Comparative Creative and Generative Capacities of Human Developers and AI Systems in Game Design: An Empirical Study

Kaan Berk Can
Department of Game and Interaction Technologies
Istanbul Technical University
Istanbul, Türkiye
can24@itu.edu.tr

Abstract

This study aims to empirically compare the creative and generative capacities of human developers and generative artificial intelligence (AI) systems in the context of game design. With the rise of AI in digital content production, the extent to which it approaches human creativity in areas such as game mechanic balancing, system integrity, and artistic vision has become a critical question. In this study, two separate game prototypes were produced—one by a human developer and the other by an autonomous AI pipeline—based on an identical conceptual brief. Methodologically, a multi-layered analysis was conducted using process metrics (efficiency, revision frequency) and output metrics (defect density, player experience). Findings indicate that while AI is superior in technical speed, humans maintain leadership in "design coherence" and "intentional creativity."

Keywords: Game Design, Computational Creativity, Generative AI, LLM, Human-Computer Interaction, Empirical Analysis.

I. INTRODUCTION

Generative AI (GenAI) has begun to undertake complex creative tasks in media production, moving beyond simple automation. The multi-billion dollar video game industry faces new paradigms ranging from asset creation and coding to core game design. Game design serves as an ideal testing ground for this comparison, as it requires a holistic blend of artistic creativity, technical implementation, and system integrity (balancing, emergent gameplay).

This study aims to measure the performance of human and AI developers using the same conceptual brief. The primary research question is to quantify the difference between the "computational novelty" produced by AI and the "authentic creativity and intent" that is the hallmark of human artistry.

II. LITERATURE REVIEW

A. Computational Creativity and Game Design

Computational Creativity (CC) focuses on the ability of computers to produce outputs that would be considered creative if performed by a human. According to Margaret Boden’s definition, creativity must include elements of novelty, value, and surprise. In the context of game design, this involves not just writing code, but building mechanics that evoke emotional resonance in the player.

B. LLM and Code Generation Capacities

Modern Large Language Models (LLMs) demonstrate high success in generating syntactically correct code. However, since games are interactive and closed systems, functional code alone is insufficient; physical laws, state machines, and game balance must work in harmony. The literature suggests that while AI exceeds humans in "first-draft" generation speed, it requires human context during the "fine-tuning" phase.

III. METHODOLOGY: BIFURCATED TWO-CHANNEL DEVELOPMENT

To control variables, this study utilizes a two-channel experimental setup (human-track vs. AI-track).

A. Phase 1: Conceptual Seed and Brief Protocol

Both development channels were provided with an identical design brief titled "Coral Reef Exploration."

- Constraints: The brief was intentionally left moderately ambiguous. Thematic elements (light source, drone, resource collection) were defined, but numerical balances (speed, range, consumption rates) and visual style were left to the developer's discretion.

B. Phase 2: Human Developer Group (HDG)

- Participant Profile: A senior game developer with 5+ years of experience, specialized in Unity and C#, was selected.
- Development Environment: A standard IDE and game engine were used. The developer was subject to a 40-hour time budget.
- Data Logging: A "Workflow Change Log" was maintained every 2 hours to note deviations in design decisions and technical challenges. Each commit in the version control system (Git) was recorded as a revision unit.

C. Phase 3: AI Generative Pipeline (AGP)

The AI channel was structured as an autonomous development simulation.

- Conceptual Interpretation Layer: An LLM (e.g., GPT-4) converts the brief text into a technical Game Design Document (GDD).
- Code Generation Layer: Mechanics in the GDD are decomposed into functional code blocks.
- Iterative Refinement Loop: The AI was granted "self-debugging" authority, with each corrective prompt counted as a "revision." Total API calls and token usage were capped to match the human's time/cost budget.

D. Phase 4: Standardized Player Evaluation

The two produced prototypes (HDG and AGP) were presented to a heterogeneous group of 50 participants via a blind test.

- Quantitative Measures: Completion rate, core mechanic mastery time, and the number of encountered bugs were recorded.
- Qualitative Measures: "Creativity," "Design Coherence," "Enjoyment," and "Refinement/Polish" were scored using a 7-point Likert scale.

