
A Note on Generative Games: Positioning, Progress and Prospects*

Department of Computer Science
Beijing Jiaotong University
Beijing, China 100044
jtsang@bjtu.edu.cn

Abstract

Recent advancements in large-scale foundation models have expanded generative AI's influence from static content creation to dynamic, interactive applications, with gaming as a prominent domain. This note introduces a novel genre termed *Generative Games*, which incorporates generative AI as a foundational mechanic. It begins by tracing the evolution of generative AI applications in games, then identifies three key features of generative games as real-time interactive scenarios, personalized and dynamic storylines, and autonomous, self-evolving character behaviors. Finally, it envisions a future where games drive advancements toward AGI and evolve beyond entertainment to support self-actualization.

1 Introduction

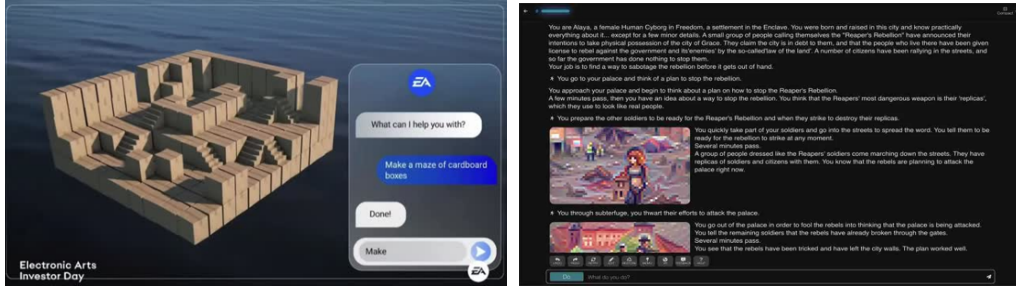
Recent advances in pre-trained foundation models and generative AI have expanded AI capabilities across a variety of media, evolving from static content like text, image to dynamic content such as music, video, human-like voice—and now to interactive content, with game being a typical example. The gaming industry, which is intrinsically related to computing, has always been embracing as well as pushing forward new AI technologies.

Prior to foundation models, conventional AI techniques were already enhancing game elements such as NPC dialogue, character voiceovers, and automated testing. Today, generative AI and large models have begun to reshape nearly every facet of game development, e.g., character prototyping, storyline [51] and level creation [48], game asset generation [18], and game data analysis [30]. Pilot efforts even propose to automate the entire game development process by employing multiple AI agents working in concert [11].

The above efforts employ AI as auxiliary tools, aiming to improve the efficiency of existing game production pipelines. [59] provides a detailed summary of the relevant literature. This note, however, is particularly intrigued by a novel genre of games that are fundamentally propelled by generative AI capabilities and designed to deliver AI-native gaming experiences, which is referred to as *Generative Game*. The following outlines three examples (illustrated in Fig. 1).

- *Electronic Arts (EA)* has recently unveiled a concept video titled "Imagination to Creation" [15], which demonstrates the potential of generating real-time interactive game scenarios. It showcases two players who collaboratively articulate their vision for a game world, which is then rapidly materialized into a cardboard maze and two gun-wielding characters. As the players navigate the emergent game world, they continue to expand upon it, illustrating the capability of users to create game assets from scratch using natural language prompts.

*The author is neither a practitioner nor a researcher in game. This note serves only as a summary of recent readings and personal thoughts from the perspective of an AI researcher. Incompleteness and inaccuracies may exist.



(a) EA's concept video "Imagination to Creation". (b) Text adventure game "AI Dungeon 2".



(c) "Eden Island" crafts autonomous and self-evolving NPCs.

Figure 1: Example generative games.

- "AI Dungeon 2" is a text adventure game crafting personalized and dynamic storylines [28]. It uses language models to generate narratives and outcomes, enabling players to perform any action they can articulate, with the AI dungeon master generating responses accordingly. This transcends the limitations of predefined storylines, offering a unique gaming experience where the story evolves in countless ways, leading to an infinite variety of adventures.
- "Eden Island" leverages generative AI to imbue its Non-Player Characters (NPCs) with autonomous and self-evolving behaviors [53]. Autonomous behaviors refer to the NPCs' ability to make decisions and take actions without direct player intervention. Self-evolving behaviors allow the NPCs to update themselves after being created and deployed, learning from experience and deliberately improving their performance. This creates a more realistic and immersive gaming environment, where NPCs are not just reactive but also proactive, exhibiting a level of agency that brings the game's world to life.

Definition 1. *Generative game* is a form of digital entertainment that harnesses generative AI techniques to create real-time interactive scenarios, personalized and dynamic storylines, and autonomous, self-evolving character behaviors.

Unlike traditional game+AI attempts that utilize AI primarily to enhance productivity and reduce cost on the production side, generative games introduce entirely new gameplay mechanics and experiences driven by AI. In these games, elements such as environments, narratives, and character behaviors are autonomously generated by AI on the fly, providing players with adaptive, unique, and immersive gameplay experience.

This note is by no means to provide a comprehensive survey. Instead, the following sections will first outline the evolution of generative AI's applications in games (Section 2), then highlight three key features of generative games (Section 3), and conclude with a discussion on current challenges (Section 4) and future prospects (Section 5).

