Generating Together: Lessons Learned from Developing an Educational Visual Novel with AI Collaboration

1st Samuel Shields
Computational Media
University of California, Santa Cruz
Santa Cruz, United States
samshiel@ucsc.edu

2nd Alexander Calderwood
Computational Media
University of California, Santa Cruz
Santa Cruz, United States
acalde24@ucsc.edu

3rd Shi Johnson-Bey
Computational Media
University of California, Santa Cruz
Santa Cruz, United States
ismajohn@ucsc.edu

4th Noah Wardrip-Fruin
Computational Media)
University of California, Santa Cruz
Santa Cruz, United States
nwardrip@ucsc.edu

5th Michael Mateas
Computational Media
University of California, Santa Cruz
Santa Cruz, United States
mmateas@ucsc.edu

6th Edward Melcer
Computational Media
University of California, Santa Cruz
Santa Cruz, United States
emelcer@ucsc.edu

2024 IEEE Conference on Games (CoG) | 979-8-3503-5067-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/CoG60054.2024.10645553

***Abstract*—Visual novels are a popular game genre for educational games. However, they often feature pre-authored plot structures that cannot dynamically adjust to the player's progression through learning objectives. Employing procedural storytelling techniques boosts plot dynamism, but this comes at the cost of needing a larger repository of content (dialogue and images) to support different learning progressions and objectives. In this paper, we present postmortem-style case studies describing the lessons we learned from attempting to integrate large-language models (LLMs) and text-to-image models into the development of an educational visual novel about responsible conduct of research. Specifically, we discuss our experiences employing generative AI in our dialogue, character sprite, and background image creation processes.**

***Index Terms*—Procedural Narratives, Interactive Narratives, Visual Novels, Educational Games, Game Production, Generative AI**

I. Introduction

Visual novels (VNs) are a popular game genre for building educational games [1]. VNs are interactive narrative games characterized by their use of dialogue presented within text boxes, multiple-choice buttons, 2D character art, and static background images [2]. What makes VNs suitable tools for educational games is the potential to align narrative progression with learning objectives. However, they often feature pre-authored plot structures that cannot dynamically adjust to the player's progression through learning objectives [1]. Thus, if players must repeat content to reinforce their learning, they will be required to revisit the same content in the same order. This limitation presents an opportunity to use dynamic procedural narrative techniques, which could provide different gameplay without strictly repeating content.

Procedural narrative is the process of "creating a narrative over time, where that narrative is not determined before the process begins" [3]. Procedural narrative games often suffer from requiring large amounts of authored content to provide players with varied and responsive gameplay experiences [3]. This predicament has canonically been referred to within the Interactive Storytelling Community as The Authoring Problem [4]. However, the proliferation of new generative AI technologies like large language models (LLMs) and text-to-image models such as Stable Diffusion has opened new opportunities for rapid content creation for text and images. Thus, it is critical to understand how these technologies might integrate within video game development to support narrative and artistic productivity.

Generative artificial intelligence has provided an alluring space for individuals producing games, as it has the potential to produce assets with high enough fidelity to use in published games at a much faster clip than traditional approaches [5]. This advantage, however, has increased the fear that these tools will be used to replace traditional game development roles such as writers and artists [6]. This perspective devalues the creative process and the unique expressivity and meaning-making that only humans can provide [7]. Conversely, there is a complementary fear that these tools may not prove useful as people attempt to employ them to produce actual games rather than speculative tools and demos [8]. The allure of generative AI is driven by a faith in the long-term quality, coherence, and steerability of these approaches that our presented case studies find unearned.

In this paper, we describe a production process that po-

This material is based upon work supported by the National Science Foundation under Grant No. 2202521. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

sitions game content creators and generative AI systems as collaborators in a mixed-initiative co-creative process, aiming to create content that could not be easily created independently [9]. Specifically, we provide postmortem-style case studies on how generative AI, particularly generative text and image models, were incorporated into an interdisciplinary game development team of writers, artists, and programmers. We outline this experience through our revamping of the interactive narrative serious game Academical [10] from a statically authored interactive narrative game to one authored for dynamic content. This process has involved supporting collaboration and procedural narrative authoring as writers and artists use different generative technologies for their tasks. While we found some success integrating generative artificial intelligence into the system, we ultimately required heavy human oversight to include generated content in our prototype.

This paper discusses our team structure and how we involve writers and artists to collaborate while leveraging generative technologies to support their creative process. We focus mainly on three small case studies: how we used Stable Diffusion for character sprites, our attempts to use Stable Diffusion for background image generation, and our prototype for leveraging LLMs to assist writers in creating dialogue utterances that change based on the speaker and social context. We close with a discussion on our limited success using these tools during the production process.

