Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 3, 2026

ARTIFICIAL INTELLIGENCE IN VIDEO GAME DEVELOPMENT: A SYSTEMATIC REVIEW OF APPLICATIONS, PRODUCTION PRACTICES, AND SOCIO-TECHNICAL CHALLENGES

Ali Hassan[*1], **Muhammad Umar Hameed**[2]

[*1]*University of Central Punjab*
[2]*Principal Lecturer, University of Central Punjab*

[*1]byteslicergames@gmail.com, [2]Umar.hameed@ucp.edu.pk

DOI: https://doi.org/10.5281/zenodo.24330411

Keywords
artificial intelligence; video game development; systematic review; procedural content generation; creative labour; responsible AI.

Article History
Received: 19 January 2026
Accepted: 04 March 2026
Published: 18 March 2026

Copyright @Author
Corresponding Author: *
Ali Hassan

Abstract
AI can aid in content creation, adaptive behaviour, balancing, and testing in video-game production, but there is little evidence of its benefits in production workflows outside of technical performance. In this study, 25 publications were critically synthesized from 45 articles in the period of 2009–2026. Compares computational experiments, practitioner studies, deployment evaluation, reviews, conceptual frameworks, and ethical arguments. The results show the ability to create constrained content, model behaviors, and detect anomalies in images, whereas any other generalizations regarding productivity and creativity are context- and evaluation-dependent. Solvability, predictive accuracy, and detection performance do not directly dictate player enjoyment, creative originality, and improvements throughout complete production workflows. Integration is a specification, curation, interoperability, and validation process, and thus a human control process for the organization. In these arrangements, a range of issues of labour insecurity, attribution, data governance, and accountability arise, but anticipated harms and normative arguments need to be separated from observed outcomes. The review suggests that the production value of AI relies on the way AI is integrated into creative processes, how effort is defined for correction and integration, and who is in control of the final product. Evaluation should then relate technical outcomes to production practices and the responsibilities involved in the use of these.

1. INTRODUCTION

A level can be generated that can be solved without fitting the game's pacing; dialogue can be fluent while conflicting with the game's world; an accurate detector can detect only a small subset of defects. These differences are what make it difficult to judge AI in a game development scenario solely by the performance of the program's algorithms. Liapis et al. (2019, pp. 63–65) identify a central challenge to coordinating content that is generated across various facets of the game. Chia (2022, pp. 398-402) extends the problem to production: Automation includes decisions of control and the evaluation of creative work. These views collectively indicate that capability is useful in the context of the design and organizational arrangements.

Other techniques exist for the field, with varying requirements and histories. Search-based procedural content generation relies on representations and evaluation functions, while machine learning generation relies on patterns found in existing content (Togelius et al., 2011, pp. 175–176; Summerville et al., 2018, sec. I). While generative language models enable conversational interfaces and new authoring opportunities, they are not a subset of all the previous adaptive agents, planning systems, and reusable AI architectures. It is important to note that not all procedural generation is AI, as that would make these differences hard to see.

Where relevant, runtime systems are relevant, as they are used for authoring, integrating, balancing, testing, and/or governance.

There is substantial technical knowledge provided in existing surveys. Liu et al. (2021, pp. 30–32) examine deep-learning generation and its limitations, while Gallotta et al. (2024, sec. 3.9) distinguish forms of language-model design assistance. The classifications are useful for identifying capabilities, but do not replace evidence of how developers use tools in work. However, experiences and decisions that cannot be captured by benchmark evaluations are touched upon by Alharthi's (2025) practitioner study and by Eden Doshi and Knight's (2026) ethnography. Ethical analysis introduces the notions of responsibility and ownership, but these arguments are not as well supported by evidence as the observed outcomes.

This review connects these perspectives through three questions:

1. Which AI applications support video-game development, and how are their outcomes evaluated?
2. How is AI integrated into production workflows, and what human oversight and creative control do these practices require?
3. What socio-technical challenges arise around labour, ownership, data, and responsibility, and how strong is the supporting evidence?

The contribution is an important discussion of the different kinds of research that are found on game development. The review focuses on how measurable outputs relate to production practices, including the human decisions and responsibilities related to the use of outputs.