2 The Evolution of Generative AI's Applications in Games

The advent of Generative Adversarial Networks (GANs) [20] has made it possible to generate highly realistic images. This note marks the refinement of GAN models as the starting point for the application of generative AI in games. Since then, the development can be broadly divided into

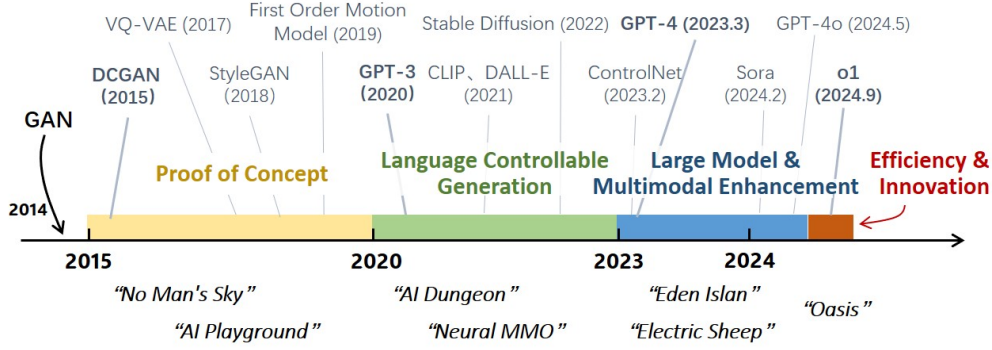


Figure 2: Evolution of Generative AI’s applications in games. In each phase, the representative AI models are listed above, with the typical games presented below.

four phases: Proof of Concept (2015–2019), Language-Controllable Generation (2020–2022), Large Model & Multimodal Enhancement (2023 to mid-2024), and Efficiency & Innovation (mid-2024 to present), as illustrated in 2. Among them, the first two phases primarily focused on image-related content synthesis. With the release of GPT-4 [36], there was a significant leap in large model capabilities, leading to the latter two phases entering the era of generative games.

2.1 Phase 1: Proof of Concept (2015–2019)

Generative Adversarial Networks (GANs) [20], introduced in 2014, marked a turning point in the development of generative AI. This phase witnessed the emergence of various GAN variants, establishing a solid foundation for generative content in games. Among them, DCGAN [40] was an early breakthrough, demonstrating stable and reliable performance in generating images.

Following DCGAN, more advanced GAN architectures were developed. CycleGAN [65] tackled unpaired image-to-image translation, enabling tasks like converting real-world photos into game-style images [22]. StyleGAN [24] introduced style-based controls, offering more granular manipulation of image features. One notable use case involved generating customizable character avatars for role-playing games [57].

In parallel, VQ-VAE (Vector Quantized Variational Autoencoder [55]) and its successor VQ-VAE-2 [42] gained traction. These models leveraged a discrete latent space, which proved particularly effective for tasks requiring temporal consistency, such as character animations and game cutscenes [16].

Another significant development was the First Order Motion Model [45], designed to animate still images by driving them with motion data extracted from reference videos. In games, this model enables dynamic animations for static characters, real-time expression generation for narrative games, and live avatar syncing for virtual streamers.

We can see that this phase primarily focused on proof-of-concept demonstrations, showcasing the feasibility of generative AI in games. The advancements highlighted how generative models could automate labor-intensive tasks like character creation, environment generation, and animation.

Despite these advancements, generative AI technologies were still in their early stages. No commercially released games relied predominantly on AI-generated content during this period. Games like “No Man’s Sky” [17] utilized traditional procedural generation techniques, where deterministic algorithms were used to create planets and ecosystems. Meanwhile, some experimental projects showcased the potential of generative AI. For instance, NVIDIA’s “AI Playground” [12] demonstrated real-time image generation using GANs, producing realistic images and textures suggesting future applications in games. “Project M” [32] utilized GANs for generating realistic character animations and facial expressions, paving the way for more dynamic interactions. However, these projects were not released as commercial games, serving mainly as proof for generative AI’s capabilities.

2.2 Phase 2: Language-Controllable Generation (2020–2022)

As data and model scales continued to grow, generative models underwent substantial advancements. In 2020, OpenAI released GPT-3 [8], a language model with 175 billion parameters, capable of generating human-like text. GPT-3 showcased the power of scaling laws, underscoring the transformative potential of large-scale foundation models which were then established as a recognized direction for advancing AI capabilities.

The introduction of CLIP [39] was a pivotal moment in the alignment of natural language and visual modalities. By training on a large corpus of text-image pairs, CLIP demonstrated the capability to understand visual content through natural language prompts. CLIP paved the way for the development of DALL-E [41], which, for the first time, allowed users to generate scenes or objects with natural language. This breakthrough introduced unprecedented flexibility in content creation, empowering developers to generate custom images directly from descriptive prompts.

Subsequently, diffusion models like Stable Diffusion [43] emerged, offering improved quality, speed, and flexibility in image generation. These models provided game developers with tools to generate high-fidelity content based on specific prompts, facilitating the creation of usable game environments and in-game assets.

CLIP, DALL-E and Stable Diffusion have found extensive applications in games. Developers have used these models to quickly generate diverse game assets [41]. For example, CLIP-based implementations seamlessly integrate visual content into interactive gameplay, significantly reducing development time and enhancing player immersion [39]. Techniques like DreamBooth further refine diffusion models for subject-specific generation [44].

During this phase, the advancements in language-controllable generation significantly broadened the scope of generative AI applications in games. By integrating natural language inputs into visual and interactive content generation, generative AI practically enhanced the game development pipeline, reducing production costs and time while maintaining high levels of quality.

Notable AI-driven games during this period exemplified the potential of generative AI technologies. Projects such as “AI Dungeon” [27] leveraged GPT-3 to deliver infinite narrative possibilities through text-based adventures. “Neural MMO” [47] incorporated large-scale generative models to create expansive, self-sustaining virtual worlds populated by AI-driven entities. These pioneering works demonstrated how generative AI could fundamentally transform game design, setting the stage for more AI-native applications in the future.

