Human-In-the-Loop RL for Image Generation

Aziza Ergasheva and Cathy Joseph

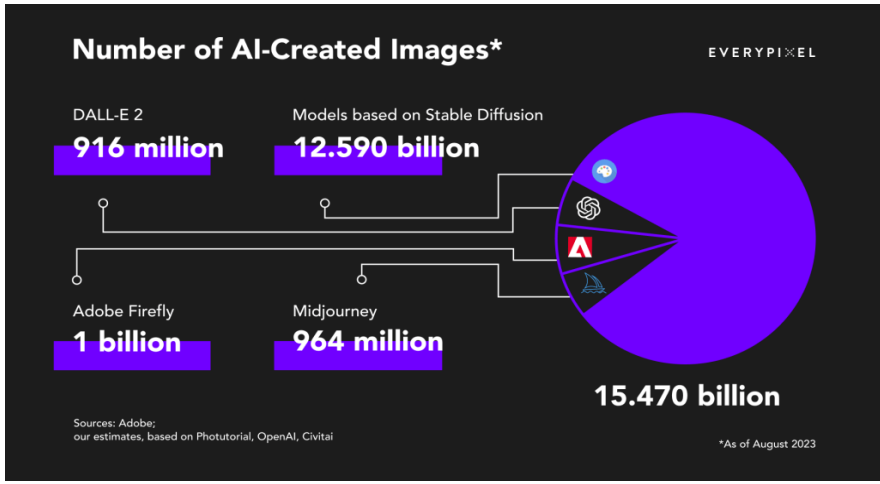
December 16th, 2025

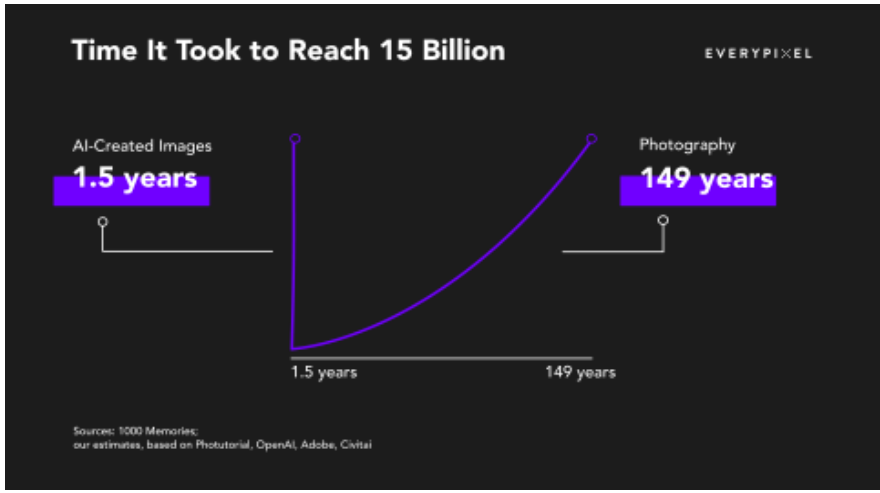
Our Motivation

- Due to its ease, efficiency, and access to vast data, Generative AI (Gen AI) has become more prevalent in creative industries.
- Yet, Gen AI has been disrupting the roles of artists
  - image generators facilitate economic loss, exacerbate inequality and bias, and promote art forgery (Jiang et. al, 2023)
  - “habitual reliance on AI appears to cultivate cognitive offloading mindset that gradually undermines the self-regulatory and generative processes essential for creativity” (Kwan &Hung, 2025, 13)
- To mitigate the harms of cognitive offloading and decrease in creative output, our goal is to find means to give artists more autonomy in this streamlined process.