2. Methods

2.1 Review design and scope

2.1 Review design and scope

Thematic synthesis was adopted to identify the main themes in the use of AI in video-game development, its integration into the video game production process, and the socio-technical issues. There were 25 scholarly publications in the corpus with technical experiments, empirical practitioner research, surveys, conceptual frameworks, and normative analysis.

Different kinds of evidence would be distinguished since they have different questions, and this will lead to different conclusions. Technical evaluations are for tasks that are clearly specified, practitioner studies relate to experiences in specific contexts, and conceptual or ethical analyses focus on implications and responsibilities. They were also understood in terms of their research designs, rather than as one type of evidence to support industry-wide effects.

The review concentrated on AI that could support game-content creation, agent design, balancing, testing and production integration, and related tasks. Any research that dealt with game playing performance was only relevant in the context of informing these development activities. There was no minimum publication year.

2.2 Search strategy and information sources

2.2 Search strategy and information sources

A literature search from 30 August to 1 September 2026 was completed. The search of information content sources was done on IEEE Xplore, ACM Digital Library, ScienceDirect, Springer Nature Link, SAGE Journals, MDPI, and Cambridge Core. These databases and publisher platforms covered the areas of computing research, game production, practitioner experiences, and socio-technical issues.

The search strategy spanned three domains of the intervention and innovation field, namely: AI applications, production practices, and socio-technical issues. AI techniques were searched in conjunction with the words “video game” and “video game development” or “video game governance”. Two search strings are provided; these are given in Tab. 1.

Of the 45 articles, 12 were selected for review. There were 45 articles considered for inclusion, 12 of which were selected for review. These 25 were chosen for detailed analysis and synthesis based on the relevance to the three research questions.

Table 1. Search strategy and information sources

Search focus	Search string	Information sources
AI applications	("video game" OR videogame OR "digital game") AND ("artificial intelligence" OR "machine learning" OR "reinforcement learning" OR "large language model" OR "generative AI") AND (develop OR design OR "procedural content generation" OR testing OR balancing)	IEEE Xplore; ACM Digital Library; ScienceDirect; Springer Nature Link
Production practices	("video game" OR videogame OR "digital game") AND ("artificial intelligence" OR "generative AI" OR "procedural content generation") AND (production OR workflow OR "mixed initiative" OR "mixed-initiative" OR "design assistant" OR "creative control")	ACM Digital Library; ScienceDirect; SAGE Journals; MDPI
Socio-technical challenges	("video game" OR videogame OR "digital game") AND ("artificial intelligence" OR "generative AI" OR automation) AND (ethic OR labour OR labor OR authorship OR copyright OR privacy OR governance)	SAGE Journals; Cambridge Core; ACM Digital Library; IEEE Xplore; MDPI

2.3 Selection and full-text assessment

45 articles were reviewed, and 25 papers were selected for the synthesis. Each of the 25 publications that were retained was read entirely and evaluated with regard to its relevance to the use of AI in video-game development, the production process, and the socio-technical challenges arising from them.

Full-text assessment was used to analyze the research goal, research method, results, assessment method, and limitations of each publication. The entire set of retained publications comprised the data for extraction, qualitative appraisal, and thematic synthesis.

2.4 Extraction, appraisal, and synthesis

Data were extracted on the development activity, type of AI technique, context of a study, human involvement, assessment measures, main findings, and limitations of each publication. Traceability between the source material and the claims made in the review was facilitated by the use of page or section locators.

Appraisal was qualitative rather than based on a standardized numerical rating. Task restrictions, comparison baselines, evaluation measures, and validation were considered in the context of technical studies. The recruitment methods, self-report dependency, and contextual constraints were investigated for practitioner research. Survey assessment included coverage, survey methods, and whether the survey overlapped with other evidence included. The scope and grounding of conceptual and normative publications were studied.

The findings were presented in relation to the three research questions and grouped into themes covering AI applications and evaluation, production integration and creative control, and socio-technical challenges. Within these themes, the distinctions between technical ability and production effectiveness, between the experiences reported by participants and measured outcomes, and between ethical concerns and demonstrated harms were noted. Secondary summaries were not included as further independent confirmation of primary studies on which they commented.