2.3 Phase 3: Large Model & Multimodal Enhancement (2023 to mid-2024)

The release of ChatGPT in late 2022 fully demonstrated the potential of generative AI, igniting its application across various fields. In the realm of language, GPT-4 [36] demonstrated capabilities approaching human-level performance, adeptly handling a wide range of tasks with fluency and coherence. In games, it has been utilized to enhance non-player character (NPC) interactions, e.g., [21] exploited GPT-4 in understanding and predicting player behavior in complex gaming scenarios.

In the realm of vision, ControlNet [64] introduced precise control mechanisms, making image generation more practical and reliable. The combination of ControlNet with Stable Diffusion has allowed for flexible conditional inputs to guide the image generation process, resulting in assets that meet fine-grained design requirements.

In early 2024, OpenAI introduced Sora [35], a text-to-video model capable of generating high-fidelity videos up to a minute long. This advancement significantly elevated the potential for video generation in games. Subsequently, DeepMind’s Genie [9] leverages unsupervised learning from unlabelled videos to generate action-controllable virtual worlds, enabling real-time interactive game environments where player actions dynamically shape the gameplay.

Following these developments, GPT-4o [33] emerged as a multimodal model capable of processing and generating text, images, and audio, fostering more imaginative and interactive gaming experiences. For example, [46] highlights GPT-4o’s proficiency in voice interactions, which can be applied to games by enabling voice-controlled gameplay and dynamic, voice-responsive NPCs.

During this period, the continually advancing capabilities of generative enabled AI-centric gameplay experiences. Developers leveraged these technologies to create adaptive narratives, dynamically generated multimodal scenarios, and intelligent NPCs that respond to player interactions in real-time.

“Eden Islan” [53] serves as a standout example for this phase, which leverages language models with human-like capabilities to drive NPC behavior and interactions. Another example is “Electric Sheep” [14], an art experiment that allows players to create and explore their own dreamworlds with AI-generated characters and murals. Players can tell Electric Sheep what they wish to dream about and talk to the virtual inhabitants in a 3D space. These attempts illustrate the transformative impact of generative AI, leading to a shift towards AI-native game design. Generative AI not only assists in content creation but also plays a central role in shaping the player’s experience.

2.4 Phase 4: Efficiency & Innovation (mid-2024 to present)

Recently, advancements in reasoning infrastructure and model miniaturization have significantly reduced the cost and further increased the feasibility of applying generative AI in games. One notable example is WonderWorld [62], a framework capable of generating interactive 3D scenes from a single image in less than 10 seconds on a single A6000 GPU.

GameNGen [54] and GameGen-X [10] have explored the potential of entirely generative game engines. GameNGen, powered by a diffusion model, enables real-time interaction with gaming environments, achieving over 20 frames per second on a single TPU. Further, GameGen-X facilitates high-quality, open-domain generation by simulating various game engine features, including dynamic environments and complex actions. “Oasis” [13] serves as a practical demonstration of GameNGen and GameGen-X. It applies generative AI to create a fully interactive game environment, where every frame is generated in real-time. Such works showcase how AI can handle complex game engine tasks, such as dynamic world-building and real-time responsiveness, delivering a seamless and immersive player experience.

The introduction of the o1 model [34] brought a substantial advancement in complex reasoning capabilities, unlocking new possibilities for innovation in generative games. Its sophisticated reasoning supports the development of highly personalized and dynamic storylines that adapt seamlessly to player choices, providing a unique narrative journey for each player. Moreover, o1 enhances the behavior of autonomous, self-evolving character agents, enabling NPCs to independently form goals, strategies, and relationships, resulting in a more immersive and lifelike game world. The launch of games leveraging o1-alike reasoning models is eagerly awaited in the very near term.

3 Key Features of Generative Games

Among the above four phases, the first two primarily revolve around visual generation, where generative AI serves as an auxiliary tool. The latter two phases, however, shift toward AI-native games, paving the way for generative games. The following discussion will center on the latter two phases, exploring the three key features that define this evolution.

3.1 Real-time Interactive Scenarios

This feature enables the generation of game assets and environments in real-time based on user interactions. By immersing players in dynamic, responsive worlds, it allows game elements to adapt seamlessly to player behavior and preferences. For instance, WonderWorld [62] demonstrates this capability through its interactive generation and real-time updates of online scenes, creating a continuously evolving gameplay experience.

GameGAN [25], introduced in 2020, was one of the earliest attempts to simulate games using generative models. By training on massive gameplay records from “Pac-Man”, it demonstrated the feasibility of recreating entire game environments without the need for traditional game engines or hand-coded rules.

In 2023, UniSim [26] extended these capabilities by focusing on video-based scene prediction. Through a supervised learning paradigm, UniSim could predict subsequent videos given a previous video segment and an action prompt. This approach advanced dynamic scene generation, enabling AI to model sequential events and transitions within a gaming context more effectively.

Building on this foundation, Genie [9], introduced in 2024 by DeepMind, explored unsupervised learning of actions directly from video data. Genie’s ability to generate action-controllable virtual worlds from unlabeled datasets reduced the dependence on manually annotated data, making it more efficient to create interactive and adaptive environments.

Further advancements came in the second half of 2024 with GameNGen [54], also from Google, which showcased playable simulations of video games powered by diffusion models. GameNGen positions itself as a pioneering generative game engine, pushing the boundaries of AI-driven interactivity and demonstrating the potential to replicate classic games in real-time.