Background

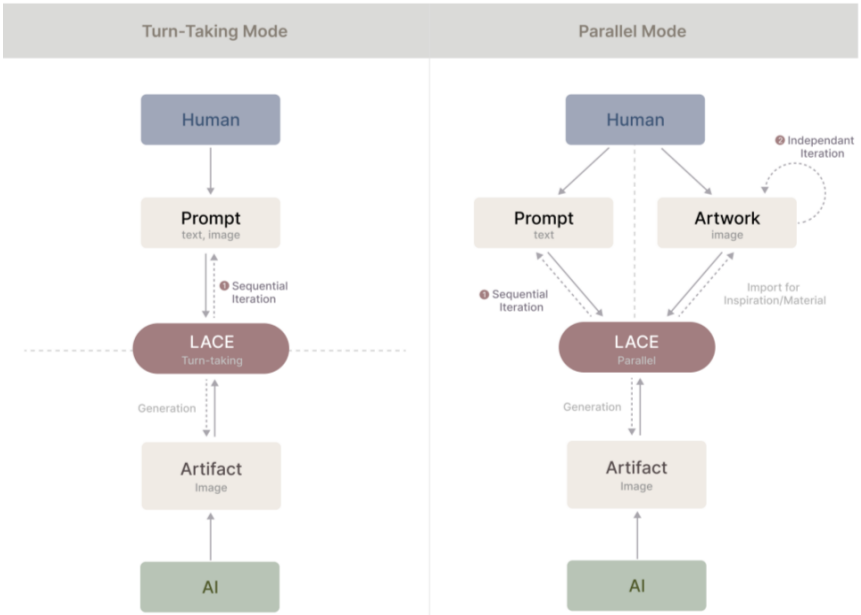
- From the early 2020s, several AI Image Generators were released and more followed:
  - 2021 OpenAI's DALL-E 1
  - 2022 Open AI's DALL-E 2
  - 2022 Stability AI's Stable Diffusion
  - 2022 Midjourney
  - 2025 Google's Nano Banana Pro
- Gen AI is a probabilistic model, meaning it learns the underlying patterns in data to generate new variations ("(PDF) Generative AI," 2025), without an inherent understanding of the task
- A recent study, approved and funded by the UCL School of Management and the University of Exeter, found a collective decrease in novelty among written short stories that utilized Gen AI as a form of assistance.





Related Work

- As a means to make Gen AI systems more interpretable to artists, concept of Experiential AI emerged (Hemment et al., 2023).
  - Artists experience the effects of changing dimensions and seeing the outputs of the model.
  - Reduces the black box feel.
- Modifying interactions with Image Generators to grant artists more autonomy (LACE) (Huang et al., 2025).
  - Co-creative approach where artists provide feedback, draw parts of an image, etc.
  - Turn Taking- AI draws an image and artist provides feedback and this process repeats.
  - Parallel Interaction: Artists draws part of an image, and the AI gives inspiration in real time.



Target Task: Learning Artistic Preferences from Human Feedback

Our claim

- Currently, modern AI systems are powerful, but inflexible and prone to replacing human judgement (Musser, 2019). This project responds to this rising problem by keeping humans in the learning process.
- It is imperative to add more human decision making to ensure it does not undermine their creative outlet and perpetuate cognitive offloading. We incorporated this by using a **preference rating scale**, as standard loss functions fail for subjective creative tasks (Lambert, et al., 2022).

