Lottery and Sprint: Generate a Board Game with Design Sprint Method on AutoGPT

Maya Grace Torii*
toriparu@digitalnature.slis.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Japan

Takahito Murakami*
takahito@digitalnature.slis.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Japan

Yoichi Ochiai
wizard@slis.tsukuba.ac.jp
University of Tsukuba
Tsukuba, Japan

ABSTRACT

In this paper, we introduce "Lottery and Sprint", a board game creation methodology that cooperates human design intuition with the structured Design Sprint framework executed by the AutoGPT system. By aligning AI-driven processes with human creativity, we aim to facilitate a collaborative game design experience. A user study is conducted to investigate the playability and enjoyment of the generated games, revealing both successes and challenges in employing systems like AutoGPT for board game design. Insights and future research directions are proposed to overcome identified limitations and enhance computational-driven game creation.

CCS CONCEPTS

• **Human-centered computing** → *Interactive systems and tools*; • **Applied computing** → **Arts and humanities**.

KEYWORDS

board game design, auto game creation, design sprint, generative AI

1 INTRODUCTION

Creating novel, enjoyable and effective board games typically requires a detailed understanding of game mechanics, player engagement, and strategic balance [5]. Inexperienced individuals often face challenges in designing board games that cater to various play styles, objectives, and constraints [11]. To address this, we present the "Lottery and Sprint" method. This approach allows human designers to work with the AutoGPT system—an AI model based on Generative Pre-trained Transformers (GPT)—wherein the Design Sprint framework is embedded into the AutoGPT's prompt mechanism. Through this method, the collaboration between human and AI leads to the creation of innovative and balanced board games.

---

*Both authors contributed equally to this research.

This research explores the application of the AutoGPT system in board game design, detailing our "Lottery and Sprint" approach. The process begins with AutoGPT generating a variety of game designs in a metaphorical "lottery", allowing users to select a favored design. Once a game is chosen, it enters a customized Design Sprint [8] process to refine and improve the game concept. The central research question of this study is whether the AutoGPT system, operating within the proposed framework, can effectively create board games that are enjoyable, playable, and captivating, even for individuals with limited experience in game design.

Our investigation encompasses a review of related works. We then present the system design using AutoGPT and customized Design Sprint. This is followed by a user study evaluating the generated board games in terms of playability and entertainment by Creativity Support Index (CIS) and qualitative analysis.

2 RELATED WORKS

Game creation tools have advanced significantly in recent years, with techniques like Monte Carlo tree search enabling automated game content generation [3]. However, these systems often focus on automation and may not effectively support human designers' creative ideation. For instance, GDL [15] employs complex formalisms that could hinder non-experts' intuitive expression of ideas [4], while Monte Carlo tree search in Regular Boardgames [9] primarily targets AI techniques rather than assisting designers. As a result, there is room for improved human-computer collaboration in the game design process.

In this context, AI-driven game design has introduced new possibilities for collaborative game creation using AI [7].

The rapid growth and advancements in GPT and Large Language Models (LLMs) have expanded creative processes across various domains [18]. For example, Buncho employs GPT-2 to enhance writers' enjoyment in crafting novels [12], and similar capabilities have been applied to adjust video game level difficulty [16].

Considering these developments, AutoGPT, an open-source project for automatic task decomposition and problem-solving, emerges as a promising candidate for creative endeavors [13]. Yang et al. conducted a comprehensive benchmark study on AutoGPT agents in real-world scenario simulations [17]. Building upon this foundation, our study aims to explore AutoGPT's potential in board game design through the "Lottery and Sprint" method. This approach seeks to facilitate collaboration between human designers and AI, with the goal of creating engaging and playable board games for individuals with limited design experience while maintaining a rigorous analytical perspective.

