop not only speeds ticket closure but also systematically reduces the share of hallucinations, since the algorithm learns from live deviations rather than synthetic sets.

Finally, live ops processes are shifting from intuitive tuning to continuous split testing. Within Unity Gaming Services, the designer defines variants of reward cadence or item pricing, and the platform automatically segments the audience and computes statistical significance, streaming results to a dashboard nearly in real time (Unity, n.d.). A similar mechanism in PlayFab enables parallel trials of multiple configurations and instant shutdown of losers, thereby minimizing risk (Microsoft, 2025). A large language model here serves as strategist: it generates hypotheses from historical regularities, and the split-testing system verifies them, turning adaptation into a determinate scientific process in which each iteration lifts retention not by inspiration but through data-backed inference.

Since every third team already uses generative models in production and 87% of developers utilize AI agents for automating their routine operations, the discussion about adaptive design is no longer abstract (Kachwala, 2025a). It requires specific rules on how to transform accumulated precedents into a reproducible practice.

Begin by automating the embryo, not the final build. Large models can be used to undertake coarse embossing of intent, e.g., quest graph, level mock-up, and UI skeleton, while constraints and test goals are still authored by someone who is not a designer. Relate dynamic difficulty to retention indicators. To truly resolve the paradox, have your model optimize for return probability rather than pure victory. Adjust your algorithmic difficulty precisely in order to minimize risk of churn, retain signal choice, and the bounds of intervention within the hands of your designer. Human responsibility lies within the test loop; therefore, the optimal configuration proposed by the agent is confirmed by the expert, wherein false positives themselves are incorporated into the training data to mitigate recurrence.

The fourth principle treats AI-NPCs as a user service rather than a decorative feature. PUBG Ally, built on ACE, demonstrates that a co-author bot that drives vehicles, shares loot, and suggests tactics bolsters early retention more reliably than a content-free chatbot; the pivotal element is an assistance intent embedded in the game as a layer of instruction and feedback (Peters, 2025). The low latency of local ACE inference enables such a helper to act without noticeable pauses, thereby preserving the flow.

The fifth principle mandates continuous experimentation instead of static balance. Once a prototype is ready, its variants enter an A/B cadence in which the Unity UGS platform measures the impacts on retention and monetization. At the same time, PlayFab launches parallel split-tests with confidence bounds sufficient to prevent losing configurations. This auto-funnel converts game updates into a list of guesses that can be checked, reducing the likelihood of gut-feeling fallbacks (Unity, n.d.). The method begins by stating the goal: the team transforms a vague task—such as reducing early drop-off—into something that can be measured and identifies signs that clearly show progress. At this point, it is crucial to determine which action signs—such as the rapid acquisition of a primary skill, frequent back-to-back misses, and a combination of economic transactions—correlate with interest, and which merely introduce more confusion. The clearer the link between the signal and the expected behavior, the lower the risk of morphing the design around a false proxy.

After signal selection, an experimental frame is created. The UGS and PlayFab platforms enable the specification of parallel variants of the game experience, automatically allocate audiences, and compute the statistical validity of the differences. It is vital to pre-commit to success and stoppage boundaries. If the target indicator fails to improve to the predefined level of confidence, the experiment is closed to avoid wasting the available user volume and diluting the inference.

The next layer involves integrating SDKs and configuring logging. System hooks stream telemetry, including latencies, match outcomes, and purchases, onto a unified bus, where data are tagged with attempt identifiers. Such detailed logging enables not only post-factum evaluation of hypotheses but also rapid patch-fixing when anomalous spikes of errors or drops in desired metrics are detected.

At last, all signs are measured against the present three-month main series by type and area. The group checks its own performance, interest, and profitability against outside averages, adjusts target ranges, and updates estimates for the next round. The market moves quicker than products grow; thus, changing aims every three months is the shortest cycle that keeps up sharp competition without causing too much trouble in the team. The overall architecture is presented in Figure 3.