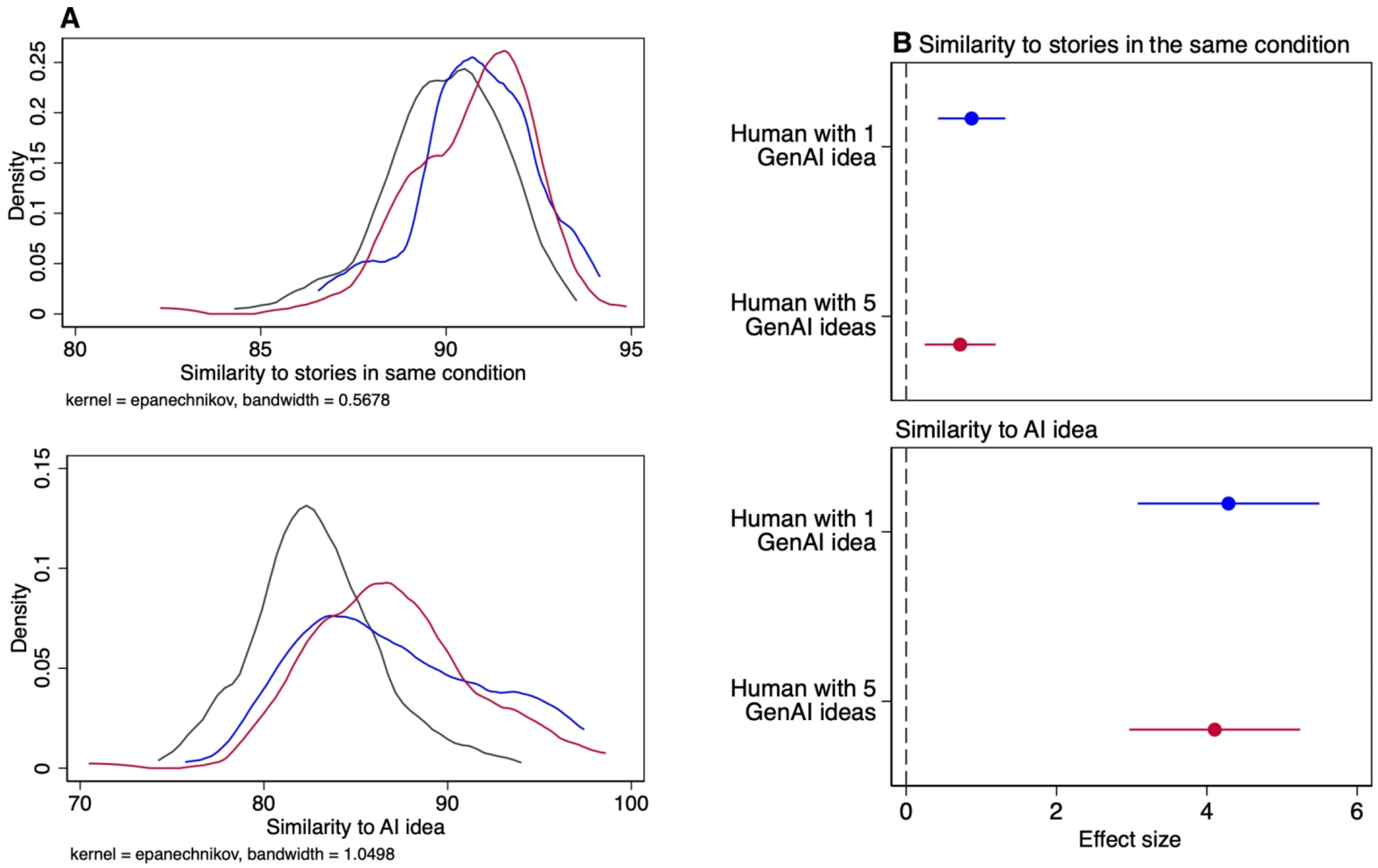
Target task

- Our goals is to add more human-decision making in art generation, starting with human-feedback to preserve artistic style and personal style

ML Objective

- Input: Current artistic state (ex: color, style, mood, etc.)
- Action: Modify the artistic attributes
- Reward: Using human feedback to learn artist styles (ratings from 1-5)
- Goal: Learn a policy that generates image aligned with the human artistic choices

Figuring Depicting Why



Proposed Solution

- Train an agent to learn artist's specific style, in which the artist/user rates each image. The agent will then refine their policy and retry another prompt.
- Once artist's style is chosen, the artist can load the style and use it when generating images.
- While basic, this is a great start for us to experiment with reinforcement learning and has plethora of applications
  - Recommending color palettes
  - Recommending inspirations similar to their styles

Implementation

[Link to demo]

- Used free Segmind Stable Diffusion 1B model (optimized for speed)