II. Related Work

A. Game development postmortems in academic research

This paper presents three case studies of our experiences in a postmortem style, similar to presentations at the Game Developers Conference. Specifically, our presentation of content follows the style of Ian Horswill's postmortem paper on the successes and failures of the MKUltra research game [11]. Horswill specifically defines the purpose of postmortems as highlighting "goals, technology, what went right, what went wrong, and what [we] should have done instead". Importantly, Horswill also notes that the point of utilizing this methodology in academia is not to document specific technical innovations, since those have been written about elsewhere, but rather to talk about why those innovations were insufficient and how we can apply those lessons to future games. Given the research focus of this paper, i.e., understanding the strengths and weaknesses of incorporating generative AI approaches into game development, we believe this methodology is ideal.

B. Mixed-initiative co-creative AI systems

Research on mixed-initiative co-creative AI systems focuses on developing workflows that allow humans and computational generative systems to work collaboratively to create things that neither could easily create independently [9], [12]. This workflow is where generative AI tools shine the best, as they can empower people to augment their typical workflow while still retaining a sense of ownership over their creations. Our case studies describe using generative AI to enhance production speed and creative writing exploration.

Fig. 1: A screenshot showing the interface of Academical 1.0

C. Generative AI in game development

Generative AI-powered mixed-initiative co-creativity tools have been the subject of significant investments from large technology companies. In 2023, Unity unveiled its suite of generative content tools for helping developers create 2D textures/sprites, character animations, game code, and interactive virtual characters. [13] Similarly, NVIDIA recently unveiled NVIDIA ACE for Games [14], a suite of tools to help game developers create animations and synthesize/process speech. Finally, Google announced Google Genie, an image-to-game model that generates entire games from input images [15].

There has been a significant quantity of recent work using large language models to address various game design tasks – for instance, generating interactive stories [16], 2D Game Levels [17], and quest generation [18]. Nascent research into generative AI for game asset generation has extended to dynamic asset generation during gameplay [19] (differing from our asset pipeline that includes human review or revision before publication).

III. Academical

A. Academical 1.0

The initial version of Academical was a serious game intended to teach Responsible Conduct of Research through an interactive narrative that presented ethical dilemmas that users could role-play through [10]. Academical was designed to have nine unique playable scenarios, each pertaining to a specific topic in Responsible Conduct of Research (RCR) [20], [21]. Each scenario in Academical centers on a conversation between two stakeholders in the RCR issue at hand, one of whom is controlled by the player by selecting dialog options (see Figure 1). Through these choices, the player will ultimately reach one of several possible endings, a subset of which represents successful navigation of the situation. Generally, these scenarios become more complex (and difficult to navigate) as the game proceeds. They are structured to show how seemingly obvious answers around questions of research ethics can be complicated by factors such as power dynamics and marginalized identities and experiences [10]. Prior research on Academical also found the game to successfully enhance moral reasoning, knowledge, and attitudes towards

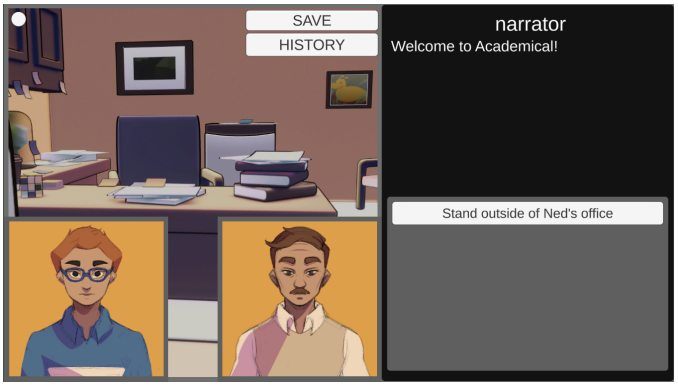

Fig. 2: A screenshot showing the interface of Academical 2.0.

RCR [22]. While featuring branching interactive narratives, we will refer to this form of Academical as "Academical 1.0" throughout this paper, as the artifact produced had a static approach to narrative content, constructing and displaying the same scenarios across scenes to the player.

Recent work on Academical has examined it through the lens of Self-Determination Theory [23] to better understand how certain elements of Academical's design influence different Self-Determination Theory factors and, subsequently, player experience, as well as how such elements could be changed to further improve the game [24]. Self-Determination Theory consists of three key factors that have been demonstrated to be potent drivers of enjoyment and engagement, i.e., Autonomy, Relatedness, and Competence. Specifically, autonomy relates to a "...sense of control or volition", relatedness concerns "...sense of social connectedness", and competence relates to "... a sense of mastery" [24]. Notably, while the "puzzle-like" scenarios for players to work through led to strong feelings of autonomy and competence, feelings of relatedness were notably low for players. This suggested that a procedural narrative approach to Academical's design could help players feel more connected to and engaged with the storylines and characters presented by Academical 1.0.

B. Academical 2.0

Given the success of Academical 1.0 over traditional RCR training methods, additional game features were scoped to help increase feelings of player autonomy, relatedness, and competence even further in a new version of Academical, i.e., Academical 2.0 (see Figure 2). Namely, we hypothesized that increasing the variety of conversational topics the player could encounter while more fluidly presenting pedagogical goals could help improve feelings of player-relatedness.

