

The Generative AI Image Consistency Experiment

BRITAIN THOMAS*, Texas A&M University, USA

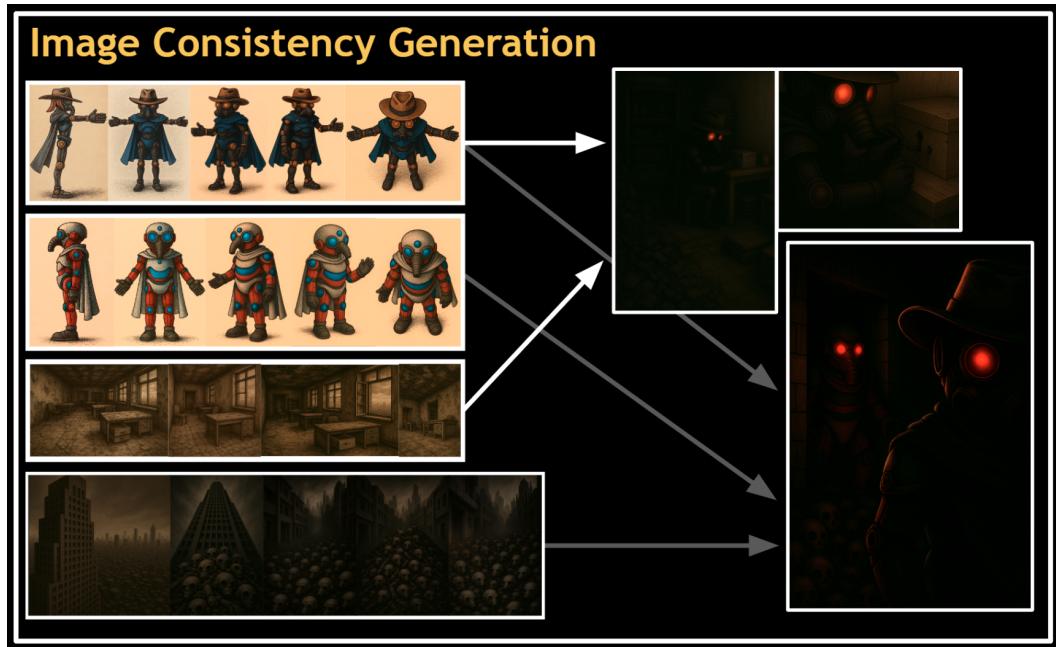


Fig. 1. The Final result of the application of all five phases to produce consistent images for storyboard creation.

The Generative AI Image Consistency Experiment focused on applying the concept of user defined prompts, prompt narrowing, prompt flooding, and prompt referencing. The application of these series of steps to produce consistent generated images, allowed for more cohesive storyboard-asset creation with consistent characters and environments. The goal of combining these approaches was to simplify and speed up the storyboard-asset creation process for more logical consistency between images. The application of user defined prompts proved significant in mitigating anomalies and illusions produced by the AI, and by applying prompt referencing along with this method, allowed for consistency between panels. Applying the concept of image-to-image based prompting, gave more successful results than traditional text-to-image only prompts.

*Both authors contributed equally to this research.

Author's Contact Information: Britain Thomas, britainthomas@tamu.edu, Texas A&M University, College Station, Texas, USA.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Conference acronym 'XX, College Station, TX

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-XXXX-X/2018/06

<https://doi.org/XXXXXXX.XXXXXXXX>

ACM Reference Format:

Britain Thomas. 2025. The Generative AI Image Consistency Experiment. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 8 pages. <https://doi.org/XXXXXXX.XXXXXXX>