Due to the different research designs, contexts, and outcome measures, results from the publications were not used to create an effect estimate. The synthesis focused on the consistency of findings, type of evidence differences, and the constraints in moving findings outside of their context.

3. Findings

3.1 Characteristics of the evidence

Historically established adaptation and generation research is merged with recent

language-model and production studies in the corpus. The publication dates do not reveal the maturity of the technology: the older architectures solve integration issues that are still relevant, whereas the newer generators might be tested in a more limited prototype. Surveys increase the number of records but are not necessarily unique in their evidence and in their selection procedures. Hooshyar et al. (2018, pp. 90:3–90:5) describe a systematic selection process, whereas Gallotta et al. (2024, secs. 1–2) explicitly adopt an expert-led typology and manual review. They are not the same measure of prevalence of research and would misrepresent their designs to treat them as such.

3.2 Applications and the meaning of their evaluation

Choices of quality measures in procedural generation include a design decision. Togelius et al. (2011, p.176) provide an explanation of how difficult it is to formalise goals like fun. It does not go away in later learned generators. Summerville et al. (2018, sec. IV.A) recognize playability as a limitation, whereas Liu et al. (2021, p. 31) demonstrate the focus of research on (tile-based) content and familiar game domains. The success of the structures is therefore more indicative of the specific generation task than of the overall creative ability.

Nam et al. (2022, secs. III, V–VI) investigate reinforcement-learning generation in a simplified version of an RPG, employing predetermined
quality measures. These kinds of proxies allow comparisons without player enjoyment being established. Zakaria et al. (2021, pp. 8–10, 21–22) explore generation in 7×7 Sokoban, a game in which the solvability and controllability of a game is computable under explicit rules. They have published a preprint that can be used to conclude in that experimental space, not across commercial-level design pipelines. These studies illustrate why there is a need to have a separate account of whether the evaluation is the intended human experience in the case of generation.

This is echoed by the distinction that can be found in narrative generation. On 114 judgments from 38 students, De Lima et al. (2022) report 48.68% authorship-classification accuracy. The near-chance judgments suggest that it is difficult to distinguish authors in this prototype. They do not create anything similar in terms of originality, professional quality of the narrative, or replayability. The issues addressed by Salmaze et al. (2026) are inconsistent evaluation and limited human baselines in language-model narrative research. There are 51 outcomes in their Table 2, but the number of publications is only 55, so the two numbers do not match, and it is not possible to translate these figures into percentages for the total corpus. The general remark on the comparability is still valid, but without the addition of an unknown factor.

Table 2. Application groups, evaluation endpoints, and limits of production inference

Development function	Representative sources	Evaluation focus	Inference boundary
Content generation	Togelius et al. (2011); Liu et al. (2021); Nam et al. (2022); Zakaria et al. (2021)	Solvability, diversity, control	Proxy quality does not establish enjoyment or pipeline savings
Narrative and dialogue	de Lima et al. (2022); Gallotta et al. (2024); Salmaze et al. (2026)	Coherence, authorship judgments, reported experience	Fluency and uncertain authorship are incomplete quality measures
Testing	Albaghajati and Ahmed (2023); Paduraru (2026)	Approach suitability, detection, inspection effort	Coverage remains dependent on fault class and deployment

Behaviour and balancing	Bakkes et al. (2009); Pfau et al. (2023); Zohaib (2018)	Adaptation, simulated combat, difficulty	Behavioural outcomes require contextual human validation
Integration	Dignum et al. (2009); Safadi et al. (2015); Liapis et al. (2019)	Coordination, reuse, coherence	Component portability does not demonstrate whole-game automation

As seen in Table 2, there are differences among application categories as to what is considered successful. Testing offers a better production endpoint, given that detection is linked to actual inspection work. The attribute-based framework by Albaghajati and Ahmed (2023, Abstract) addresses an appropriate approach to testing, rather than being evidence of a universal efficiency effect. According to Paduraru (2026), the inspection time is reduced by around 38% in an internal QA environment. This operational outcome is more related to production than to classification accuracy only, and it has organizational and game-specific conditions that limit transfer.