We also hypothesized that repeated sprite content, the 'spoke and hub' pattern [25], and players' limited options impacted feelings of player competence and relatedness in the static version of Academical. To this effect, we aimed to incorporate sets of non-repeating sprites for each player, making each traversal of the interactive visual novel unique, even when the same player repeats the same pedagogical goals. This heavily increases the authorship burden on artists and writers but also presents an opportunity to leverage modern tools in generative AI and interactive storytelling to increase the scope and coverage of sprite and narrative creation. Using a storylet architecture for the game also allows us to provide a greater volume and diversity of choices for each conversation, to improve the autonomy and relatedness players feel while playing the game.

IV. Production Environment

Academical 2.0, though still in active production, has provided enough experience to indicate which practices are working in our process, common pain points, and future production issues that need to be addressed. Our contribution is a series of case studies of integrating generative AI into artistic and writing practices to enhance productivity and maintain the integrity of creators' work.

This section describes the general team structure, key team activities and tools, and how we center "concept creation" as the root for all downstream, technical, and generative processes.

A. Lightweight Agile Team Structure

The contributors to Academical 2.0 roughly fell into one of the following categories with the following number of contributors: 3 writers, 1 2D/3D artist, 1 generative artist, 3 system engineers, and 1 producer. These roles are structured around a lightweight project management infrastructure that leverages the Manifesto for Agile Software Development [26] principles and project management techniques, including a Kanban board, writer's room, daily standups, and a recurring demo/feedback session. The reasoning behind this structure was twofold: it is responsive to the low-resource, high-intensity, and dynamic environment produced during academic work (value team members, interactions, and working code over processes and strict hierarchies), and it secondly enables work tracking and celebration of progress as team members complete major features. Table I presents an overview of key team members and activities described in the following sections.

Beyond meetings, team members engaged in a shared Discord server, where topics such as "What does Ned's office look like?" help drive interplay between writers and artists, enabling the production and system engineers to implement a singular vision of Nedness.

V. Concept Formation and Iteration

The focus on shared character, background, and world concepts ensures the team is on the same page across different production methods and that generative artists have a strong, coherent wellspring of content to inform their prompts and fine-tuning efforts. We call this development phase "Concept Formation," when collaboration between the teams across different modes of image and narrative content production leads the team to share an internal, fuzzy concept of a given character, plotline, or location. Concept formation is constantly iterated through team discussions, shared asset

TABLE I: Team responsibilities and meetings

Role	Responsibility	Stand-Up	Demo	Planning	Writer's Room
Writer	Authors branching narrative content	O	X	O	X
2D/3D Artist	Create character portraits, 3D renders of environments, and 2D captures of said environments	X	X	X	O
AI Artist	Works with artists and writers to augment and increase the volume of content delivery using Stable Diffusion and LLMs	X	O	X	O
Producer	Structures team meetings, manages Kanban board and other team tooling	X	X	X	X

"X" indicates recommended attendance

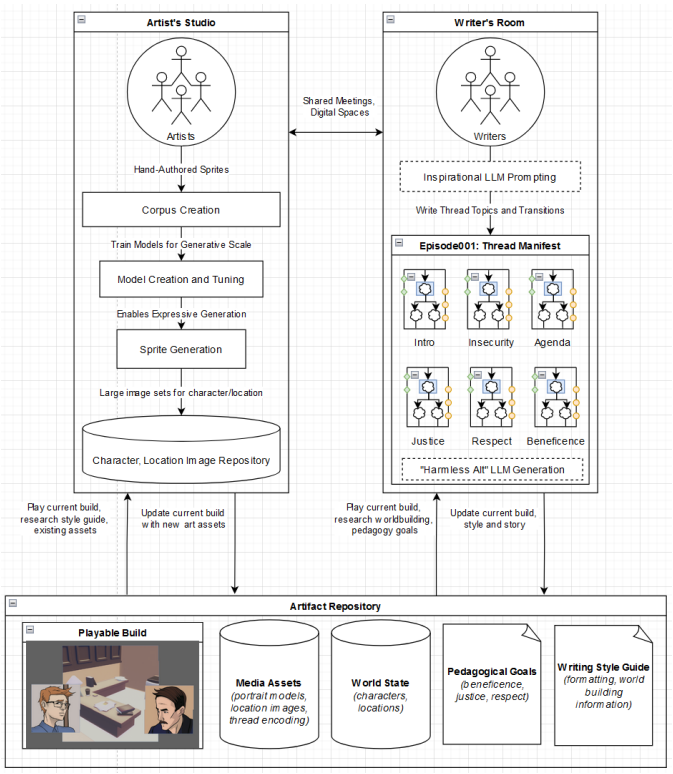

Fig. 3: Team interplay is critical to delivering a coherent and meaningful game artifact. Though artists and writers have unique development processes, exchange through shared digital spaces and meetings enable dialogues to occur that unify artistic and narrative components of Academical 2.0.

content across role domains, and feedback from demo sessions and presentations. An example of a type of artifact that deals with improving knowledge of character and location concepts is a mapping between character pose and feeling/emotion wheel, which creates a grid of the minimum required sets of hand-authored character images for a text-to-image model (Stable Diffusion 1.5) fine-tuning corpus. This approach was inspired in part by using a feeling/emotion wheel in Nic Junius' Puppitor project and allows writers to request new forms of character expression seamlessly and scope overall character expressivity for the authored story [27]. Figure 3 shows an overview of team structure and artifacts used during production.