Figure 1: Overview of the "Lottery and Sprint" process, which is inspired by the Design Sprint framework. Step 1, the Lottery phase (Prompt 1), consists of (1) G1: AutoGPT surveying online board games, (2) G2: AutoGPT creating board games based on user constraints, (3) G3: Reviewing and refining the generated games, and (4) G4: Updating the board game according to the refining process. This process follows a cyclical pattern, inspired by the flexibility and iterative aspects of Design Sprint, and continues until the generated rules are bug-free. Then, in G5, the single user (GM) selects a game from the generated options then gives feedback to AutoGPT (Prompt 2). In Step 2, the process follows a similar pattern, with (1) G1: GM reading user feedback in detail, (2) G2: AutoGPT regenerating the selected game based on feedback provided by the GM, (3) G3: Reviewing and refining the game again, and (4) G4: Updating the board game according to the refining process. This cycle repeats until the game is ready for test play (G5) with multiple user groups (GM and players). In Step 3 and onward, the process repeats the pattern inspired by Design Sprint as in Step 2, with feedback from test play being incorporated into the game by the GM, leading to further refinement and updates with the help of AutoGPT.

Figure 2: Using the creation prompt, 15 board games were randomly generated in the same condition. The generated text was then evaluated through a TF-IDF test, followed by Cosine Similarity, and the results are presented.

3 SYSTEM OVERVIEW

3.1 Design Sprint and Its Integration

Design Sprint is a time-bound, structured process that aims to solve complex problems through design, prototyping, and testing ideas with users. Developed by Google Ventures, it is a five-day process where each day is dedicated to a specific phase: Understand, Diverge, Converge, Prototype, and Test [1, 8]. In the context of the AutoGPT board game creation system, the principles of Design Sprint are integrated into the architecture, enhancing the efficiency and effectiveness of the game design process. This integration can be mapped onto the four key stages of the system's architecture as follows:

1. Research existing board games (Understand and Diverge): In this stage, the system surveys existing board games, gathering insights and generating ideas, similar to the Understand and Diverge phases of Design Sprint. The system learns about diverse rule sets, game balance, and strategic depth to inform its game design process. 2. Create a board game with constraints (Converge and Prototype): This stage corresponds to the Converge and Prototype phases of Design Sprint. The system uses the insights gathered from the research stage to create a new board game based on given materials and constraints, akin to building a prototype for testing. 3. Reflect on game design (Test): The Test phase of Design Sprint is mirrored in this stage, where the system evaluates the created game based on usability, rule clarity, design consistency, strategic balance, and enjoyment. This serves as an internal evaluation mechanism for the system's output. 4. Update and iterate (Iterate): The final stage of the AutoGPT system emphasizes iteration, a core principle of Design Sprint. Informed by the reflections from the previous stage, the system refines and updates the game to ensure that all materials, instructions, and game elements are clear and consistent.

3.2 Prompts

In the AutoGPT board game creation system(Fig. 1) [1], the four key stages of the architecture are supported by two types of prompts, creation prompts and feedback prompts. These prompts guide the AutoGPT system during the design process and are divided according to their role within the system architecture. Creation Prompts play a role in the first step of the system architecture: they survey, create an initial draft of board game rules, reflect on the created game, and update the game based on that reflection.

Prompt 1: Creation prompts

(G1) In order to make an interesting board game, survey board games focusing on basic various rules of board game, how other games are keeping game balance, and how other games exist.

(G2) Create a board game with the material and constraint written in text file "<board game material and constraint>. txt". Output the game explanation including all information to play the game with Title, Materials, setup, gameplay rules, game ending conditions, board layout and design, the unique point, the enjoyable point and strategy to win the game as text file so you can memorize or refer any time.

(G3) Step–by–step reflect the game you have created and consider if there are any failure in following the materials and constraints, difficult instructions for human to keep track while playing the game(usability in games), unclear rule instruction for situation potentially to occur(rule design), failure in rule, inconsistent in between the winning strategy and intention of the game(board game design and game balance, game theory) and export your result in text file. If there are not any problems, write so. use_gpt4 for better reflection.