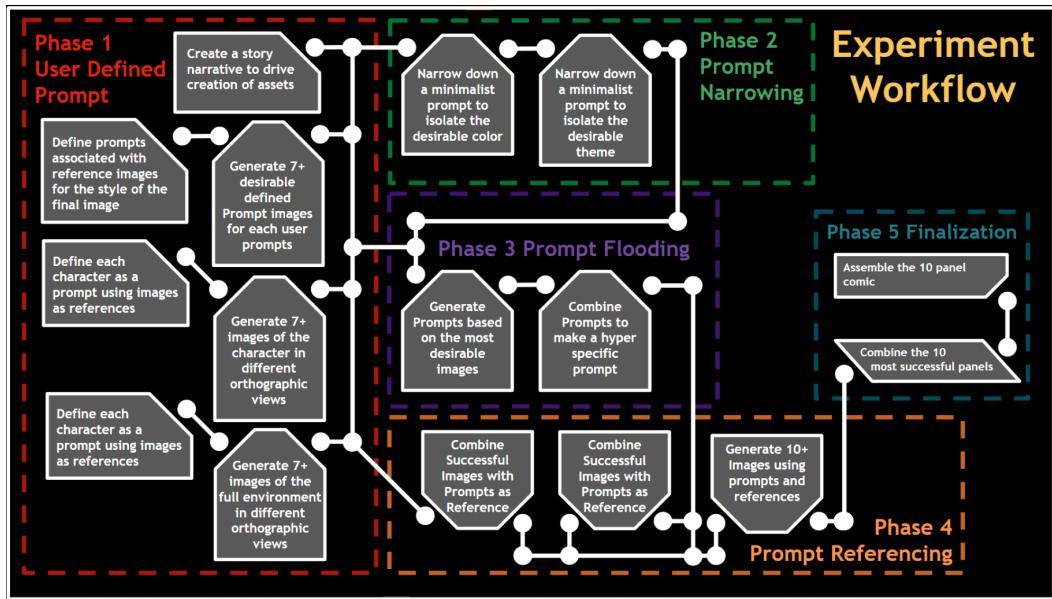


Fig. 2. The Experiment Workflow to generate the consistency comic through the use of the five phases of the experiment through user defined prompts, prompt narrowing, prompt flooding, prompt referencing to then be assembled in the the finalization process seen in Phase 5.

1 Introduction & Related Works

Consistent characters and environments give access to more application of AI assisted storyboard-asset creation, the concepts of applying user defined prompts, prompt narrowing, prompt flooding, and prompt referencing developed the way for consistent board creation. Traditional methods that do not utilize these series of steps and utilize some of the concepts in isolation show a room for improvement [3]. The benefit of OpenAi's DALL-E 3.0 model allows for the combination of all these steps to improve image quality being a transformer model when combined with ChatGPT o4 prompt generation. DALL-E 3.0's image generation suffers from inaccuracy, but maintains artistic abstraction between tests with the methods of combining text and image generation [2]; therefore, the benefit of combining the methods gives more room for appealing images at the cost of accuracy, but with the introduction the series of steps in the Generative AI Image Consistency Experiment paves way to consistency of these appealing artistic images once defined by the user and saved in memory. The benefit of DALL-E is its incorporation of ChatGPT and the functionality shared between the two AI models, because of this integration and the introduction of DALL-E 3.0 the quality of images the model produces are higher [1]. With the introduction of user defined prompts, once saved into memory, utilized in this experiment gave an increase in consistency while maintaining quality between the images.