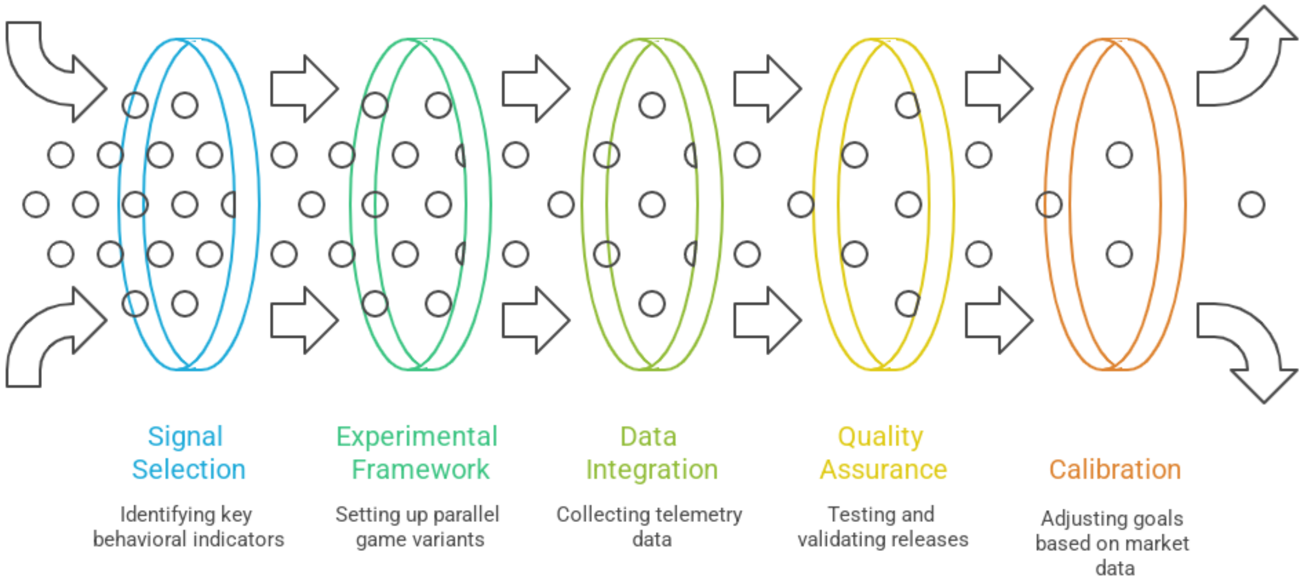

Fig. 3. Game Development Methodology Funnel

Completing the map of opportunities, we must acknowledge that the domain of automated high-level design is held not only by the vector of progress but also by the relief of constraints. Foremost is the question of rights to machine-generated material and to the raw data on which the machine was trained. A confident mistake can easily find its way unchallenged into a design document, particularly when teams have become accustomed to relying on the speed of the model. This means that second-order defects make their way to production: though the system may work formally, semantic assumptions have already violated the original intent. To keep trust going, there has to be an auto-anomaly validator logging statistically unusual answers and an expert-qualifier making the final decision thereby bringing responsibility back to humans but with a lag.

The ethical and regulatory field is also changing. Difficulty modulation that autonomously adjusts based on the player's emotional state can cross over from support to manipulation if not made transparent, and may suddenly promote extended playing time. Already a topic for discussion among regulators about just where those boundaries lie, with adaptation principles certification standards available to developers to disclose their adaptation principles. The general data protection regime that sits atop this forbids the harvesting of behavioral signals without explicit consent and directly impacts just how incomplete telemetry can be that the algorithm would require. Thus, every studio would have to invent not only new mechanics but also continuously balance personalization as a benefit against user autonomy, with an opt-out right from any analytical tracking.

Conclusion

The study demonstrates that artificial intelligence, integrated into high-level game-design processes, has ceased to be a laboratory experiment and has become an everyday instrument reshaping the very architecture of development. However, the risks identified demonstrate that until there is a parallel construction of constraint systems, the technologies' potentials cannot be fully realized. The matter concerning rights to content and source data requires transparent versioning, as well as an easily reproducible transformation log. Model hallucinations mean that there should be a dual verification circuit in which the human link is decisive. Ethical and regulatory challenges underscore that gameplay adaptivity must always remain a means of engagement – not of covert manipulation – and, more practically, the degree of personalization allowed to players versus their autonomy that will determine the industry's social legitimacy in the future.