There is also a balance to be struck between the modelled and lived experience. Pfau et al. (2023, secs. III–IV, VII) take 6 months of data from 213 Aion players and model combat behaviour in a setting where it is only PvE with equipment being constant, which limits generalization. Bakkes et al. (2009, pp. 99–103) show adaptation under perfect-information assumptions for three Spring maps, and the identification of human difficulty is left for the future. Zohaib (2018, "Conclusions" section, para. 2) clearly displays selected (non-exhaustive) DDA examples. Together, these sources facilitate context-specific balancing strategies instead of a general ability to infer optimal difficulty.

3.3 Production integration and creative control

Design changes that are made in a mixed-initiative manner involve where the decisions are being made. There are three types of assistance: conceptual assistance, procedural assistance, and production assistance, according to Gallotta et al. (2024, sec. 3.9). There is still plenty of room for interpretation by the designer in an idea generator, while an artifact generator still requires constraints and decisions about accepting its outputs. Dutra et al. (2025, Abstract) present a reinforcement-learning framework that uses level models provided by experts. This makes human specification part of generation, but does not establish a general improvement in designer satisfaction or studio productivity. In contrast, Alharthi (2025) found that out of 42 participants, 90.5% reported using generative AI for ideation, 83.3% for asset creation, and 11.9% for QA. The pattern suggests the presence of different applications in the sampled community, and not industry-wide. The study's generalisability is limited by the regional restrictions and constraints, and the recruitment through the development-oriented WhatsApp group (Alharthi, 2025, p. 6–7, 17). Therefore, the results of the reported usefulness must be interpreted as a subjective evaluation by participants, and not as a comparison of the production methods.

There is work to do aside from the generative model that also needs to be done for integration. Dignum et al. (2009) state that intelligent gameplay is related to synchronization and coordination between agents and game systems. In two environments, as demonstrated by Safadi et al. (2015), there is reuse of targeting logic, but in this case there are still problems of compatibility, overhead, and developer control. Reuse can lead to less repetition of implementation and be dependent on shared representations/ adapters. Liapis et al. (2019, pp. 63–65) apply this question to the creative side: local success should be in harmony with the story, game mechanics, and sound and image logos of the whole game.

Paduraru (2026) presents an example of a state-aware QA system with a retained human QA with his Paduraru system. The results, however, must be used with caution: The expensive verification step is assumed to be triggered; hence, the amortized throughputs of 38-49 FPS are reported, which, if it is not triggered, would

result in 59.7 milliseconds of full hierarchical processing. Efficiency is not a faster model alone in part; it has to do with pipeline arrangement.

Chia (2022, pp. 398–402) explains why the remaining work matters culturally. Procedural systems involve the dispersion of agency amongst the technical and artistic aspects of the work, with professional discourses valuing the different contributions at different levels. Kim and Kim (2024, pp. 57–61) discuss collaborative possibilities, but their overview's reported examples do not independently verify industry-wide gains. In all these sources, the term "human involvement" is not adequate unless it names who is involved in the specification, correction, approval, and maintenance of the outputs. Those arrangements are the key to production value.

3.4 Socio-technical challenges

Creative authority is related to the author and dissemination of work. Chia (2022, pp. 396, 402) focuses on the relationship between the discourse on automation and the hierarchy of production and the difference between valued creative judgement and supposedly mechanical execution. What is important for AI is how tools can be used to redistribute recognition as well as tasks. However, Chia's analysis of procedural automation is not necessarily evidence of all workflows depicted by contemporary language models.

Alharthi (2025) notes issues of originality, dependence, ownership, and collaboration. Eden Doshi and Knight (2026) demonstrate that there is no contradiction between automation anxiety and interest in assistance and experimentation. Their ethnographic cases, in contrast, add a wrinkle to the notion of enthusiastic uptake and resistance. They create locally negotiated meaning, for futures of uncertainty, rather than a calibrated causal connection between AI and unemployment.