VI. Case Study 1: Narrative Generation

This case study details how we structured our writer's room and integrated ChatGPT to inform and inspire authors on content creation. This process was introduced to accelerate authored content, as writing many dialogue utterances was required for the procedural and dynamically served content needed to create our prototype. We looked to generative AI techniques because the writer's room activity produced high quantities of structured content that could be used to scaffold prompts to return specific character utterance examples.

ChatGPT was able to generate utterances that met the high-level goals we specified, such as introducing topics or continuing the plot as desired. However, the final content needed high amounts of revision to be included in the production game. The overall style of the text was verbose and literal, despite our numerous iterations, offering too much framing text and not enough dialogue. When it did generate dialogue, it was stiff and did not follow game writing principles (such as verb-first choice labels, or direct over meandering dialogue). Despite this, the placeholder text it generated led to a quick revision loop, and encouraged the writers to arrive at each utterance with a prior understanding of its role within the broader game.

A. Writer's Room

Taking inspiration from the structure of the 'writer's rooms' typical in narrative television and game production, we instituted a biweekly writer's room. We recruited a group of undergraduate and master's students in games, literature, and HCI who received independent study course credit for participation. Assignments rotated through playtesting, storyworld ideation, story skeleton authoring, and storylet (or fragment) writing. Our writers, each experienced in branching narrative writing and unfamiliar with generative AI practices, were able to adapt writing skills from linear and branching narratives to our storylet framework (further explained in the next section).

B. StoryAssembler Affordances

Academical 1.0 was written in Twine, a choose-your-own-adventure game authoring platform. Authoring for the framework resembles manipulating sticky notes on a corkboard; player choice links are statically authored, as is common in hyperfiction.

Academical 2.0 makes use of StoryAssembler [28], a narrative engine based on the storylet pattern [29]. Users of the engine author a collection of fragments. A fragment is defined by its content, or text to be displayed upon visitation, and the choice label that will be displayed alongside other fragment's labels during the player's action. Each fragment has a set of preconditions that must be satisfied before the fragment can be displayed and effects applied once a fragment has been visited. Additionally, each fragment may be templated to allow varying instantiations of a given story beat.

Authors specify a list of story goals or desirable game states, and a forward search algorithm performs an exhaustive search through the space of possible player actions to select the allowable player actions that will lead to the most expedient satisfaction of the goals. The original StoryAssembler implementation was written in Javascript in 2019. Our reimplementation of the engine is in Step, a logic programming language for text generation [30].

Authoring content for a highly combinatorial storylet architecture like StoryAssembler, which allows many possible paths to reach a given fragment, is known to be challenging [31]. While the system allows hard-authored links between fragments (as in a traditional branching fiction engine such as Twine), the raison d'être of the system is not realized until the fragment count is high. Consequently, more art and text assets are needed than is typical, even for branching narrative experiences. This need corresponds to our search for an asset pipeline.

C. Non-repeating Beats

Failure is part of the learning process, and we expect that players will have to repeat parts of the game related to certain learning objectives. A premise of our work is that static story content has reduced player relatedness [24] compared with a system with (stateful) content diversity.

The maintenance of fragment preconditions as ground state variables means that each beat is paired with structured data ideal for LLM prompting. We used an early prototype prompt pipeline that works as follows: 1.) an LLM is provided with a narrative context (fragment content that is likely to have preceded the current fragment), as well as the values of the ground variables from the global blackboard. 2.) Text is generated to serve as a baseline content 3.) Writers revise the initial text. Often the text is far off the mark, but its distance from the appropriate text jumps the writers into the revision stage of writing. It is often discarded entirely, and very little of the LLM-generated text made it through our multiple rewrites. In particular, LLM-generated text was overly verbose compared to the needed affordances of a visual novel and often missed required plot points or embellished new ones. For example, when prompted for an exchange where Ned notes Brad's tardiness, the LLM produced the following text:

"Ned, you're late," Professor Ned remarked without looking up from the paper he was grading. His tone was casual, yet held a subtle scolding undertone. The clock on the wall behind him read 10:32, the digital numbers mocking Brad's tardiness. Ned had clearly not missed Brad's late arrival. His heart pounded in his chest, he knew he couldn't stall anymore. Squaring his shoulders, Brad shut the door behind him and prepared to face his professor."