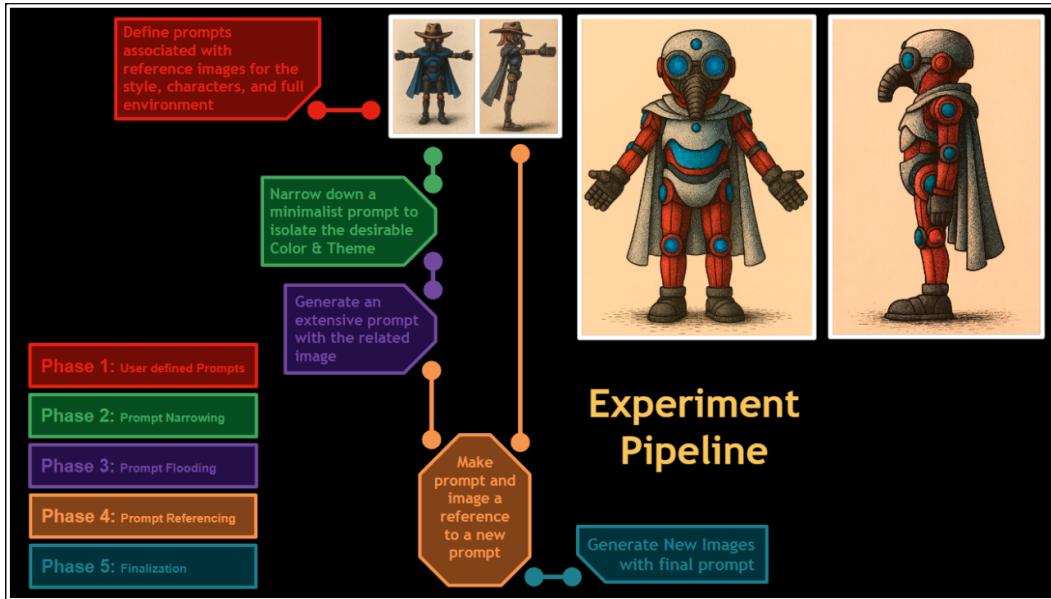


Fig. 3. The Experiment Pipeline used to generate the comic's consistency for each of the ten panels showcasing the steps of the entire experiment in a condensed view.

2 Methodology

The Generative AI Image Consistency Experiment consisted of a five phase process to generate each image, seen in figure 1. The workflow of the entire experiment, figure 2, showcases each step in creating the individual panels, while the pipeline is shown in a condensed format seen in figure 3. The process to enrich the desirable images seen in figure 1 are showcased in figure 4 of the process in action. Phase 1 of the experiment consisted of providing images created by the user to associate keywords for DALL-E to identify what the user is referring to when a prompt is stated by the user [figure 5]. Phase 2 of the experiment consisted of prompt narrowing to optimize prompts to its associated user prompt to produce the most successful and desirable output for the user, and each prompt is then saved in memory with its associated images [figure 6]. Phase 3 of the experiment consisted of prompt flooding, which takes the saved memory user defined and narrowed prompts to then hyper specify what the desirable image is, this is to mitigate illusions between future iterations in the prompt enrichment process seen in prompt flooding [figure 7]. Phase 4 of the experiment consisted of prompt referencing, which takes the final images and tweaks them for edits defined by the user for more desirable images [figure 8]. Phase 5 is the finalization step of the experiment where all images are then combined to fit a narrative finalizing the images [figure 9 10]. The final panel creation process is showcased in figure 11, along with all the edits and modifications needed to produce a final panel.

3 Result and Future Work

The final result of the experiment proved a successful strategy to generate consistent images for future and higher quality storyboard-asset creation. The introduction of user defined prompts seen in figure 3, gave editing and tweaking the final image easier when anomalies did occur. Without user defined prompts, the correcting process of images that had anomalies would drastically reduce