(G4) Reading the reflection, updating the game. Check if the game is perfect and ready to play by taking the step in (G3).

Feedback Prompts are involved in the later stages of the system architecture. In these stages, the system processes user feedback and updates the game accordingly: it reads the feedback, recreates the board game, reflects on the game, and then updates it.

Prompt 2: Feedback prompts

(G1) Read the board game rule which was created and written in "<Target board game>.txt", material and constraints defined in "<board game material and constraint>.txt".

(G2) You have feedback from players of the game in this file "< feedback>.txt". Read the file and recreate the game. output and save the reflection to the text file so you can refer any time.

(G3) Step–by–step reflect the game you have created by considering the materials used (only defined materials should be used), rules or instructions to be very clear for first time players to play, clear winning conditions. output and save the reflection to the text file so you can refer any time.

(G4) Reading the reflection, updating the game by cooperating with the reflections and output and saving the updated game to a text file so you can refer any time. Make sure to include all the rules in the text file.

3.3 System Output and Evaluation

As defined in the prompts (Prompt 1), the system generates board game rules including elements such as title, materials, setup, gameplay rules, game ending conditions, etc[1].

Our analysis shown in Fig. 2 evaluate the diversity of game rules generated by the lottery phase using the AutoGPT system. We generated 15 different game rules and analysed their similarity using NLTK for morphological analysis [10] and the TF-IDF method for text vector representation [14].

By using cosine similarity, which ranges from 0 to 1, to compare the TF-IDF representations of the game rules, we were able to evaluate the diversity of the generated content over randomness. Our results show a mean cosine similarity of 0.59, a standard deviation of 0.08, with minimum and maximum cosine similarity values of 0.39 and 0.76 respectively. These values suggest that there is at least a reasonable level of diversity in the proposed designs, given the limitations of the data and constraints provided to users, such as the limited number of players and items available.

4 USER STUDY

4.1 User Study Design and Methodology

The user study was designed as a two-game experiment. Each game consisted of a playtest and a survey. Participants wrote feedback in the format of listing "problems" and "requirements" as a set in their native language, and the study conductor then translated it into English for the AutoGPT system. The overall view of the user study is explained in Fig. 3.

At the beginning and end of each phase, participants completed a survey designed to gather quantitative and qualitative data on

Figure 3: The user study design: Step1-board game preference survey Step2-select a board game generated and give first feedback to the AutoGPT system. Step3-playtest the game the first time and answer post-game play survey. Step4-game master gives second feedback to the system. Step5-playtest the game the second time and answer post-game play survey, final survey and board game preference survey.

the AutoGPT system's usability, the playability of the generated games, and the overall enjoyment and engagement of the games.

In the user study, 12 participants were grouped into three teams of four: one volunteer Game Master (GM) and three players per team. The participants, with ages ranging from 20 to 27 years and a balanced representation of both genders, were recruited from the University of Tsukuba. Participants had varied board game experiences, ranging from less than a year to over a decade, with play frequencies from a few times a year to once a month. The GMs, responsible for guiding sessions, chose a board game from three generated rules, provided system feedback, and supervised playtesting. The players were responsible for playtesting the game, providing feedback, and participating in the discussion.

4.2 Data Collection

Data was collected through board game preference surveys, post-gameplay surveys, final surveys, Creativity Support Index (CSI) [2] for Game Masters (GMs), and user study conductor observations. This approach allowed us to gather insights into the usability, playability, and enjoyment of the AutoGPT-generated games. Surveys were administered to gauge participants' initial expectations and subsequent impressions. Post-gameplay surveys covered various aspects of game experience, while the final survey allowed participants to reflect on their overall experience. The CSI [2] was used to evaluate the level of creativity support provided by the AutoGPT system to the GMs. GM's feedback, survey responses, and conductor observations contributed to the data collection regarding the system's usability and game development process.