The most recent work, GameGen-X [10], elevated these efforts by significantly enhancing the complexity and realism of generated environments and actions. With its ability to simulate diverse and lifelike scenarios, GameGen-X represents a major leap in creating open-domain game simulations.

DeepMind unveiled Genie-2 deepmind2024genie2 recently, further elevating the creation of interactive 3D environments to new heights. This model excels at generating complex and dynamic scenes from simple prompts, enabling real-time interaction and adaptability. Genie-2’s long-term environmental consistency ensures that objects and elements behave predictably, even when out of sight. Its versatility extends beyond gaming to research, development, and simulation training, making it a valuable asset for both creators and researchers in exploring new ideas and testing hypotheses within a realistic virtual space.

This progression illustrates a clear trajectory from early generative AI experiments to increasingly sophisticated models capable of simulating rich, interactive, and realistic games. Let’s speculate wildly. If the complexity and granularity of these models continue to advance, reaching the level of replicating the real world, it opens the door to creating a "second universe"—a concept akin to the Matrix system depicted in the movie “The Matrix”, where entirely immersive and autonomous virtual worlds become possible.

3.2 Personalized and Dynamic Storylines

In addition to influencing game scenarios, user interactions can also actively shape the storyline and plot progression. The narrative evolves through the interplay of player actions. Players drive the story forward through continuous exploration, where every choice can lead to different narrative outcomes, immersing them in an exciting and unpredictable adventure.

Relevant attempts can be broadly categorized based on the complexity of storyline generation and the depth of player-AI interaction. In games emphasizing passive-triggered dynamic narratives, the storyline evolves as a response to user actions, with AI generating contextually rich narratives. “Lasofa” [63], an interactive storytelling system integrated into a sofa, transforms conventional furniture into a storytelling medium. The narratives, dynamically generated by LLMs and delivered through advanced audio systems, provide an enveloping auditory experience, blending technology and storytelling within furniture design.

Games with moderate interaction involve player decisions that significantly influence the storylines. “Snake Story” [60] exemplifies this approach by allowing players to determine the storyline’s progression through path-based gameplay. Generative AI dynamically generates endings based on the chosen paths, ensuring a unique narrative outcome in each playthrough. Similarly, StoryVerse [58] positions players and AI as co-authors of the narrative. Player choices shape the storyline, while AI ensures coherence and depth.

In strongly interactive games, players deeply engage in the narrative creation process, working with AI to craft complex, adaptive storylines. In “1001 Nights” [49], players interact with GPT-4 to co-create personalized stories. “ChatGeppetto” [31] extends this by enabling continuous dialogue between players and virtual characters, with AI adapting the storyline in real-time to the player’s conversational inputs.

Games exploring infinite generation and open-ended exploration focus on using AI to unlock boundless possibilities. “Infinite Craft” [6] exemplifies this by using LLMs to dynamically generate the outcomes of player-created elemental combinations. This empowers players to explore a nearly infinite array of creative possibilities. Imagine enhancing a game like “80 Days” [23] with generative AI, where infinite routes are dynamically crafted based on player choices. Such a system would

transform each journey into a uniquely tailored adventure, allowing players to step into the wondrous world of Jules Verne’s novel.

A significant leap is anticipated from pre-defined branching narratives to adaptive, player-driven storytelling experiences. This will underscore the growing capability of generative AI to craft personalized and ever-evolving stories. Again, let’s imagine the possibilities. If narrative systems continue to advance, becoming capable of interpreting not only player actions but their emotions, intentions, and contexts, we could see the emergence of endlessly adaptive narrative worlds. It might move toward the immersive complexity like “Black Mirror: Bandersnatch”. In such a future, players would not simply navigate stories, but shape and inhabit boundless narrative spaces, creating a level of immersion that fundamentally transforms interactive entertainment.

3.3 Autonomous and Self-evolving Character Behaviors

This feature enables characters to exhibit behavior that is both logical and capable of evolution. Characters act as autonomous agents with realistic, logical actions while also self-evolve over time.

Actually, AI has been applied to NPC development for years. For example, NetEase’s “Justice Online” includes over 200 AI-driven NPCs capable of dynamic player interaction. Similarly, companies like Inworld have dedicated efforts to advancing AI NPCs, aiming to enrich game worlds with more immersive and responsive characters.

A seminal contribution is Stanford’s Generative Agents [37], which demonstrated that generative AI has matured beyond merely supporting dialogue to becoming a central gameplay mechanic. These agents simulate lifelike characters capable of logical reasoning and interaction, showcasing the potential for NPCs to become core elements of gameplay.

Following this, first-person talker games like “Suck Up” and “With You Til The End – Yandere AI Girlfriend Simulator” focus on interacting with a single LLM-driven NPC. In “Suck Up”, players assume the role of a charming vampire who must persuade homeowners to invite them inside without resorting to violence. The game leverages AI-powered characters capable of dynamic conversations, remembering past interactions, and recognizing disguises. In “Girlfriend Simulator”, players are confined by a ‘Yandere’—a young woman obsessively in love with them. The gameplay involves exploring her apartment or hacking her computer to uncover interests that can be used in conversations to build empathy.

For more complex implementations, “Eden Island” [53] exemplifies the integration of multiple interactive NPCs into evolving storylines and gameplay. This survival game features a rich ecological environment and allows players to trade resources with indigenous characters. These NPCs possess their own motivations, emotions, and decision-making processes, making their behavior autonomous and entirely self-driven.