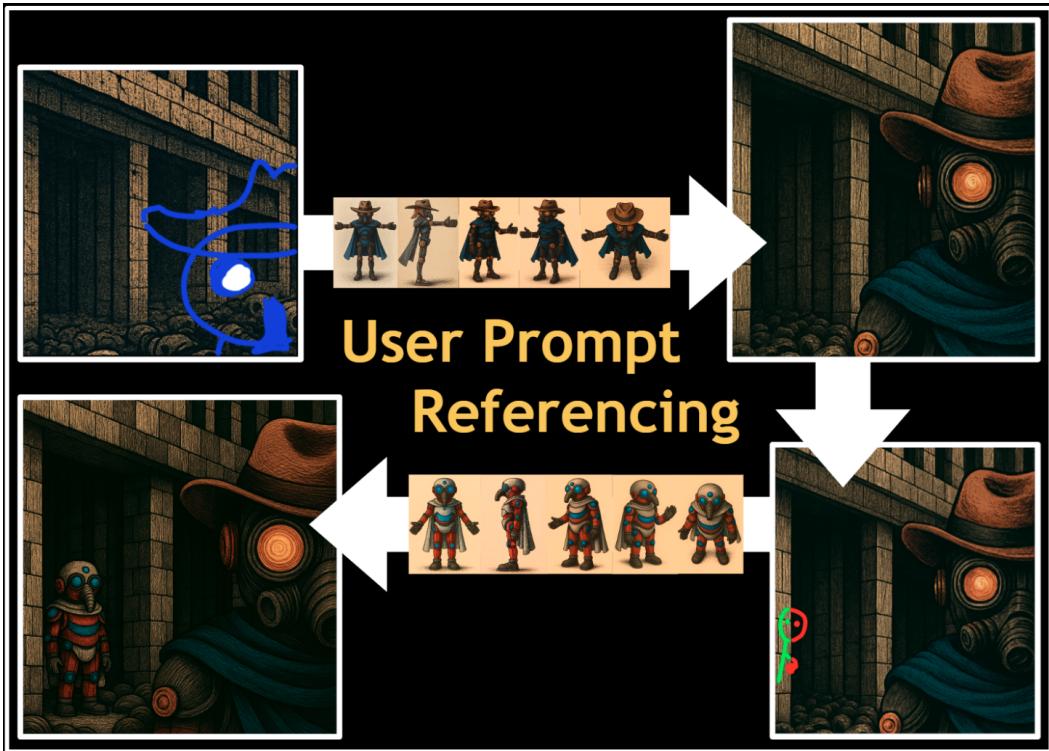


Fig. 4. The process of user prompt referencing replacing the area defined by the user to place the characters in the final environment background during Phase 4 of the experiment.

desirability: This was showcased in previous experiments that utilized all of these methods in isolation or without this experiment's full workflow. The original vision of this experiment was consistency between characters and environment, but the introduction and effectiveness of user defined prompts paved the way for the future experiment of AI assisted image editing. The ability to remove anomalies by defining what specifically needs modifying by circling affected areas, similar to how art directors suggest editing scenes in 3D work, gives the ability more flexibility to fix anomalies and illusions within AI generated artwork. Potential room for improvement of this experiment is in the dataset of the user defined prompts, this area could vastly improve when introducing successfully generated images to train DALL-E tailored to the user, and their own bias.

4 Conclusion

The overall experiment proved to be effective in producing successful image consistency along with correcting anomalies when they did occur. By incorporating user defined prompts, allowed to match the bias between user and DALL-E to reduce illusions when generating images. Prompt narrowing, prompt flooding, and prompt referencing allowed the maximization of the incorporation between DALL-E and ChatGPT, and the introduction of user defined prompts, gave way to mitigate and correct illusions. The goal of this experiment was to showcase the combination of text-to-image and image-to-image referencing that gives OpenAi's image generation model leverage over other competitors; however, when these methods are done in isolation they show the platforms

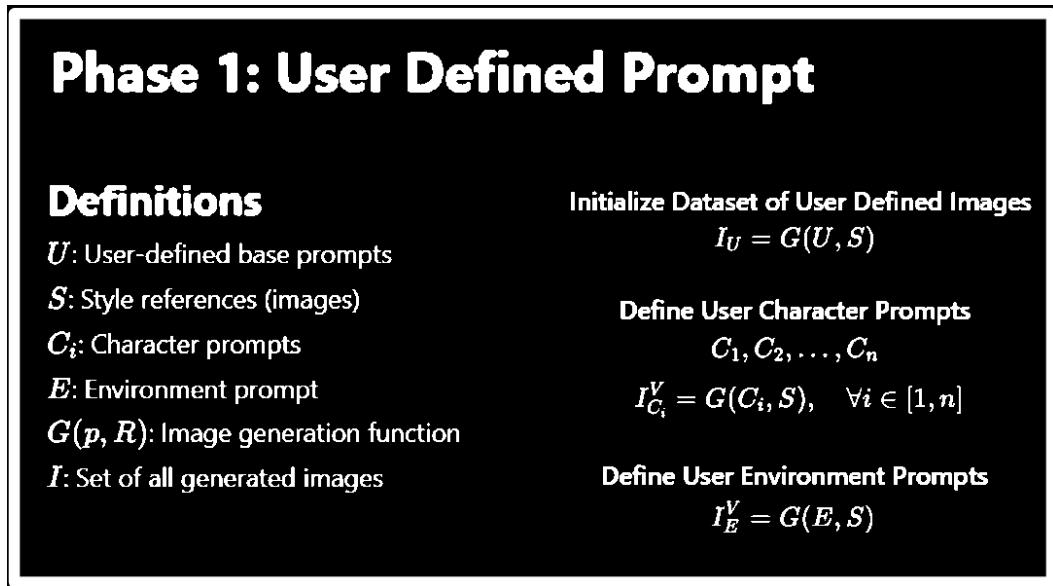


Fig. 5. The mathematics formula for user prompt defining during Phase 1 of the experiment. The user defined prompts (U) are combined with style referenced images (S) to generate images that match the users defined prompt (G). The character is then generated with several views (C) in a set (n) of varying angles. The environment is then generated with several views (E) in a set (n) of varying angles. Then the set of images is combined to save in memory as the users prompt as a set of all images (I).