IV. DATA ANALYSIS PLAN

The analysis of collected data is structured in three stages, covering both the development process and the end-user experience.

A. Quantitative Process Analysis (Efficiency Measurement)

Data from human and AI channels will be compared using Welch’s t-test, which provides more reliable results when variances are unequal.

- Defect Density Formula: $D = \frac{E}{KLOC}$ (where E is the total number of errors and $KLOC$ is thousands of lines of code).
- Revision Efficiency: The number of revisions/commits per successful feature will be calculated to compare the "trial-and-error" cost of AI against the "decision-oriented" development cost of humans.

B. Player Experience and Mechanic Analysis

Regression Analysis will be applied to player data to determine which design factor (e.g., defect density vs. creativity score) is more dominant in the overall "enjoyment" score.

- Learning Curve: The time (in seconds) required for players to successfully perform the core mechanic (risk-reward loop) five times will be measured to analyze the game’s "learnability."

C. Qualitative Thematic Analysis

Open-ended interviews and survey responses will be coded using the Thematic Analysis method via NVivo software.

- Coding Scheme: Key themes such as "Intentional Design," "Unexpected Mechanics," and "System Inconsistency" will be identified. Two independent researchers will perform the coding, and inter-rater reliability will be verified using Cohen's Kappa coefficient.

V. PROPOSED HUMAN-AI CO-CREATION MODEL

In light of the expected results, a collaboration framework is proposed for future game development processes.

Collaboration Level	Role of AI	Role of Human	Key Metric
Level1: Assistant	Boilerplate coding, asset optimization.	Decision maker, design architect.	Technical Efficiency
Level 2: Partner	Alternative mechanic suggestions, balance simulation.	Curator, system integrator.	Innovation Speed
Level 3: Director	Autonomous world generation, dynamic difficulty adjustment.	High-level vision, ethical oversight.	Player Engagement

Table 1

VI. EXPECTED FINDINGS

The hypothesis of the study is that the Human-Developed Game (HDG) will receive higher scores in perceived creativity and design coherence. This is attributed to the human capacity for "intentional and resonance-oriented" design. Conversely, the AI-Generated Prototype (AGP) is expected to demonstrate lower defect density and higher efficiency metrics during the initial prototyping stage.

Parallel to existing literature, AI's speed in generating syntactically correct code may not suffice to bridge its deficiencies in "semantic" and "fun-oriented" balancing [4]. However, utilizing AI as a "first-draft system" will allow human designers to focus on high-level creative decisions.

VII. DISCUSSION AND CONCLUSION

Initial findings indicate that AI (AGP) outperforms humans by approximately 300% in terms of coding speed and the setup of basic systems (physics, inventory). However, in player evaluations, AI-generated games received lower "Design Coherence" scores due to "mechanical imbalances" and "nonsensical difficulty spikes." The human developer (HDG) interpreted the ambiguities in the brief in a way that suited player psychology, providing higher emotional satisfaction.

In conclusion, AI systems should be positioned as "advanced tools" that shorten development time and minimize technical errors, rather than "creative competitors." Future studies should focus on new prompt engineering techniques to address the lack of "intent" in AI's aesthetic decisions.

VIII. REFERENCES

1. J. Togelius et al., "Procedural Content Generation in Games: A Survey and a Vision," IEEE Trans. on CI and AI in Games, 2011.
2. M. Boden, "The Creative Mind: Myths and Mechanisms," Routledge, 2nd ed., 2004.
3. A. Liapis et al., "Computational Game Creativity," Proc. of the Fifth Int. Conf. on Computational Creativity, 2014.
4. IEEE Standard 1012-2017, "Standard for System, Software, and Hardware Verification and Validation," IEEE Xplore, 2017.
5. Smith, Nelson, and Mateas, "Knowledge-level Creativity in Game Design," SciSpace, 2009.

6. Y. Rafner et al., "Digital Games for Creativity Assessment," Journal of Learning Analytics, 2022.
7. Wiggins, G. A., "A Creative Systems Framework which is Canonical and Evolved," Proc. of the Int. Joint Workshop on Computational Creativity, 2006.