Whereas the final edit from the writers had a substantially different tone, was more succinct, and listed the correct character addressing the other:

"Ah, there you are!" Ned says cheerfully. "I was beginning to wonder if we'd canceled this week."

VII. Case Study 2: Character Portrait Generation

Character expressivity in Academical is driven by the ability of character portraits to emote appropriately according to narrative content and the uniqueness of expression of the character throughout the story. From a sprite-authoring perspective, we needed to create and index a set of portraits for any given character. In Academical 2.0, we instead tasked the artist with creating a small, roughly 20-image corpus of representative sprites that can be used to fine-tune a generative system (Stable Diffusion 1.5) such that it can consistently produce character portraits across a wide range of expressions at a high volume of images. Figure 4 provides an overview of the portrait generation pipeline used for this case study.

A. Character Persona Building

For character portraits to remain consistent through both hand-authored and generated content, concept formation for characters must be performed and iterated throughout both writers and artists. This took the form of shared discord chat spaces and informal conversations throughout meetings, where team members would ask questions about characters to decide what relevant emotions and character properties needed to be represented in portraits. A simple example: when discussing an early episode of Academical 2.0, writers and artists sought to represent a subplot where Brad (and, by extension, the player) deals with insecurity around their advisor. The writers informed the artist of the need for a variety of nervous expressions of Brad, building up details about why Brad may express insecurity. This process informed the art team what portraits to create and created flavorful details for the overall story. As writers and artists constructed new subplots and episodes, the concept of Brad's persona emerged.

B. Corpus Creation and Model Tuning

For the generative artist, the above process created a corpus to tune Stable Diffusion for future image generation. We found good results with training sets of images between 15-20 images, providing a corpus for a Low Rank-Adaptation (LoRA) fine-tuning set for creating subsequent images that will eventually be used at runtime. These images were fed into the content-selection architecture of the game, enabling writers to leverage new character expressions without directly engaging in Academical 2.0's technical systems. To achieve consistency, fine-tuned models were trained in a single art style (A graphic-novel-inspired style with heavy inking on borders and bright, contrasting colors) and a single outfit and pose.

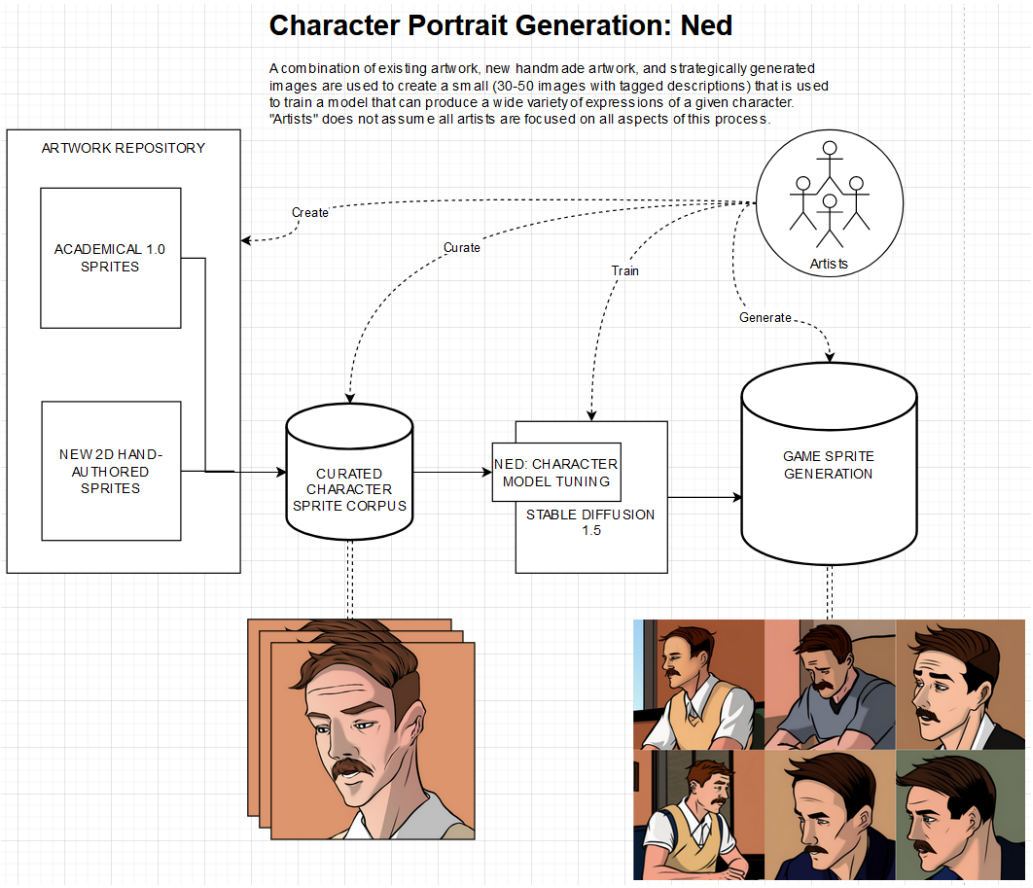

Fig. 4: An overview of the generative portrait process used to create portraits for character expression. These initial rounds of images were ultimately excluded for their inconsistency of style and perspective.