5 RESULT AND DISCUSSION

Overall Assessment of the AutoGPT System: The results of the final survey showed that participants provided positive evaluations of the AutoGPT board game creation system. In the Game Master (GM) category, high scores were recorded for Q1 (Median: 4.00, SD: 0.58), Q2 (Median: 4.00, SD: 0.00), and Q3 (Median: 4.00, SD: 0.00). Similarly, in the player category, high scores were observed for Q1 (Median: 4.00, SD: 0.87) and Q2 (Median: 4.00, SD: 1.45) in the final survey. These high scores indicate that the generated games

met participants' expectations and that the AutoGPT system was effective in creating board games that were generally well-received (Fig. 4).

The system scored an average of 77.58 (n=3) on CSI scores ranging from 0 (worst) to 100 (best). The highest factor was Collaboration with an average of 30.67, followed by Exploration with an average of 27.67. Collaboration evaluates if the system allowed other people to work with the user easily, and Exploration evaluates if the user could explore many different ideas, options, designs, or outcomes. Expressiveness had the lowest score of 5.29, evaluating if the user was able to be very expressive and creative while doing the activity. This result is supported by a comment from the final GM survey, where the respondent was asked to identify which aspects of the system they found convenient or supportive and any difficulties they encountered: "I can't predict what kind of instructions make what kind of difference."

Comprehensibility of Game Rules: In the post-gameplay survey, both GMs and players reported increased median scores for the understandability of game rules in the second game compared to the first game. GMs showed an increase in median score from 2.00 in the first game to 3.00 in the second game (SD1: 1.00, SD2: 1.00), and players displayed an increase in the median score from 2.00 to 3.00 as well (SD1: 1.33, SD2: 0.97). However, this improvement could be attributed to participants playing the game twice with similar rules. It remains unclear whether the generated game itself enhanced its comprehensibility (Fig. 4). Several participants provided qualitative comments, with one noting, "It would be better if the content was a bit more detailed." The lack of clarity in the generated rules was also identified as a potential factor affecting comprehensibility.

Strategic Elements and Fairness: The evaluation of strategic elements revealed a discrepancy between GM and player perceptions, with GMs reporting a higher score in the second game, increasing the median score from 3.00 to 4.00 (SD1: 1.53, SD2: 0.58). In contrast, players' evaluations remained unchanged, with their median score staying at 3.00 in both the first and second game (SD1: 1.00, SD2: 1.27). This suggests that the improvement in strategic elements, based on GM input and expectations, did not translate into an enhanced experience for the players. Similarly, evaluations of fairness

Figure 4: (A) Quantitative results of "post-gameplay survey" GM- Q1 Overall, the game was what I expected. Q2 the game rule was understandable. Q3 the rule documents were written for easy to graphic design. Q4 The game had enough strategic elements. Q5 the length of the game was appropriate. Q6 the rules of the game were fair. Q7 the game had the originality. Q8 the choice of theme for the game was what I expected. Q9 there was appropriate interaction with other players. Q10 I could prepare a contingent event. Q11 There was adequate player control elements in the game. **(B)** Player- Q1 I enjoyed. Q2 the game rule was understandable. Q3 the rule documents were written for easy to graphic design. Q4 the length of the game was appropriate. Q5 The game had enough strategic elements. Q6 the rules of the game were fair. Q7 the game had the originality. Q8 the theme of the game was great. Q9 The game had interaction with other players. Q10 The game had contingent events. Q11 the game had adequate control over what I could control in the game. Q12 I want to play the game again. Q13 I want to play games with friends and family. **(C)** Quantitative results of "final survey" GM- Q1 The board games output by GPT were fascinating. Q2 the board games output by GPT were as expected. Q3 comparing the first and second sessions, GPT's improved board game was as expected. **(D)** Player- Q1 the board games were fascinating Q2 Comparing the first and second sessions, the board game was improved. **(E)**Board of the board game drawn by the participant for playtest.

improved for GMs, with an increase in the median score from 2.00 in the first game to 4.00 in the second game (SD1: 1.73, SD2: 2.08); however, player evaluations did not exhibit a significant change between the first and second games (Median1: 3.00, Median2: 3.00, SD1: 1.27, SD2: 1.23). This suggests that only GMs were satisfied with the improvements in fairness while players did not share the same sentiment (Fig. 4). Player feedback, such as "The game balance isn't fully adjusted," supports this observation. Even with GMs having the authority to adjust rules, ensuring universally accepted fairness was challenging.