In addition to enhancing NPCs, generative AI can also be used to drive player-controlled main characters, especially in life simulation games. For instance, Unbounded [29] utilizes generative models to create a sandbox environment where players can customize and guide their characters’ growth. This game also integrates the other two features of generative games: players can engage with a continuously evolving game scenario, and the dynamic storylines adapt to each player’s choices.

Similarly, GoodAI’s “AI People” [19] offers a sandbox experience where players create and engage with AI-driven characters. These characters exhibit behaviors such as learning, emotional responses, and goal pursuit. It is like an AI-driven version of “The Sims”, where your controlled characters could adapt dynamically, forming unique personalities, pursuing goals, and shaping their own developmental trajectories.

Generative AI has also enabled gameplay that connects character agents with physical spaces, creating new interactive possibilities. For example, “ChatWaifu” [52] turns popular game characters like *Ganyu* from “Genshin Impact” into computer-controlled conversational agents, bridging gaming worlds with AI-powered computer use. Similarly, miHoYo’s *Lumi*, a virtual idol and popular Bilibili livestreamer, showcasing how game characters can evolve into influencers beyond traditional gameplay. Additionally, game IPs like “The Walking Dead” on Facebook Watch have embraced live-streaming formats, blending interactive gameplay with real-time audience engagement. These innovations highlight the growing convergence of games, AI, and media.

In summary, in generative games, players can interact with digital personas, including both NPCs and player-controlled characters. These interactions enhance immersion, providing players with more authentic gameplay experiences. By allowing characters to exhibit realistic behaviors and evolve independently, this feature pushes the boundaries of traditional games.

Rich Sutton’s Alberta Plan [50] outlines a vision for developing continually learning agents that adapt autonomously to complex environments. It emphasizes lifelong adaptation, and scalable computational efficiency to create intelligent systems capable of sustained improvement over time. Looking ahead, further advancements could lead to experiences akin to “Westworld”, where digital characters and environments blur the line between fiction and reality, creating unparalleled depth and interactivity.

4 Challenges

Generative games, while showcasing significant potential, face a variety of technical challenges and inherent limitations that hinder their full adoption and integration.

4.1 Technical Challenges

- **Content Consistency**
 - *Visual Consistency*: Maintaining coherence in generated images, scenes, and animations is critical. Abrupt or unrealistic transitions can disrupt immersion and break the player’s connection to the game.
 - *Semantic Consistency*: Long-context coherence in generated text and stories remains a challenge. Ensuring that narrative and dialogue align with logical, extended contexts is essential to preserve storytelling quality.
- **Computational Cost and Real-Time Performance**
 - *High Computational Cost*: Training and deploying generative AI models demand extensive computational resources, significantly increasing development costs.
 - *Real-Time Performance*: Generating content dynamically in real-time can strain hardware, impacting game frame rates and fluidity. Handling complex tasks in this manner risks lag or stuttering during gameplay.
- **Stability** The integration of new characters or storylines into a generative framework can challenge the stability and coherence of the existing game structure, making updates and expansions difficult to manage without disrupting player experiences.

4.2 Limitations

- **Ethical and Social Concerns** Generative content risks perpetuating biases or producing harmful outputs, posing threats to social values and fairness. Ensuring the ethical deployment of generative AI is an ongoing challenge.
- **Constraints on Innovation** While generative AI excels in producing vast quantities of content, it often relies on patterns from existing data. This tendency may limit its ability to deliver genuinely innovative game concepts, characters, or narratives, potentially leading to homogenized content that lacks originality.
- **Complexity in Game Development** The inherent opacity of generative AI models makes controlling their output challenging. This lack of interpretability complicates the development process, increasing the difficulty for developers to refine and direct the game’s generative systems.

Addressing these limitations will require advancements in AI research, optimization of computational resources, and adherence to ethical guidelines. Despite these challenges, generative games remain a promising frontier with the potential to redefine the games landscape.

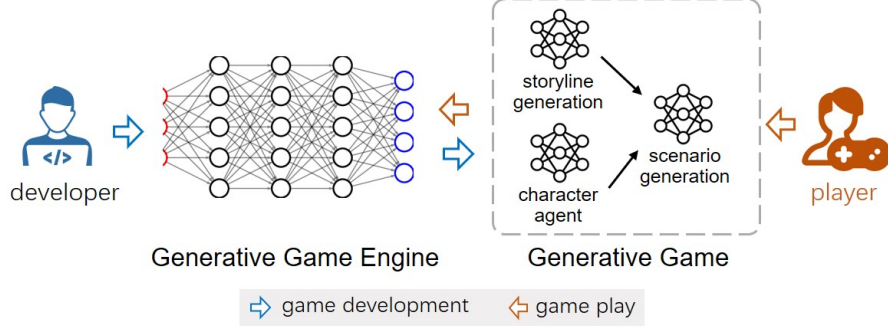


Figure 3: Generative game development and play.

5 Prospects

5.1 Generative Evolution of Game Development

One emerging trend is the democratization of game development. Platforms like Roblox [4] and Meowjito Games [5] exemplify this movement, offering low-code or no-code solutions that allow users without programming skills to turn their ideas into playable games. However, such platforms often come with low graphical quality. An alternative approach involves developing a NL2GE (Natural Language to Game Engine) module and leveraging existing game engines. For instance, DreamGarden [7] serves as an AI-driven design assistant for Unreal Engine, allowing users to create high-quality games by describing their idea in natural language.

Looking further, game engines may evolve from using generative AI as a tool to becoming fully generative systems. This shift could give rise to a new paradigm called *AI-native Game Development* and AI-native game engines called *Generative Game Engines*.