This means that the fine-tuning corpus consisted only of a single character wearing a single outfit and was always framed from the head and shoulders of the character. The limitations of pose and outfit ensure that we generate usable images for the overall game. That being said, images did require post-processing for pattern errors and slight irregularities from the desired output. Additionally, a drawback of this approach is that generation required many rounds of attempts to get usable images as the generation was non-deterministic and thus often generated images with artifacts or unrelated entirely. Another drawback was that curating and training the initial set of images took many days of tweaking. Table II provides an interview with a contributing artist concerning overall successes and failures associated with our approach.

VIII. Case Study 3: Location Background Generation

Similar to characters, location concept formation, and iteration were performed according to their appropriateness of the narrative. Writers and artists collaborated to create flavorful details around relevant locations. Our prototype episode centered on a character's office (Ned); as such, details generally related to how Ned's practice and interests would translate to details represented in their office. An example of this is the inclusion of a "brain in a jar" and "disorganized post-its" around the office of a Psychology professor. These details are documented in our team's shared chat and forum spaces, with a conversation between the creating artist and the writers driving the inclusion of flavorful properties.

A. Stable Diffusion and Backgrounds

While Stable Diffusion 1.5 provided a consistent representation of characters with additional training, initial testing of fine-tuning locations produced images with high levels of inconsistency. Objects were excluded or moved around a location, and spatial layout changed from image to image. In addition, perspective was often incorrect and objects had varying levels of scale that did not reflect a realistic scene. We evaluated that the tool was not appropriate for generating background images for these reasons - a production decision to avoid "shiny object syndrome" or the usage of a tool due to its fad usage across processes. Avoiding shiny object syndrome allows us to still meet the project's requirements without losing inordinate time to tuning a tool for processes that it does not excel at in the first place. [32] This decision still gave us the required expressivity for the system while reducing the overall time to deliver assets.

TABLE II: Responses from Contributing Artist on Stable Diffusion

Question	Answer
How would they have approached their task without using generative images?	I would have looked through a list of references of as many expressions as I could find, and I would adjust each sprite to create the desired expression. I would also work on new expressions for sprites as needed.
How have the generated images stacked up to their authored samples? Do they have to do any post-processing?	The generated images do have some inconsistency with patterns such as on clothing, and would require post-processing to fix that (if it was necessary to have consistent clothing or other accessories). The generated images' styles appear consistent for faces.
How have they had to adjust their approach to authoring to accommodate stable diffusion?	I've had to include different face angles for authoring to accommodate stable diffusion. I've worked on visual novel games in the past, and those sprites were drawn in only one angle (front facing).
Do they believe that they could be more productive without it from both a time and quality perspective?	For a game such as Academical that requires a ton of expressions and various facial angles, I do think that stable diffusion saves time, as rendering sprites is rather time-consuming. The quality of the generations appears to be acceptable - I believe a current issue with image generation is consistent patterns, while a major strength is style consistency. If the character designs were more complex and had a ton of patterns, then I think the amount of time for post-processing could be equal or more than the time spent hand-drawing.

IX. Discussion

A. Generative Techniques as Creative Augmentation Instead of Artist Displacement

A core theme of this project is that generative techniques are not strong replacements for core creative activities but are better at augmenting creative output to solve key tensions in traditional game development — such as authoring similar portraits for a single character. This is reflected in the literature where human-centric approaches to AI show more promise than granting total creative output to generative systems. [8] Some have discussed that generative AI is excellent at boosting brainstorming, concepting, and unblocking technical issues, but struggles with thematic consistency needed in game development. [33] In a study interviewing professional game artists, text to image generation systems were noted to be used mostly in inspirational and conceptualization stages of art, and could not produce production-level quality. [6] A case study of game proposals demonstrated that generative AI was a major boon to creating concept and prototyping assets, bolstering our findings that these systems are ideal for prompting imagination and game vision. [34] We show that emphasizing creative "bedrock," such as concept formation, creates an environment to safely use text-to-image models and LLMs to increase the expressivity of a system without replacing these critical roles in a game production pipeline. Framing these modern tools as being bent towards expressivity instead of being pitched as time/cost savings also helps assuage concerns that these tools are only intended to replace foundational roles in video game development.

A core takeaway from Stable Diffusion is that it did not accelerate the overall pace of development on the project. While it successfully produced new sprites for the game, the time to create a base set of images to train from, train models to use those images, and subsequent oversight from artists to release seemed to match a traditional art production pipeline. Additionally, while ChatGPT was useful for prompting ideas and establishing a baseline for how characters would speak, it consistently produced language that was too verbose and esoteric to be included in the final game outright. Human oversight and review seemed to be critical to producing quality content when using both of these systems.