Originality and Game Theme: The AutoGPT system successfully delivered novel and original game concepts, as evidenced by consistently high scores provided by both GMs and players. In the post-gameplay survey, GMs recorded median scores of 4.00 for both games (SD1: 0.58, SD2: 0.00), and players reported median scores

of 4.00 for both games as well (SD1: 1.24, SD2: 0.87). In contrast, evaluations of game theme showed differing opinions between GMs, who perceived a positive improvement, and players whose assessment declined in the second game (Fig. 4). GMs offered qualitative comments such as, "It's unpredictable how instructions will alter the game," and "It brings ideas humans might not immediately consider." These remarks highlight the system's unique ability to introduce originality in game creation.

Contingency and Board Graphic Design: The decrease in player evaluations of contingency in the post-gameplay survey, with the median score dropping from 4.00 in the first game to 3.00 in the second game (SD1: 1.13, SD2: 1.30), and the increased importance placed on contingency in the board game preference survey indicate that contingent or random events play a crucial role in the overall enjoyment of generated board games (Fig. 5).

Figure 5: (A)quantitative results of "board game preference survey" Questions:Q1 Concepts Q2 Board graphic design Q3 Game themes Q4 Game strategy Q5 Communications Q6 speed of progression Q7 Easy rule Q8 Randomness Q9 Controllable many objects Q10 Fairness rule. p ** < 0.05. and (B) Average of weighted factors and total CSI result.

Furthermore, high-quality graphic design may not be as essential to board game enjoyment as initially thought (Fig. 4A,B,D). The evaluation of board graphic design in the "board game preference survey" (Q2) changed, with the median score decreasing from 4.50 (SD1: 0.89) to 4.00 (SD2: 1.22)(Fig. 5). Despite its simplicity, the hand-drawn board provided an engaging gaming experience. This finding aligns with Kieran Hicks et al.'s observation that visual embellishments do not always improve the user experience [6].

6 LIMITATION AND FUTURE WORK

Our study of the AutoGPT system for board game creation showed promising results. However, there are some limitations to address, and we suggest potential areas for future research. First, the system had difficulty creating complete game board designs, often resulting in issues with overlapping components and missing elements. Future work could focus on improving the graphical capabilities of the AutoGPT system for better board designs. Second, sometimes the system generated games with contradictory rules or materials not allowed by the given constraints. To make the games more playable, future research should focus on refining the system's ability to understand constraints and provide clear rules. Lastly, the limited working memory of the AutoGPT system prevented it from researching existing board games thoroughly. Enhancing its working memory by developing database system could help the system better understand existing games and create more engaging designs.

7 CONCLUSION

In conclusion, our study demonstrated the potential of the Auto-GPT system in generating engaging and original board games. By addressing limitations and refining the system's capabilities, it can better match players' expectations and deliver satisfying gaming experiences. Future work can focus on enhancing the system for various game types, contributing to the advancement of human-computer interaction in the gaming industry. Overall, the AutoGPT

system shows promise in revolutionizing the game creation process and offering enjoyable, well-rounded board games for diverse audiences and gaming preferences.