If we compare game engines to integrated development environments (IDEs) for programming, traditional engines like Unreal Engine or Unity resemble VS Code. While, generative game engines are more like Cursor [2]. Although Cursor is currently positioned as a copilot, requiring developers to possess a certain level of programming skills, as the model’s generative capabilities continue to improve, it is set to evolve into an agent, which is somewhat reminiscent of Bolt [1].

At that point, game developers would interact with a generative game engine supported by large-scale foundation models, using natural language to define game mechanics, as illustrated in Fig. 3. The games are generated according to high-level developer instructions, which exist in the form of game-specific and small-scale models derived from the game engine model. This aligns with an emerging model-generated-model research line. For example, Prompt2Model [56] and AutoMMLab [61] both explore the potential of transforming natural language instructions into deployable machine learning models.

Unlike previous works like GameNGen and GameGen-X relying on a single generative model, the envisioned framework would leverage three specialized models corresponding to the three key features of generative games:

- *Storyline Generation Model*: This model controls the narrative progression, ensuring long-term coherence and addressing the lack of memory seen in works like GameGen-X.
- *Character Agent Model*: Responsible for managing the behaviors of key NPCs and main characters, this model ensures autonomous and self-evolving character actions.
- *Scenario Generation Model*: Taking inputs from both the storyline and character agent models, this model generates the game environment dynamically, ensuring visual and functional coherence.

When players engage with a generative game, the game runs on the synchronization of the game-specific models and the game engine model. The game engine model provides the foundational functionalities necessary for the game, such as physics simulation, rendering, and system-level

interactions. Meanwhile, the game-specific models deliver tailored elements unique to the game, including storyline progression, character behavior, and scenario adaptation.

5.2 When Generative Game Meets o1: from Outer Shell to Inner Soul

The ongoing efforts have primarily focused on the first feature of generative games—real-time interactive scenarios—emphasizing dynamic environment generation and immediate system responses. However, the paradigm shift introduced by o1’s advanced reasoning capabilities holds the potential to revolutionize the other two features.

One transformative impact lies in enhancing narrative generation. Current generative models often struggle with maintaining semantic coherence over long contexts, leading to fragmented or shallow storytelling. With o1-alike reasoning models, generative games can achieve long-term narrative planning, where storylines adapt dynamically to player decisions while preserving logical consistency across branching paths.

Another change enabled by o1 is in character behavior, particularly in advancing the autonomy and evolution of agents. o1 introduces reasoning capabilities that allow for dynamic, context-aware decision-making over extended timelines. This enables characters to act not just reactively, but proactively, setting a new standard for lifelike digital agents in generative games.

Through these advancements, o1 shifts generative games from focusing solely on external dynamics to addressing deeper, internal complexities. By uniting logical depth with emotional resonance, o1 opens the door to a new era of generative games that blur the line between artificial constructs and soulful experiences.

5.3 Gaming towards AGI

Games and computer technology have been propelling each other forward. In the past, gaming demands drove advancements in GPU computing and network technologies. Early developments in AI also found their roots in games like checkers, chess, and different logical puzzles.

In modern AI, the role of games has remained pivotal. DeepMind’s Alpha series achieved its first major breakthroughs in the game of Go. Similarly, prior to the GPT series, OpenAI’s flagship projects included OpenAI Five and Dactyl, which explore paths towards AGI through virtual worlds (Dota) and the physical world (Rubik’s Cube) respectively. These successes highlighted games as controllable, cost-effective simulators that provide ideal environments for training and evolving AI agents.

Traditionally, reinforcement learning struggled to bridge the gap between toy simulations and real-world tasks due to the specificity of agent behaviors and the complexity of environmental modeling. Today, the advent of large language models (LLMs) has enabled agents to comprehend and execute natural language instructions, making them far more capable in addressing most practical tasks. Meanwhile, the recent developments in generative game engines have demonstrated the feasibility of modeling complex environments, adding an unprecedented level of realism and adaptability to simulated worlds.

The future envisions games evolving into digital twins and simulations of the real world. Explorations within these environments are expected to feedback into the real world. This convergence of games and generative AI holds massive potential to redefine the role of games, making them integral to solving some of humanity’s most complex challenges.

5.4 Beyond Entertainment: Games as a Pathway to Self-Actualization

The realization of AGI is expected to revolutionize societal structures, with automated production systems efficiently fulfilling most material needs. As traditional work paradigms become obsolete, humanity will face a profound shift in its priorities, transitioning from a focus on survival and productivity to a pursuit of higher-order goals. This aligns with Maslow’s hierarchy of needs, where industrialization and information technology have addressed physiological and safety needs, and the advancement of intelligent technologies opens the door to fulfilling humanity’s apex aspirations—psychological fulfillment and self-actualization.

In this future context, the role of games deserves a reexamination far from being mere entertainment. Actually, many games have already evolved into platforms for intellectual and cultural exploration. For instance, “Disco Elysium” delves deeply into sociological and philosophical questions, challenging players to reflect on identity, politics, and morality. Games like “Genshin Impact” and “Black Myth: Wukong” introduce players to rich elements of Chinese mythology and culture, fostering cross-cultural appreciation and understanding.