Table 3. Socio-technical issues and the status of their supporting evidence

Issue	Evidence and sources	Critical implication
Creative authority and labour	Production discourse, practitioner accounts, and ethnographic cases: Chia (2022); Alharthi (2025); Eden Doshi and Knight (2026)	Control and recognition may shift; employment effects remain unquantified.
Ownership and provenance	Practitioner concerns and ethical analysis: Alharthi (2025); Melhart et al. (2024); Gallotta et al. (2024)	Attribution and responsibility for training material require explicit arrangements.
Player data and autonomy	Modelling limitations and ethical arguments: Hooshyar et al. (2018); Yannakakis and Melhart (2023); Melhart et al. (2024)	Predictive accuracy alone does not justify data collection or intervention; meaningful consent and control over subsequent uses must also be considered.
Governance and reproducibility	Responsible-AI proposals and mapping findings: Canca et al. (2024); Salmaze et al. (2026)	Accountability requires documented practices; proposed safeguards are not proven effective.

The difference in Table 3 is fundamental: while a measured prevalence estimate is not necessary to a credible ethical concern, it is not possible to report a credible ethical concern as a measured prevalence estimate. Melhart et al. (2024, secs. IV–VI) examine privacy, transparency, and ownership within adaptive systems. Whether gathering data is justified cannot be determined by the system's usefulness in predicting behaviour alone; it also requires consideration of meaningful consent and users' control over subsequent uses.

Hooshyar et al. (2018, p. 90:16) name data access and generalizability restrictions, and Yannakakis and Melhart (2023, secs. The different aspects of affective sensing, detection,

and adaptation are discussed and distinguished (II, X), and their ethical boundaries are explored. Linking those sources makes one see that uncertainty is an issue of governance, and that a context-dependent inference is not necessarily a reliable account of emotion and can misdirect interventions. Improved predictive capacity does not address the question of who sets the goal of adaptation.

Ownership issues also apply to an artefact that has been created. Gallotta et al. (2024, sec. 6) discuss training material, output attribution and model opacity while Melhart et al. (2024, sec. VI) consider responsibilities in co-creation. These analyses aim to find tensions, not to resolve existing copyright law. Canca et al. (2024) suggest integrated ethical risk assessment, as safeguards can be contradictory; for example, more moderation can lead to more surveillance. The proposal provides a reason for governance, but not evidence that a specific process has lessened harm.

Lastly, Salmaze et al. (2026) cite underreporting of prompts and model versions. This reduces reproducibility and makes it difficult to determine who is responsible for different results. Technical documentation thus becomes a link between the integrity of research and its management rather than just a mere requirement for academic reporting.

4. Discussion

4.1 From capability to production value

The synthesis identifies a recurring relationship between AI capabilities and the additional work needed to use them. The output from generators needs to be specified and validated, the behaviour in behavioural models needs to be interpreted, the reusable architecture needs adapters, and the outputs of the detection need to be triaged. This doesn't prove that automation simply relocates a workload that didn't change before it was automated. It demonstrates why the benefits need to be measured throughout a process instead of from the performance of a single component.

It is instructive to compare with Alharthi (2025) and Paduraru (2026). Questions answered by perceived and observed inspection-time reduction are not the same. Their findings are not mutually dependent and do not require a common estimate of productivity. Better comparisons would involve the correction effort, output, organizational context, and tasks involved.

Creative control should also be considered something that is done through decision-making, not automatically gained when there's a human supervisor. Gallotta et al.'s (2024, sec. 3.9) assistance categories define the levels of initiative, as well as Chia (2022, pp.398-402) describe how authority and recognition can be imbalanced. Practitioners may correct outputs, thus performing essential processes, but not necessarily guiding the tool's goals. That is why it is important to embed production organization under an AI evaluation instead of conducting production organization after the technical evaluation.

4.2 Evidence gaps and research priorities

Independent and sustained observation of whole production workflows is the most missing item in this collection. Computational evaluations tend to focus on the task, and practitioner studies shed light on experience without separating out causal effects. These research studies should be related to each other, in the sense of investigating what is accepted, rejected, repaired, and maintained over time by developers, as well as technical results.

There would also be a benefit to human-authored baselines and transparent model or data records for the sake of comparison. Salmaze et al. (2026) show how challenging it is to combine the incompatible assessments in the narratives. Liapis et al. (2019, pp. 63–65) explain why whole-game coherence complicates assessment of individual generators. Future research should thus allow a determination of production suitability in terms of the role of the artefact and not be based on one general creativity or quality value. These priorities are identified within this collection and not the claim that no relevant research is to be found elsewhere.