REFERENCES

[1] Carlos Araújo, Ivon Santos, E.D. Canedo, and Aletéia Araújo. 2019. *Design Thinking Versus Design Sprint: A Comparative Study.* 291–306 pages. https://doi.org/10.1007/978-3-030-23570-3_22

[2] Erin A. Carroll, Celine Latulipe, Richard Fung, and Michael Terry. 2009. Creativity Factor Evaluation: Towards a Standardized Survey Metric for Creativity Support. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition* (Berkeley, California, USA) *(C&C '09).* Association for Computing Machinery, New York, NY, USA, 127–136. https://doi.org/10.1145/1640233.1640255

[3] Guillaume Chaslot, Sander Bakkes, Istvan Szita, and Pieter Spronck. 2008. Monte-carlo tree search: A new framework for game ai. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 4. 216–217.

[4] Giuseppe De Giacomo, Alejandro Catala, Bistra Dilkina, Michela Milano, Senén Barro, Alberto Bugarín, and Jérôme Lang (Eds.). 2020. Ludii - The Ludemic General Game System. 325 (2020), 411–418. https://doi.org/10.3233/FAIA200120

[5] Van Eck, R. N., V. J. Shute, and L. P. Rieber. 2017. Leveling up: Game design research and practice for instructional designers. Pearson, New York, NY, 227–285.

[6] Kieran Hicks, Kathrin Gerling, Patrick Dickinson, and Vero Vanden Abeele. 2019. Juicy Game Design: Understanding the Impact of Visual Embellishments on Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Barcelona, Spain) *(CHI PLAY '19).* Association for Computing Machinery, New York, NY, USA, 185–197. https://doi.org/10.1145/3311350.3347171

[7] Ahmed Khalifa, Michael Cerny Green, Diego Perez-Liebana, and Julian Togelius. 2019. General Video Game Rule Generation. arXiv:1906.05160 [cs.AI]

[8] J. Knapp, J. Zeratsky, and B. Kowitz. 2016. *Sprint: How to Solve Big Problems and Test New Ideas in Just Five Days.* Simon & Schuster. https://books.google.co.jp/books?id=rV0JCgAAQBAJ

[9] Jakub Kowalski, Maksymilian Mika, Jakub Sutowicz, and Marek Szykuła. 2019. Regular boardgames. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 1699–1706.

[10] Edward Loper and Steven Bird. 2002. Nltk: The natural language toolkit. *arXiv preprint cs/0205028* (2002).

[11] Michal Mochocki and Piotr Milewski. 2018. *Game Design Curriculum White Paper 2.0.*

[12] Hiroyuki Osone, Jun-Li Lu, and Yoichi Ochiai. 2021. BunCho: AI Supported Story Co-Creation via Unsupervised Multitask Learning to Increase Writers' Creativity in Japanese. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 19, 10 pages. https://doi.org/10.1145/3411763.3450391

[13] Significant-Gravitas. 2023. Significant-Gravitas/Auto-GPT. https://github.com/Significant-Gravitas/Auto-GPT

[14] Amit Singhal et al. 2001. Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.* 24, 4 (2001), 35–43.

[15] Michael Thielscher. 2017. GDL-III: A description language for epistemic general game playing. In *The IJCAI-16 workshop on general game playing*. 31.

[16] Graham Todd, Sam Earle, Muhammad Umair Nasir, Michael Cerny Green, and Julian Togelius. 2023. Level Generation Through Large Language Models. In *Proceedings of the 18th International Conference on the Foundations of Digital Games* (Lisbon, Portugal) *(FDG '23)*. Association for Computing Machinery, New York, NY, USA, Article 70, 8 pages. https://doi.org/10.1145/3582437.3587211

[17] Hui Yang, Sifu Yue, and Yunzhong He. 2023. Auto-GPT for Online Decision Making: Benchmarks and Additional Opinions. arXiv:2306.02224 [cs.AI]

[18] Jingfeng Yang, Hongye Jin, Ruixiang Tang, Xiaotian Han, Qizhang Feng, Haoming Jiang, Bing Yin, and Xia Hu. 2023. Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond. arXiv:2304.13712 [cs.CL]