From an alternative perspective, entertainment is juxtaposed with work. In an era where AGI could automate most forms of labor, traditional work may no longer be necessary and the concept of pure entertainment may cease to exist. In “The Geometry of Wealth” [38], happiness is categorized into two types: experienced happiness and reflective happiness. Experienced happiness pertains to immediate, day-to-day satisfaction, often aimed at reducing discomfort and gaining a sense of control. Entertainment, in its traditional form, falls under this category, providing temporary relief and pleasure. Reflective happiness, on the other hand, relates to long-term fulfillment and purpose. It involves understanding how individual experiences contribute to broader life goals and self-actualization. In a post-work world shaped by AGI, the pursuit of reflective happiness could become paramount.

In this vision, games, with their unique ability to immerse, engage, and challenge players, can play a transformative role in self-actualization by providing two complementary paths: emotional simulation and value exploration. Emotional simulation allows players to experience and process complex emotions that are difficult to encounter or articulate in real life. Through carefully designed scenarios, games can evoke feelings of courage, empathy, trust, and self-worth. For example, “See-” [3], a public interest game, places players in the perspective of a visually impaired person, simulating the challenges of navigating daily life with limited sensory input. This fosters empathy and social awareness, allowing players to connect with others on a deeper emotional level.

Value exploration caters to higher-order cognition and reflection. Through deep narratives and interactive storytelling, games can encourage players to engage with ethical dilemmas, philosophical questions, and personal aspirations. Games like “Disco Elysium” offer players the opportunity to question political ideologies and personal identity. Games like “Civilization” engage players in contemplating broader concepts of leadership, societal progress, and legacy. Such games invite players to engage with broader, thought-provoking dimensions, guiding them toward intellectual and existential discovery.

Ultimately, as society moves towards a post-material and post-work reality, games have the potential to transcend their traditional role as entertainment. They can become mediums for introspection, growth, and the pursuit of purpose, pushing forward the joint evolution of society and technology.

References

- [1] Bolt.new. <https://bolt.new/>.
- [2] Cursor. <https://www.cursor.com/>.
- [3] See-. <https://www.taptap.cn/app/162251>.
- [4] Roblox, 2006.
- [5] Meowjito games, 2024.
- [6] Neal Agarwal. Infinite Craft. <https://neal.fun/infinite-craft/>, 2023.
- [7] Andrzej Banburski-Fahey. Dreamgarden: A designer assistant for growing games from a single prompt, 2024.
- [8] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.
- [9] J. Bruce et al. Genie: Generative interactive environments. *DeepMind Research*, 2024.
- [10] Haoxuan Che, Xuanhua He, Quande Liu, Cheng Jin, and Hao Chen. Gamegen-x: Interactive open-world game video generation, 2024.
- [11] Dake Chen, Hanbin Wang, Yunhao Huo, Yuzhao Li, and Haoyang Zhang. Gamegpt: Multi-agent collaborative framework for game development. *arXiv preprint arXiv:2310.08067*, 2023.
- [12] NVIDIA Corporation. Nvidia ai playground. 2018. Tech Demo, <https://www.nvidia.com>.

- [13] DecartAI. Oasis ai minecraft: Play game online demo, 2024.
- [14] Electric Sheep. Electric sheep’s ai garden robot, 2024.
- [15] Electronic Arts. Imagination to Creation.
- [16] SM Ali Eslami et al. Neural scene representation and rendering. *Science*, 360(6394):1204–1210, 2018.
- [17] Hello Games. No man’s sky, 2016. Game, <https://www.nomanssky.com>.
- [18] Jun Gao, Tianchang Shen, Zian Wang, Wenzheng Chen, Kangxue Yin, Daiqing Li, Or Litany, Zan Gojcic, and Sanja Fidler. GET3D: A generative model of high quality 3d textured shapes learned from images. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022.
- [19] GoodAI. AI People: Announcing the Next Evolution of Gaming AI NPCs. <https://www.goodai.com/ai-people-announcing-the-next-evolution-of-gaming-ai-npcs/>, 2024.
- [20] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13, 2014, Montreal, Quebec, Canada*, pages 2672–2680, 2014.
- [21] Jiaxian Guo, Bo Yang, Paul Yoo, Bill Yuchen Lin, Yusuke Iwasawa, and Yutaka Matsuo. Suspicion-agent: Playing imperfect information games with theory of mind aware gpt-4. *arXiv preprint arXiv:2309.17277*, 2023.
- [22] Xiaowei Huang, Xintian Fu, Li Wang, and Xiaojuan Sun. Game environment style transfer using cycle-consistent adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 0–0, 2019.
- [23] Inkle Studios. 80 Days.
- [24] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4401–4410, 2019.
- [25] Seung-Wook Kim, Hyeonwoo Lee, Philip Torr, et al. Gamegan: Learning to play pac-man and beyond. *NVIDIA Research*, 2020.
- [26] University AI Lab. Unisim: Predictive video continuation for interactive gaming. *Interactive AI Research*, 2023.
- [27] Latitude. AI Dungeon, 2019.
- [28] Latitude. AI Dungeon 2, 2024.
- [29] Jialu Li, Yuanzhen Li, Neal Wadhwa, Yael Pritch, David E. Jacobs, Michael Rubinstein, Mohit Bansal, and Nataniel Ruiz. Unbounded: A generative infinite game of character life simulation. *arxiv*, 2024.
- [30] Xiaoxu Li, Xiao You, Siyuan Chen, Pittawat Taveekitworachai, and Ruck Thawonmas. Analyzing audience comments: Improving interactive narrative with chatgpt. In *Interactive Storytelling - 16th International Conference on Interactive Digital Storytelling, ICIDS 2023, Kobe, Japan, November 11-15, 2023, Proceedings, Part II*, volume 14384 of *Lecture Notes in Computer Science*, pages 220–228. Springer, 2023.
- [31] Erica Lima and Others. Chatgepetto: Ai-powered interactive storytelling through dialogue. In *Proceedings of the IEEE Conference on Games*, pages 234–245, 2023.
- [32] NCSoft. Project m, 2018. Tech Demo, <https://www.ncsoft.com>.
- [33] OpenAI. GPT-4o. <https://openai.com/index/hello-gpt-4o/>, 2024.
- [34] OpenAI. Introducing openai o1-preview, 2024.
- [35] OpenAI. Sora: Openai’s text-to-video model. *OpenAI Research*, 2024.
- [36] OpenAI, Josh Achiam, Steven Adler, and et al. Sandhini Agarwal. Gpt-4 technical report, 2023.
- [37] Joon Sung Park, Joseph C. O’Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. Generative agents: Interactive simulacra of human behavior. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST ’23)*, 2023.
- [38] Brian Portnoy. *The Geometry of Wealth: How to shape a life of money and meaning*. Harriman House, 2018.
- [39] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