Image Generator *Unlimited Image Generations + Pretty Fast

```
#Huggin Face provided a pipeline to interact and use the image generator
# (in this case I am using Segmind Stable Diffusion 1B (SSD-1B) which is free and fast!
pipe = StableDiffusionXLPipeline.from_pretrained(
    "segmind/SSD-1B",
    torch_dtype=torch.float16
)
pipe.to("cuda")
```

- For learning artist styles:
  - Began with **Q-learning** with 4 choices ("style," "mood,""composition," and "colors") with 5 values for each

```
agent = HumanInLoopAgent(learning_rate, start_epsilon, epsilon_decay, final_epsilon, discount_factor)
env = ArtCreationEnv(pipe, {
            'color': ['warm', 'cool', 'vibrant', 'muted', 'monochrome'],
            'style': ['realistic', 'painterly', 'abstract', 'minimalist', 'surreal'],
            'mood': ['peaceful', 'energetic', 'mysterious', 'joyful', 'melancholic'],
            'composition': ['centered', 'rule-of-thirds', 'asymmetric', 'symmetrical', 'dynamic']
        })
```

  - **Deep-Q Learning** for working with larger action and state space

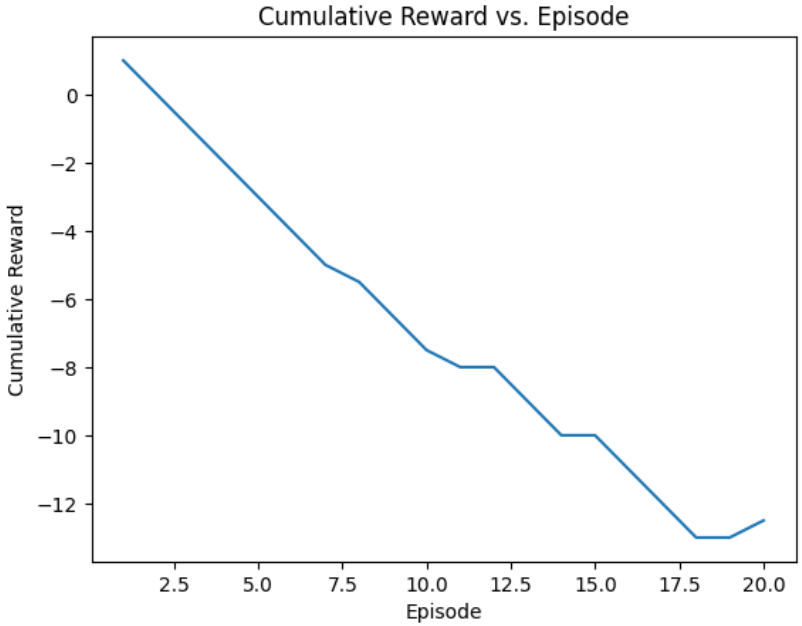
```
class DQN(nn.Module):
    def __init__(self, n_observations):
        super().__init__()
        # shared "Knowledge" Base
        self.start = nn.Sequential(
            nn.Linear(n_observations, 256),
            nn.ReLU(),
            nn.Linear(256, 128),
            nn.ReLU()
        )
        # seperate "Expert" Heads for each category
        self.color = nn.Linear(64, 5)
        self.style = nn.Linear(64, 5)
        self.mood = nn.Linear(64, 5)
        self.composition = nn.Linear(64, 5)
```

```
# in select_action():
# ...
for dim, logits in logits_dict.items():
    # 1. create a probability distribution
    dist = Categorical(logits=logits)
    # 2. sample an action
    action = dist.sample()
    # 3. tracking the Log-Probability (used for the gradient update)
    log_prob = dist.log_prob(action).squeeze()
```