4.3 Review limitations

Limitations of the findings are the 25 publications that were selected, as well as the treatment of the search and selection. The corpus includes various types of contributions: computational experiments, practitioner research, reviews, and conceptual or normative

contributions, all with their different purposes and evaluation modes. This variety makes it harder to compare them thematically, but impedes the ability to estimate the effectiveness of AI. Repeat surveys may also cover the same studies and thus not provide independent verification.

Limited transferability due to simplified experimental contexts, game-specific data sets, and practitioner accounts that are context-specific. Technical measures are not necessarily indicative of player experience or complete production costs, while self-reported benefits can not determine the causal gains in productivity (Alharthi, 2025; Liu et al., 2021, pp. 31–32). Further, findings of models are limited in their continued relevance due to rapid technological change. Consequently, the review is able to make statements on selected evidence, rather than on the levels of industry-wide adoption, employment effects or production efficiencies.

5. Conclusion

The chosen literature provides examples of the use of AI in content generation, narratives, behaviour modelling, balancing, and testing. They are judged by varying criteria, and a successful production of a limited technical task does not necessarily demonstrate the effectiveness of production overall. The integration relies on specifications, shared representations, editing, and validation, and the allocation of the integration tasks creates an allocation of creative control.

Socio-technical concerns are found in the same arrangements: training and player data impact on ownership and consent; adaptive objectives impact autonomy; changes in the output of the model impact on accountability; production hierarchies impact on recognition. These concerns are intelligible, in the sense of being conveyed by practitioner accounts and ethical arguments, but they are separate from any measured employment or productivity effects.

The most important take-home message from the review is that AI should be evaluated in the context of the creative and organizational processes that enable its products to be used. In all 25 publications included, the synthesis emphasizes the need to consider technical performance along with correction and integration effort,
human decision-making, and production and use responsibilities.

REFERENCES

Albaghajati, A., & Ahmed, M. (2023). Video game automated testing approaches: An assessment framework. *IEEE Transactions on Games, 15*(1), 81–94. https://doi.org/10.1109/TG.2020.3032796

Alharthi, S. A. (2025). Generative AI in game design: Enhancing creativity or constraining innovation? *Journal of Intelligence, 13*(6), Article 60. https://doi.org/10.3390/jintelligence13060060

Bakkes, S., Spronck, P., & van den Herik, J. (2009). Rapid and reliable adaptation of video game AI. *IEEE Transactions on Computational Intelligence and AI in Games, 1*(2), 93–104. https://doi.org/10.1109/TCIAIG.2009.2029084

Canca, C., Ihle, L. H., & Schoene, A. M. (2024). Why the gaming industry needs responsible AI. *Games: Research and Practice, 2*(2), Article 17. https://doi.org/10.1145/3675803

Chia, A. (2022). The artist and the automaton in digital game production. *Convergence: The International Journal of Research into New Media Technologies, 28*(2), 389–412. https://doi.org/10.1177/13548565221076434

de Lima, E. S., Feijó, B., & Furtado, A. L. (2022). Procedural generation of branching quests for games. *Entertainment Computing, 43*, Article 100491. https://doi.org/10.1016/j.entcom.2022.100491

Dignum, F., Westra, J., van Doesburg, W. A., & Harbers, M. (2009). Games and agents: Designing intelligent gameplay. *International Journal of Computer Games Technology, 2009*, Article 837095. https://doi.org/10.1155/2009/837095

Dutra, P. V. M., Villela, S. M., & Fonseca Neto, R. (2025). A mixed-initiative design framework for procedural content generation using reinforcement learning. *Entertainment Computing, 52*, Article 100759. https://doi.org/10.1016/j.entcom.2024.100759

Eden Doshi, A., & Knight, D. M. (2026). On the edge of the future: Automation anxiety, generative AI, and memory in game development. *Memory, Mind & Media, 5*, Article e14. https://doi.org/10.1017/mem.2026.10040

Gallotta, R., Todd, G., Zammit, M., Earle, S., Liapis, A., Togelius, J., & Yannakakis, G. N. (2024). Large language models and games: A survey and roadmap. *IEEE Transactions on Games*. https://doi.org/10.1109/TG.2024.3461510