B. Co-Creation for Professional Skills Development

Work produced during this project occurred in an academic environment, where available resources are low, and many contributors were focused on training a specific role (2D Art, Interactive Narrative Authoring, and so on). This lent well to a production environment that had siloed deliverables for each contributor while enabling the contributors to build out portfolio items relevant to their desired career trajectory. This development style in Serious Games has shown to be an effective mechanism for professional skills development, providing both explicit and implicit motivation to complete various tasks. The benefits of using serious games as professional skills development have been observed in the literature [35], and our shared tooling and development practices also ensured that all contributors across silos are exposed to a wide variety of production techniques and modern technologies.

C. Future Work

Due to the limitations encountered during our experiments with generative AI, future versions of Academical will lessen the use of generative AI in the production process. Additionally, introduction of a social simulation system will hopefully further increase ARC factors, specifically the feelings of relatedness that players feel to multiple characters. The scope of RCR training will be increased to include more topics than human subjects research - for example, peer review processes and plagiarism. Finally, a comparison study will be conducted between the initial version of Academical and the second version to investigate if the increased dialogue options and flexibility increase learning outcomes or ARC factors.

X. Conclusion

This paper provides a postmortem of an organizational paradigm incorporating generative AI for art and text generation into a video game production process. We reviewed how we used ChatGPT for prompting writers and used Stable Diffusion to generate character portraits. We also presented

a more traditional sprite generation process to cover location backgrounds and details. We demonstrate that while there are promising aspects to these production approaches, they are not full-displacement tools for contributors and that creative collaboration is key to producing meaningful content.

REFERENCES

[1] J. Camingue, E. F. Melcer, and E. Carstensdottir, "A (visual) novel route to learning: A taxonomy of teaching strategies in visual novels," in *Proceedings of the 15th International Conference on the Foundations of Digital Games*, ser. FDG '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3402942.3403004

[2] J. Camingue, E. Carstensdottir, and E. F. Melcer, "What is a visual novel?" *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CHI PLAY, oct 2021. [Online]. Available: https://doi.org/10.1145/3474712

[3] C. Martens and R. Cardona-Rivera, "Procedural narrative generation," Game Developers Conference, 2017, available online: https://www.youtube.com/watch?v=k2rgzZ2WXKo&t=8s&ab_channel=GDC (last accessed 19 November, 2023).

[4] C. Hargood, D. E. Millard, A. Mitchell, and U. Spierling, *The Authoring Problem: Challenges in Supporting Authoring for Interactive Digital Narratives*. Springer Nature, 2023.

[5] H. H. Jiang, L. Brown, J. Cheng, M. Khan, A. Gupta, D. Workman, A. Hanna, J. Flowers, and T. Gebru, "Ai art and its impact on artists," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 2023, pp. 363–374.

[6] V. Vimpari, A. Kultima, P. Hämäläinen, and C. Guckelsberger, ""an adapt-or-die type of situation": Perception, adoption, and use of text-to-image-generation ai by game industry professionals," *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CHI PLAY, oct 2023. [Online]. Available: https://doi.org/10.1145/3611025

[7] P. Parra Pennefather, "Ai and the future of creative work," in *Creative Prototyping with Generative AI: Augmenting Creative Workflows with Generative AI*. Springer, 2023, pp. 387–410.

[8] N. Anantrasirichai and D. Bull, "Artificial intelligence in the creative industries: a review," *Artificial intelligence review*, pp. 1–68, 2022.

[9] G. N. Yannakakis, A. Liapis, and C. Alexopoulos, "Mixed-initiative co-creativity," in *Foundations of Digital Games Conference 2014*, ser. FDG '14. Foundations of Digital Games, 2014.

[10] E. F. Melcer, K. M. Grasse, J. Ryan, N. Junius, M. Kreminski, D. Squinkifer, B. Hill, and N. Wardrip-Fruin, "Getting academical: a choice-based interactive storytelling game for teaching responsible conduct of research," in *Proceedings of the 15th International Conference on the Foundations of Digital Games*, 2020, pp. 1–12.

[11] I. Horswill, "Postmortem: Mkultra, an experimental ai-based game," in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 14, no. 1, 2018, pp. 45–51.

[12] A. Liapis, G. Smith, and N. Shaker, "Mixed-initiative content creation," *Procedural content generation in games*, pp. 195–214, 2016.

[13] M. Whitten, "Introducing unity muse and unity sentis, ai-powered creativity," *Unity Blog*, 2023. [Online]. Available: https://blog.unity.com/engine-platform/introducing-unity-muse-and-unity-sentis-ai

[14] A. Burnes, "Introducing nvidia ace for games - spark life into virtual characters with generative ai," *NVIDIA GeForce*, 2023. [Online]. Available: https://www.nvidia.com/en-us/geforce/news/nvidia-ace-for-games-generative-ai-npcs/

[15] J. Bruce, M. Dennis, A. Edwards, J. Parker-Holder, Y. Shi, E. Hughes, M. Lai, A. Mavalankar, R. Steigerwald, C. Apps, Y. Aytar, S. Bechtle, F. Behbahani, S. Chan, N. Heess, L. Gonzalez, S. Osindero, S. Ozair, S. Reed, J. Zhang, K. Zolna, J. Clune, N. de Freitas, S. Singh, and T. Rocktäschel, "Genie: Generative interactive environments," 2024.