  - Expanded to **Policy Gradient Optimization (PPO)**

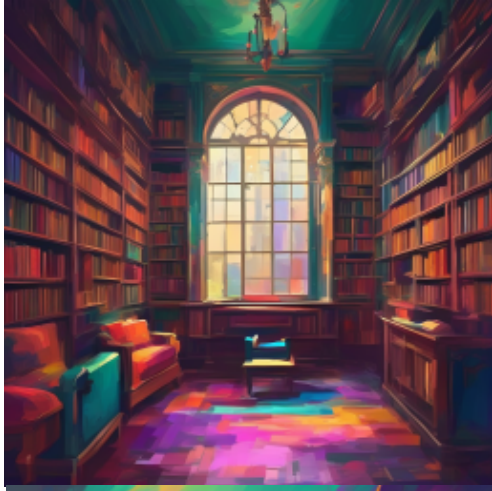
Experimental Results

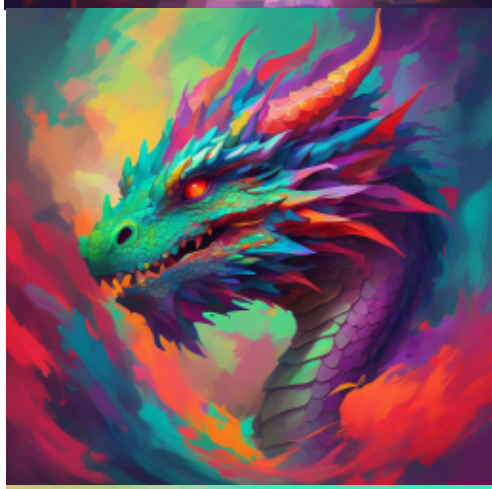
Q-learning: Rated highly for vibrant color, realistic style, peaceful mood, and centered composition



Main disadvantages:

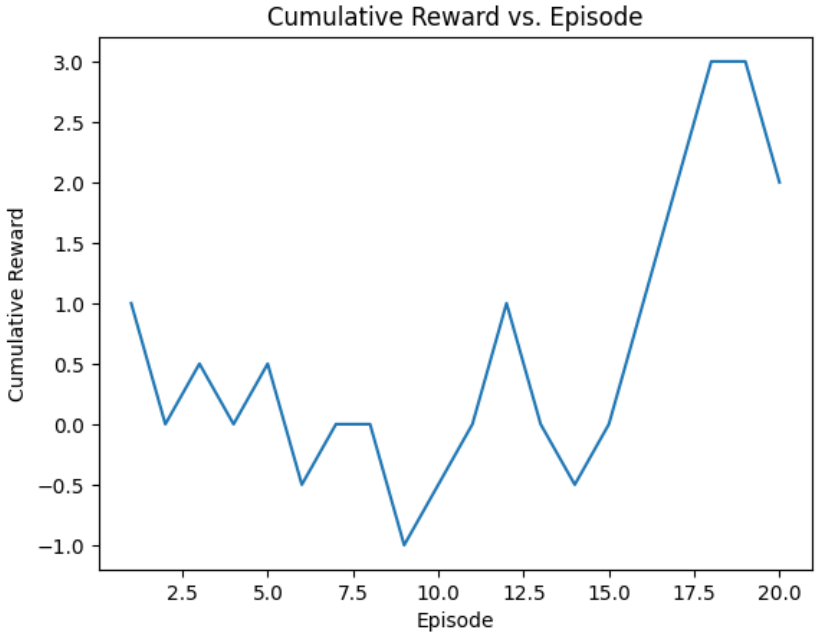
- Lacks generalization
- Curse of Dimensionality: the Q-table is huge 5 * 5 * 5 * 5 = 625 * 20 = 12,500 possible q-values