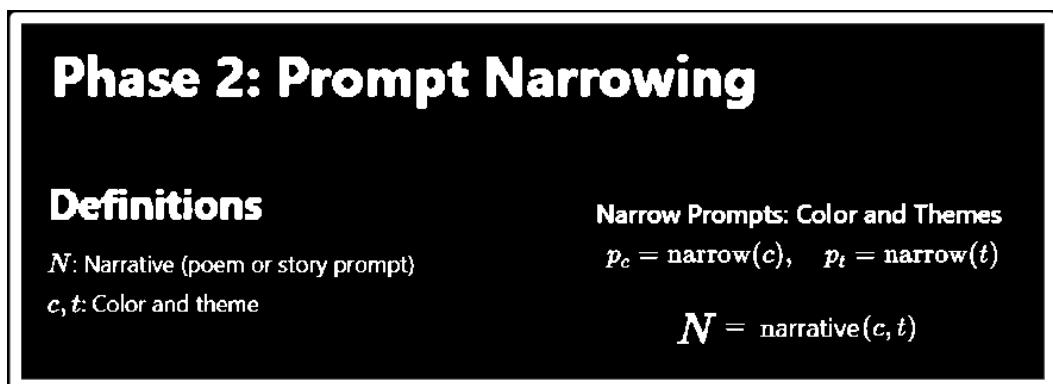


Fig. 6. The mathematics formula for prompt narrowing during Phase 2 of the experiment. Each initial prompts generated (p) are first narrowed down by color (c) and theme (t). The the initial prompts are influenced by the narrative (N) to further direct the image generation consistency.

limitations with illusions, bias, and inconsistency: When following this experiments workflow, the issues are significantly reduced.

Acknowledgments

This experiment was tested on behalf of Texas AM University.

Phase 3: Prompt Flooding

Definitions

U : User-defined base prompts

c, t : Color and theme

S : Style references (images)

$G(p, R)$: Image generation function

$E(I)$: Evaluation function

$+$: Prompt enrichment (concatenation)

Prompt Enrichment and Optimization

$$p^* = \arg \max_{p \in \{U, p_c, p_t, P\}} E(G(p, S))$$

$$p_h = p_c + p_t + p^* + P$$

Fig. 7. The mathematics formula for prompt flooding during Phase 3 of the experiment. The best prompts (p^*) are determined by the highest evaluation score ($\arg \max$) of the evaluated (E) generated images (G) when generated with the initial prompts (p). Where the initial prompts (p) are a series of iterated colors (c) and themes (t). Prompt Flooding (p_h) is then enriched by the combination of all elements of prompts influence of the final prompt (P).

Phase 4: Prompt Referencing

Definitions

U : User-defined base prompts

S : Style references (images)

C_i : Character prompts

E : Environment prompt

$G(p, R)$: Image generation function

I : Set of all generated images

$\text{ref}(I)$: Extracted reference features

Reference Successful Generated Images

$$R = \text{ref}(I_U \cup I_{C_i}^V \cup I_E^V)$$

$$I' = G(p_h, R)$$

Fig. 8. The mathematics formula for user prompt defining during Phase 4 of the experiment. The user defined series of images(I_U) is unioned with the defined character images (C) and its variants (V) and unioned along with the defined environment (E) and its variants (V). The final unioned image (p_h) is then combined with reference images (R) to tweak the final generated image result (I').

References

- [1] MM Pooja, PM Sulfath, KM Sheena, MM Pooja, and PM Sulfath. [n. d.]. DALL-E 3: Advanced AI image generation model. ([n. d.]).
- [2] Mohamad-Hani Temsah, Abdullah N Alhuzaimi, Mohammed Almansour, Fadi Aljamaan, Khalid Alhasan, Munirah A Batarfi, Ibraheem Altamimi, Amani Alharbi, Adel Abdulaziz Alsuhailani, Leena Alwakeel, et al. 2024. Art or artifact: evaluating the accuracy, appeal, and educational value of AI-generated imagery in DALL-E 3 for illustrating congenital heart diseases. *Journal of Medical Systems* 48, 1 (2024), 54.