[16] J. Freiknecht and W. Effelsberg, "Procedural generation of interactive stories using language models," in *Proceedings of the 15th International Conference on the Foundations of Digital Games*, 2020, pp. 1–8.

[17] G. Todd, S. Earle, M. U. Nasir, M. C. Green, and J. Togelius, "Level generation through large language models," in *Proceedings of the 18th International Conference on the Foundations of Digital Games*, 2023, pp. 1–8.

[18] S. Värtinen, P. Hämäläinen, and C. Guckelsberger, "Generating role-playing game quests with gpt language models," *IEEE Transactions on Games*, 2022.

[19] Y. Sun, Z. Li, K. Fang, C. H. Lee, and A. Asadipour, "Language as reality: a co-creative storytelling game experience in 1001 nights using generative ai," in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 19, no. 1, 2023, pp. 425–434.

[20] S. N. Seiler, B. J. Brummel, K. L. Anderson, K. J. Kim, S. Wee, C. Gunsalus, and M. C. Loui, "Outcomes assessment of role-play scenarios for teaching responsible conduct of research," *Accountability in Research*, vol. 18, no. 4, pp. 217–246, 2011.

[21] B. J. Brummel, C. Gunsalus, K. L. Anderson, and M. C. Loui, "Development of role-play scenarios for teaching responsible conduct of research," *Science and engineering ethics*, vol. 16, pp. 573–589, 2010.

[22] K. M. Grasse, E. F. Melcer, M. Kreminski, N. Junius, J. Ryan, and N. Wardrip-Fruin, "Academical: a choice-based interactive storytelling game for enhancing moral reasoning, knowledge, and attitudes in responsible conduct of research," in *Games and Narrative: Theory and Practice*. Springer, 2021, pp. 173–189.

[23] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The motivational pull of video games: A self-determination theory approach," *Motivation and emotion*, vol. 30, pp. 344–360, 2006.

[24] K. M. Grasse, M. Kreminski, N. Wardrip-Fruin, M. Mateas, and E. F. Melcer, "Using self-determination theory to explore enjoyment of educational interactive narrative games: A case study of academical," *Frontiers in Virtual Reality*, vol. 3, p. 847120, 2022.

[25] D. E. Millard, *Strange Patterns: Structure and Post-structure in Interactive Digital Narratives*. Cham: Springer International Publishing, 2022. [Online]. Available: https://doi.org/10.1007/978-3-031-05214-9_10

[26] M. Fowler, J. Highsmith *et al.*, "The agile manifesto," *Software development*, vol. 9, no. 8, pp. 28–35, 2001.

[27] N. Junius, *Puppitor: Building an Acting Interface for Videogames*. University of California, Santa Cruz, 2019.

[28] J. Garbe, M. Kreminski, B. Samuel, N. Wardrip-Fruin, and M. Mateas, "Storyassembler: an engine for generating dynamic choice-driven narratives," in *Proceedings of the 14th International Conference on the Foundations of Digital Games*, 2019, pp. 1–10.

[29] M. Kreminski and N. Wardrip-Fruin, "Generative games as storytelling partners," in *Proceedings of the 14th International Conference on the Foundations of Digital Games*, 2019, pp. 1–8.

[30] I. Horswill, "Step: a highly expressive text generation language," in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 18, no. 1, 2022, pp. 240–249.

[31] B. Samuel, J. Garbe, A. Summerville, J. Denner, S. Harmon, G. Lepore, C. Martens, N. Wardrip-Fruin, and M. Mateas, "Leveraging procedural narrative and gameplay to address controversial topics," in *International Conference on Computational Creativity*, 2017.

[32] L. Hills, "Management fads: How to resist shiny object syndrome." *Physician Leadership Journal*, vol. 9, no. 6, 2022.

[33] R. Panchanadikar, G. Freeman, L. Li, K. Schulenberg, and Y. Hu, ""a new golden era" or "slap comps": How non-profit driven indie game developers perceive the emerging role of generative AI in game development," in *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems*, ser. CHI EA '24. Association for Computing Machinery, pp. 1–7. [Online]. Available: https://dl.acm.org/doi/10.1145/3613905.3650845

[34] J. Lee, S.-Y. Eom, and J. Lee, "Empowering game designers with generative ai." *IADIS International Journal on Computer Science & Information Systems*, vol. 18, no. 2, 2023.

[35] S. Dini, L. Moreno, E. De Iaco, J. Zufferey, and W. Antoine, "7— co-creation of serious games to acquire a professional skill," *EXECUTIVE COMMITTEE*, 2023.