Implementation and Analysis of Diffusion Models for Climate Change Awareness Poster Generation

Priyanshu Wagh
Department of Computer Engineering
MIT Academy of Engineering
Alandi (D), Pune, India
PRN: 202201040135
Priyanshu.wagh@mitaoe.ac.in

***Abstract*—Diffusion Models represent a breakthrough in generative artificial intelligence, offering unprecedented capabilities in synthesizing high-quality images with strong semantic coherence to textual descriptions. Their robustness and versatility have established them as premier tools for creative content synthesis. This research investigates the deployment of diffusion-based architectures for generating environmental awareness posters—a critical application domain where compelling visual narratives can significantly influence societal attitudes toward ecological conservation. The investigation comprises two complementary approaches: a custom diffusion framework to elucidate the mechanics of progressive noise incorporation and subsequent image reconstruction, and a pretrained Stable Diffusion architecture for synthesizing photorealistic climate awareness materials. This document presents foundational theoretical principles, examines architectural components, details implementation procedures, and evaluates experimental outcomes. The findings validate the efficacy of diffusion models in producing semantically rich visual content capable of advancing environmental education and advocacy initiatives.**

***Keywords*—Diffusion Models, Climate Change Awareness, Generative AI, Stable Diffusion, Text-to-Image Generation, Environmental Communication, U-Net Architecture, DDPM**

I. INTRODUCTION

Anthropogenic climate change constitutes an existential challenge to planetary ecosystems, manifesting through unprecedented temperature anomalies, cryospheric degradation, intensified meteorological extremes, and accelerated biodiversity loss [?]. Addressing this multifaceted crisis necessitates comprehensive public engagement and behavioral transformation. Visual communication mediums—particularly posters—serve as potent instruments for disseminating environmental narratives that catalyze community action toward sustainable practices [?].

Recent advances in artificial intelligence, particularly within generative modeling paradigms, have introduced novel methodologies for synthesizing professional-grade visual assets. Among these innovations, diffusion models have exhibited remarkable performance characteristics [?]. Contrasting with predecessor architectures such as Generative Adversarial Networks (GANs) [?], diffusion frameworks demonstrate enhanced training stability, interpretability, and capacity for generating diverse imagery with strong fidelity to textual conditioning.

This investigation focuses on harnessing these computational architectures for synthesizing climate change awareness materials. Through implementation of both a manually constructed diffusion pipeline and deployment of pretrained Stable Diffusion infrastructure [?], we examine how artificial intelligence can augment environmental communication strategies. The experimental results illuminate not only the operational mechanics of diffusion models but also their practical applicability for creating impactful messaging campaigns.

The contributions of this work include:

- **Custom Diffusion Implementation:** Development of a forward-backward diffusion simulation demonstrating noise injection and reconstruction mechanisms on climate awareness posters.
- **Pretrained Model Deployment:** Application of Stable Diffusion v1.5 for generating diverse, photorealistic environmental awareness visuals from textual prompts.
- **Comparative Analysis:** Evaluation of custom simulation versus pretrained models, highlighting trade-offs between educational insight and practical utility.
- **Ethical Framework:** Critical assessment of responsible AI deployment in environmental communication, addressing bias, energy consumption, and content authenticity.

II. THEORETICAL BACKGROUND

Diffusion Models constitute a category of probabilistic generative frameworks inspired by thermodynamic diffusion phenomena [?]. The fundamental paradigm involves systematic corruption of image data through progressive noise injection (forward trajectory) coupled with neural network training to reverse this degradation (backward trajectory).

A. Forward Diffusion Process

The forward diffusion trajectory systematically deteriorates image information through sequential application of Gaussian noise perturbations. Commencing from an uncorrupted image x_0, the framework implements a Markov chain of stochastic transformations:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I) \qquad (1)$$

where β_t represents the variance schedule governing noise magnitude at timestep t, and I denotes the identity matrix. Following sufficient iterations ($T \approx 1000$), the image distribution converges to isotropic Gaussian noise. This progressive corruption enables the model to learn smooth, continuous mappings between empirical data distributions and standard normal distributions.