Phase 5: Finalization

Definitions

I : Set of all generated images

$T_k(S)$: Top- k images from S based on E

F : Final comic

Assemble Final Desired Images

$$I_{\text{top}} = \mathcal{T}_{10}(I')$$

$$F = \text{assemble}(I_{\text{top}})$$

Fig. 9. The mathematics formula for user prompt defining during Phase 5 of the experiment. The final images are then iterated until the best generated image result is achieved (I_{top}) based on the top ten images (T_{10}) of the final tweaked image (I'). The comic is then assembled based on the most successful results.

Final Equation

Definitions

N : Narrative (poem or story prompt)

V : Orthographic views

F : Final comic

U : User-defined base prompts

$G(p, R)$: Image generation function

$+$: Prompt enrichment (concatenation)

c, t : Color and theme

I : Set of all generated images

S : Style references (images)

$E(I)$: Evaluation function

C_i : Character prompts

$\text{ref}(I)$: Extracted reference features

E : Environment prompt

$T_k(S)$: Top- k images from S based on E

Final Desired Images

$$F = \text{assemble} \left(\mathcal{T}_{10} \left(G \left(p_c + p_t + \arg \max_{p \in \{U, p_e, p_t, N\}} E(G(p, S)) + N, \text{ref}(I_U \cup I_{C_1}^V \cup \dots \cup I_{C_n}^V \cup I_E^V) \right) \right) \right)$$

Fig. 10. The mathematics formula for the experiment to produce every stage of the experimental workflow to generate each comic panel.

[3] Jiquan Yuan, Xinyan Cao, Changjin Li, Fanyi Yang, Jinlong Lin, and Xixin Cao. 2023. Pku-i2iqa: An image-to-image quality assessment database for ai generated images. *arXiv preprint arXiv:2311.15556* (2023).

Received 22 April 2025; revised 22 April 2025; accepted 22 April 2025

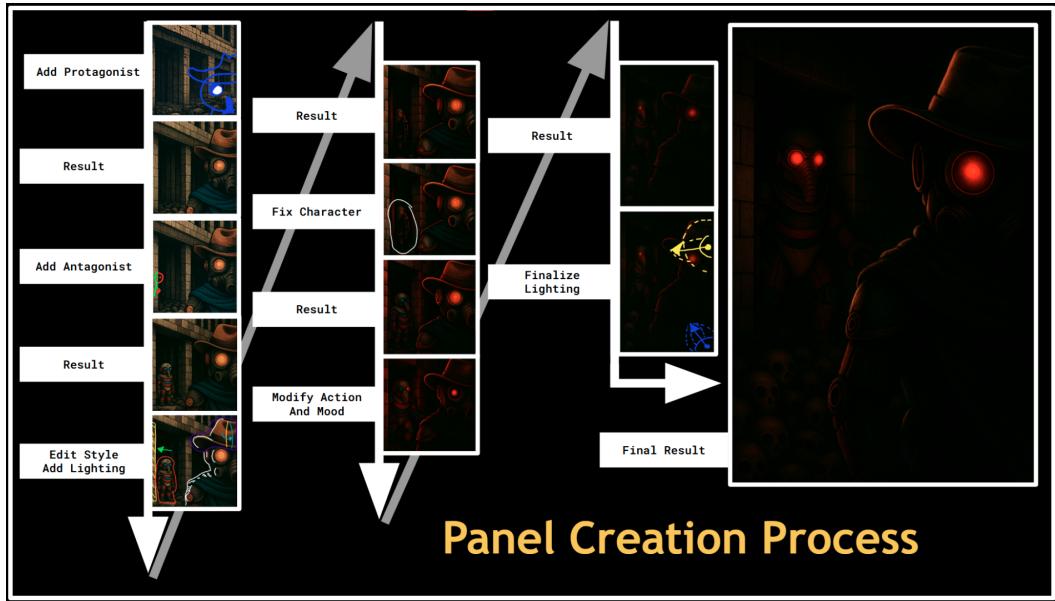


Fig. 11. The Panel Creation Process used in the final comic, and contains each iteration to produce the final panel.