So, the auto upper-level design by AI comes out as a two-level process: first, even faster prototyping added to cost reduction joined by new depths of analytics; second, and parallel with equal energy, an urgent drive to frame the process in legality, in ethics, and in methodology so that development stays within the peripheries of trust and responsibility. The ultimate vector is set not by the speed of deployment itself, but by the industry's capacity to codify standards in which technologies augment rather than supplant authorial intent—preserving equilibrium between production efficiency and the sustainability of the play experience.

References

1. Amplitude. (2025). *Interpret your retention analysis*. Amplitude. https://amplitude.com/docs/analytics/charts/retention-analysis/retention-analysis-interpret

2. Bycer, J. (2014). *Understanding High and Low-Level Game Design*. Game Developer. https://www.gamedeveloper.com/design/understanding-high-and-low-level-game-design

3. Elderkin, B. (2025). *GDC 2025 State of the Game Industry: Devs weigh in on layoffs, AI, and more*. Game Developer. https://www.gamedeveloper.com/business/gdc-2025-state-of-the-game-industry-devs-weigh-in-on-layoffs-ai-and-more

4. Fisher, N., & Kulshreshth, A. K. (2024). Exploring Dynamic Difficulty Adjustment Methods for Video Games. *Virtual Worlds*, *3*(2), 230–255. https://doi.org/10.3390/virtualworlds3020012

5. GDC. (2025). *State of the Game Industry*. Game Developers Conference. https://gdconf.com/state-game-industry/

6. Kachwala, Z. (2025a). Nearly 90% of video game developers utilize AI agents, according to a Google study. *Reuters*. https://www.reuters.com/business/nearly-90-videogame-developers-use-ai-agents-google-study-shows-2025-08-18/

7. Kachwala, Z. (2025b, June 17). The delay of GTA VI weighs on global videogame market growth, data shows. *Reuters*. https://www.reuters.com/business/media-telecom/gta-vi-delay-weighs-global-videogame-market-growth-data-shows-2025-06-17/

8. McGuire, M. (2024). *Roblox's Road to 4D Generative AI*. Roblox Corporate. https://corp.roblox.com/newsroom/2024/06/robloxs-road-to-4d-generative-ai

9. Mi, Q., & Gao, T. (2025). Engagement-Oriented Dynamic Difficulty Adjustment. *Applied Sciences*, *15*(10). https://doi.org/10.3390/app15105610

10. Microsoft. (2025, May). *Experimentation*. Microsoft. https://learn.microsoft.com/en-us/gaming/playfab/live-service-management/game-configuration/experiments/

11. NVIDIA Developer. (n.d.). *NVIDIA ACE for Games*. NVIDIA Developer. Retrieved August 9, 2025, from https://developer.nvidia.com/ace-for-games

12. Peters, J. (2025, January 7). *Nvidia's AI NPCs are no longer chatbots — they're your new PUBG teammate*. The Verge. https://www.theverge.com/2025/1/6/24337949/nvidia-ace-ai-npcs-pubg-ally-teammate

13. Sudhakaran, S., González-Duque, M., Glanois, C., Freiberger, M., Najarro, E., & Risi, S. (2023). MarioGPT: Open-Ended Text2Level Generation through Large Language Models. *Arxiv*. https://doi.org/10.48550/arxiv.2302.05981

14. Unity. (n.d.). *A/B testing*. Unity. Retrieved August 12, 2025, from https://docs.unity.com/ugs/solutions/manual/ABTest

15. Zhang, B., Xu, M., & Pan, Z. (2025). Human-AI Collaborative Game Testing with Vision Language Models. *Arxiv*. https://doi.org/10.48550/arxiv.2501.11782