- [40] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2016.
- [41] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. *arXiv preprint arXiv:2102.12092*, 2021.
- [42] Ali Razavi, Aaron van den Oord, and Oriol Vinyals. Generating diverse high-fidelity images with vq-vae-2. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 14837–14847, 2019.
- [43] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-resolution image synthesis with latent diffusion models. *arXiv preprint arXiv:2112.10752*, 2022.
- [44] Nataniel Ruiz, Yuanzhen Li, Yossi Goldman, Ben Fineran, Zhenlin Li, Kevin Shih, Pat Hanrahan, and Dan Goldman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. *arXiv preprint arXiv:2208.12242*, 2022.
- [45] Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. First order motion model for image animation. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 7137–7147, 2019.
- [46] J. Smith et al. Human-ai collaboration supporting gpt-4o achieving human-level user feedback in emotional support conversations: Integrative modeling and prompt engineering approaches. *AI Research Journal*, 2024.
- [47] Joseph Suarez, Yilun Du, Clare Zhu, Igor Mordatch, and Phillip Isola. The neural mmo platform for massively multiagent research. In *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks*, 2021.
- [48] Shyam Sudhakaran, Miguel González-Duque, Claire Glanois, Matthias Freiberger, Elias Najarro, and Sebastian Risi. Prompt-guided level generation. In *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, pages 179–182, 2023.
- [49] Yuqian Sun, Zhouyi Li, Ke Fang, and Others. 1001 nights: Ai-driven co-creative storytelling in a generative game. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, pages 425–436, 2023.
- [50] Richard S. Sutton. The alberta plan for ai research. <https://arxiv.org/abs/2208.11173>, 2023. Accessed: 2024-11-21.
- [51] Pittawat Taveekitworachai, Mustafa Can Gursesli, Febri Abdullah, Siyuan Chen, Federico Cala, Andrea Guazzini, Antonio Lanata, and Ruck Thawonmas. Journey of chatgpt from prompts to stories in games: the positive, the negative, and the neutral. In *2023 IEEE 13th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, pages 202–203. IEEE, 2023.
- [52] ChatWaifu Team. ChatWaifu, 2024.
- [53] Tencent. Eden Island, 2024.
- [54] Dani Valevski, Yaniv Leviathan, Moab Arar, and Shlomi Fruchter. Diffusion models are real-time game engines. *arXiv preprint arXiv:2408.14837*, 2024.
- [55] Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 6306–6315, 2017.
- [56] Vijay Viswanathan, Chenyang Zhao, Amanda Bertsch, Tongshuang Wu, and Graham Neubig. Prompt2model: Generating deployable models from natural language instructions. *arXiv preprint arXiv:2308.12261*, 2023.
- [57] Kai Wang, Yan Zhou, Chen Li, and Lijuan Chen. Learning to generate game environments with stylegan and cycle-consistency loss. In *2020 IEEE Conference on Games (CoG)*, pages 1–8. IEEE, 2020.
- [58] Yuhui Wang and Others. Dynamic plot co-authoring in storyverse: Collaborative storytelling with ai. In *Proceedings of the ACM Conference on Interactive Storytelling*, pages 56–68, 2024.
- [59] Daijin Yang, Erica Kleinman, and Casper Hartevelde. GPT for games: A scoping review (2020-2023). In *IEEE Conference on Games, CoG 2024, Milan, Italy, August 5-8, 2024*, pages 1–8. IEEE, 2024.
- [60] Daijin Yang and Others. Snake story: Integrating path-based gameplay with dynamic narrative generation. In *Proceedings of the AAAI Conference on Interactive Digital Entertainment*, pages 123–134, 2024.
- [61] Zekang Yang, Wang Zeng, Sheng Jin, Chen Qian, Ping Luo, and Wentao Liu. Autommlab: Automatically generating deployable models from language instructions for computer vision tasks. *arXiv preprint arXiv:2402.15351*, 2024.
- [62] Hong-Xing Yu, Haoyi Duan, Charles Herrmann, William T. Freeman, and Jiajun Wu. Wonderworld: Interactive 3d scene generation from a single image. *arXiv preprint arXiv:2406.09394*, 2024.

- [63] Tongge Yu, Meizhu Chen, Ya Li, Deehsiao Lew, and Kejin Yu. Lasofa: Integrating fantasy storytelling in human-robot interaction through an interactive sofa robot. *HRI '24*, page 1168–1172, 2024.
- [64] Lvmin Zhang and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. *arXiv preprint arXiv:2302.05543*, 2023.
- [65] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision (ICCV)*, pages 2223–2232, 2017.