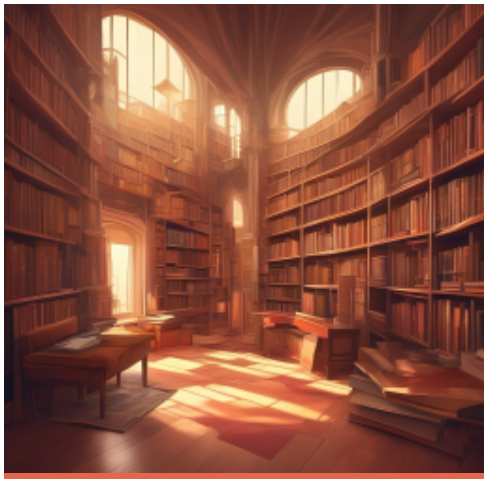
Experimental Results

Deep Q-learning: Rated highly for warm color, realistic style, peaceful mood, and centered composition



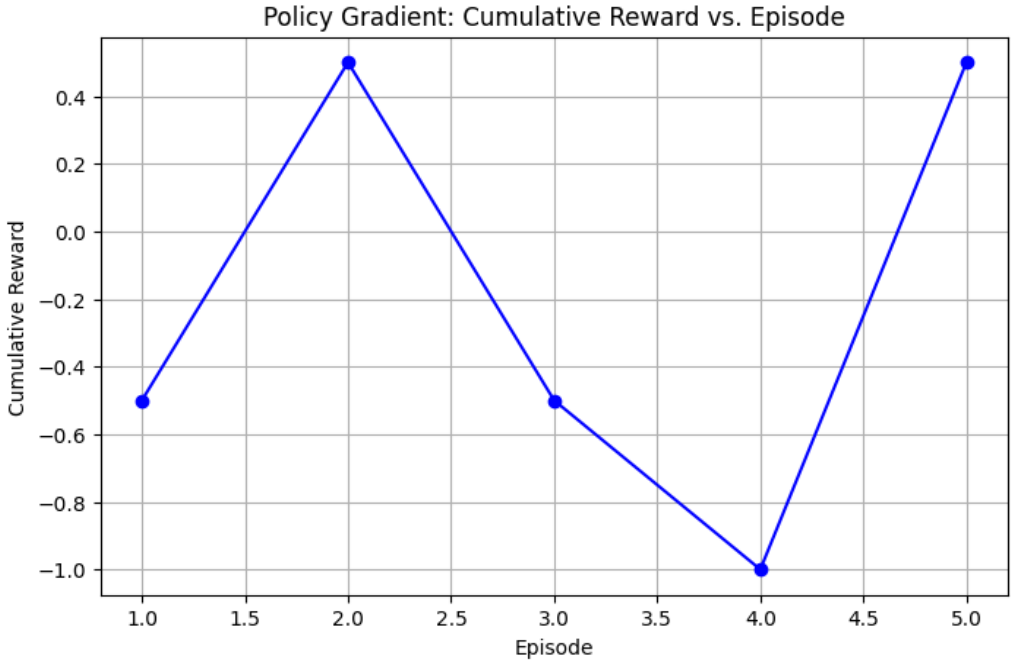
Main disadvantages:

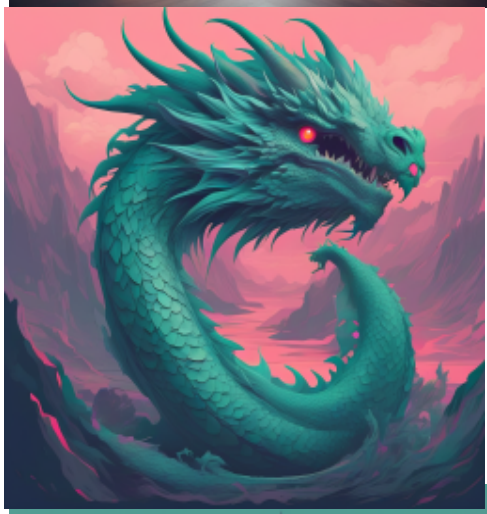
- Only ran it for 10 episodes (each with 2 iterations) --> for proof of concept



Experimental Results

Policy Gradient: Rated highly for vibrant color, surreal style, energetic mood, and dynamic composition





Main disadvantages:

- Trained with manual input and only 5 episodes

Pros:

- Although only a few episodes, the dip back to 0.5 reward shows that the model converged based on my preferences

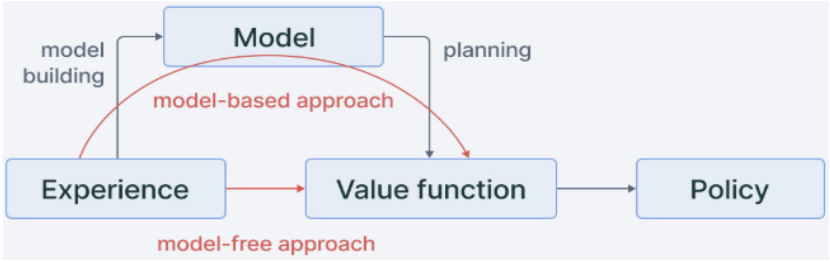
Experimental Analysis

- Q-learning doesn't perform well with large dimensions, can't generalize, high computation cost, and slow learning.
- Deep Q-learning addresses many of Q-learning's disadvantages but limited to only discrete actions.
- Policy Gradient allows for continuous actions which is much more effective (ex. 70% vibrant color paired with 30% warm) and exhibits faster convergence.

Conclusion and Future Work

- Had no interaction with the image generator; Possibly look further into modifying the image generator beyond modifying the prompts
- The image generator is older version (from 2023) targeted for speed instead of quality

Source: V7 Labs (2023)



- Future Work:
  - There is plethora more advanced reinforcement learning models we have not looked at such as
    - Model-Based RL Approaches
  - Updating model architecture to evaluate for better quality and alignment to feedback rather than performance time
  - Integration of RL-guided image generation into digital art tools to make it easier for artists to use the tool
  - Use domain-specific fine-tuning to apply the tool a specific domain like medical image illustrations, scientific images, etc.

Source: (Wikipedia, n.d.)

References

Image Credit 1 Image Credit 2 Image Credit 3. Image Credit 4

AI image statistics for 2024: How much content was created by AI. (2023, August 15). https://journal.everypixel.com/ai-image-statistics

Kwan, L. Y. Y., & Hung, Y. S. (2025). Does AI usage diminish human creativity?: How goal orientation theory moderates the negative effects between AI usage and creative output. *Social Science Computer Review*, 08944393251389125. https://doi.org/10.1177/08944393251389125

Jiang, H. H., Brown, L., Cheng, J., Khan, M., Gupta, A., Workman, D., Hanna, A., Flowers, J., & Gebru, T. (2023). AI art and its impact on artists. *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, 363–374. https://doi.org/10.1145/3600211.3604681

Generative AI: Transforming the landscape of creativity and automation. (2025). *International Journal of Computer Applications*. https://doi.org/10.5120/ijca2025924392

Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, *10*(28), eadn5290. https://doi.org/10.1126/sciadv.adn5290

Hemment, D., Vidmar, M., Panas, D., Murray-Rust, D., Belle, V., & Ruth, A. (2023). *Agency and legibility for artists through Experiential AI*. https://arxiv.org/abs/2306.02327

Musser, G. (2019, May 1). *Machine learning gets a bit more humanlike*. Scientific American. https://www.scientificamerican.com/article/machine-learning-gets-a-bit-more-humanlike/

Lambert, N., Castricato, L., Leike, J., & Amodei, D. (2022). *Illustrating reinforcement learning from human feedback (RLHF).* Hugging Face Blog. https://huggingface.co/blog/rlhf

Adobe Inc. (2023). *Adobe Photoshop logo* [Image]. Wikimedia Commons. https://en.wikipedia.org/wiki/Adobe_Photoshop#/media/File:Adobe_Photoshop_CC_2026_icon.svg

V7 Labs. (2023). *Deep reinforcement learning: A comprehensive guide*. V7 Labs Blog. https://www.v7labs.com/blog/deep-reinforcement-learning-guide