Hooshyar, D., Yousefi, M., & Lim, H. (2018). Data-driven approaches to game player modeling: A systematic literature review. ACM *Computing Surveys, 50*(6), Article 90. https://doi.org/10.1145/3145814

Kim, E. J., & Kim, J. Y. (2024). Transforming game development with generative artificial intelligence: Opportunities and challenges. *컴퓨터게임및콘텐츠논문지, 37*(2), 53–63. https://doi.org/10.22819/kscg.2024.37.2.006

Liapis, A., Yannakakis, G. N., Nelson, M. J., Preuss, M., & Bidarra, R. (2019). Orchestrating game generation. *IEEE Transactions on Games, 11*(1), 48–68. https://doi.org/10.1109/TG.2018.2870876

Liu, J., Snodgrass, S., Khalifa, A., Risi, S., Yannakakis, G. N., & Togelius, J. (2021). Deep learning for procedural content generation. *Neural Computing and Applications, 33*(1), 19–37. https://doi.org/10.1007/s00521-020-05383-8

Melhart, D., Togelius, J., Mikkelsen, B., Holmgård, C., & Yannakakis, G. N. (2024). The ethics of AI in games. *IEEE Transactions on Affective Computing, 15*(1), 79–92. https://doi.org/10.1109/TAFFC.2023.3276425

Nam, S.-G., Hsueh, C.-H., & Ikeda, K. (2022). Generation of game stages with quality and diversity by reinforcement learning in turn-based RPG. *IEEE Transactions on Games, 14*(3), 488–501. https://doi.org/10.1109/TG.2021.3113313

Paduraru, C. (2026). A state-aware, hierarchical deep learning framework for automated visual glitch detection in games. *Engineering Applications of Artificial Intelligence, 166*, Article 113497. https://doi.org/10.1016/j.engappai.2025.113497

Pfau, J., Liapis, A., Yannakakis, G. N., & Malaka, R. (2023). Dungeons & replicants II: Automated game balancing across multiple difficulty dimensions via deep player behavior modeling. *IEEE Transactions on Games, 15*(2), 217–227. https://doi.org/10.1109/TG.2022.3167728

Safadi, F., Fonteneau, R., & Ernst, D. (2015). Artificial intelligence in video games: Towards a unified framework. *International Journal of Computer Games Technology, 2015*, Article 271296. https://doi.org/10.1155/2015/271296

Salmaze, P. H., dos Santos Pereira, G., de Góes, L. M. G., Ranieri, C. M., Toledo, C. F. M., Silvestrini, M. M., & Pereira, L. T. (2026). Narrative and dialogue generation for video games: A systematic mapping. *Engineering Applications of Artificial Intelligence, 178*, Article 115041. https://doi.org/10.1016/j.engappai.2026.115041

Summerville, A., Snodgrass, S., Guzdial, M., Holmgård, C., Hoover, A. K., Isaksen, A., Nealen, A., & Togelius, J. (2018). Procedural content generation via machine learning (PCGML). *IEEE Transactions on Games, 10*(3), 257–270. https://doi.org/10.1109/TG.2018.2846639

Togelius, J., Yannakakis, G. N., Stanley, K. O., & Browne, C. (2011). Search-based procedural content generation: A taxonomy and survey. *IEEE Transactions on Computational Intelligence and AI in Games, 3*(3), 172–186. https://doi.org/10.1109/TCIAIG.2011.2148116

Yannakakis, G. N., & Melhart, D. (2023). Affective game computing: A survey. *Proceedings of the IEEE, 111*(10), 1423–1444. https://doi.org/10.1109/JPROC.2023.3315689

Zakaria, Y., Hadhoud, M., & Fayek, M. (2021). *Procedural level generation for sokoban via deep learning: An experimental study.* Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.36227/techrxiv.16640095.v1

Zohaib, M. (2018). Dynamic difficulty adjustment (DDA) in computer games: A review. *Advances in Human-Computer Interaction, 2018*, Article 5681652. https://doi.org/10.1155/2018/5681652

IEER
Institute for Excellence in Education & Research