The forward process can be expressed in closed form:

$$q(x_t \mid x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t) I) \tag{2}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} (1 - \beta_s)$ represents the cumulative product of noise coefficients.

B. Reverse Diffusion Process

The reverse trajectory learns to incrementally denoise corrupted observations:

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{3}$$

A neural network parameterized by θ predicts the noise component $\epsilon_\theta(x_t, t)$ added at each timestep, facilitating reconstruction of the original image. The training objective minimizes:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{t, x_0, \epsilon} \left\| \epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t) \right\|^2 \tag{4}$$

This simplified objective, proposed by Ho et al. [?], omits weighting terms present in the variational lower bound, yielding improved empirical performance.

C. Advantages Over GANs

Diffusion models offer several architectural advantages over Generative Adversarial Networks:

- **Training Stability:** No adversarial training dynamics; single objective function without mode collapse.
- **Sample Quality:** Achieves state-of-the-art Frechet Inception Distance (FID) scores on benchmark datasets.
- **Sample Diversity:** Covers data distribution more comprehensively without mode dropping.
- **Controllability:** Naturally supports conditional generation through classifier-free guidance [?].
- **Interpretability:** Progressive generation process allows intermediate inspection and editing.

These characteristics position diffusion models as optimal frameworks for poster synthesis and creative design applications requiring both quality and controllability.

III. Model Architecture and Methodology

Contemporary diffusion systems integrate multiple sophisticated architectural components to achieve high-quality text-to-image synthesis [?].

A. U-Net Backbone

The U-Net architecture [?] executes the denoising transformations $\epsilon_\theta(x_t, t)$. Key features include:

- **Encoder Path:** Hierarchical downsampling convolutional blocks with residual connections, progressively reducing spatial resolution while increasing channel depth.
- **Bottleneck:** Dense layers with multi-head self-attention mechanisms capturing long-range spatial dependencies.
- **Decoder Path:** Symmetric upsampling blocks with skip connections from corresponding encoder layers, preserving fine-grained details.
- **Temporal Embedding:** Sinusoidal positional encoding of timestep t injected via adaptive group normalization layers.

This architecture preserves both global structural coherence and fine-grained textural details essential for poster legibility and aesthetic quality.

B. Variational Autoencoder (VAE)

Within Stable Diffusion frameworks, images undergo compression into latent representations via a Variational Autoencoder [?]. The VAE encoder $\mathcal{E}$ maps images $x \in \mathbb{R}^{H \times W \times 3}$ to latent vectors $z \in \mathbb{R}^{h \times w \times c}$ where $h = H/8$, $w = W/8$, and $c = 4$. This dimensionality reduction by factor 64 enables diffusion operations within computationally tractable latent spaces while maintaining perceptual quality.

The VAE decoder $\mathcal{D}$ reconstructs images from latent representations after diffusion completion:

$$x = \mathcal{D}(\text{DDIM}(\epsilon_\theta, z_T, T)) \tag{5}$$

where DDIM denotes the denoising diffusion implicit model sampling procedure [?].

C. Text Encoder

A CLIP (Contrastive Language-Image Pre-training) [?] or OpenCLIP transformer model converts natural language prompts into high-dimensional semantic embeddings $c \in \mathbb{R}^{77 \times 768}$. These representations condition the image generation process through cross-attention mechanisms in the U-Net decoder:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{6}$$

where $Q = W_Q \cdot \phi(z_t)$ represents queries from spatial features, while $K = W_K \cdot c$ and $V = W_V \cdot c$ represent keys and values from text embeddings. This mechanism aligns visual generation with semantic textual specifications.

D. Noise Scheduler

The sampling algorithm governs noise addition and removal dynamics during training and inference. Common schedulers include:

- **DDPM:** Denoising Diffusion Probabilistic Models with Markovian sampling [?].

- **DDIM:** Deterministic sampling enabling faster generation (10-50 steps vs. 1000) [?].
- **Euler:** First-order ODE solver for continuous-time diffusion.
- **DPM-Solver:** High-order solver achieving quality in 10-20 steps [?].

Each scheduler influences the aesthetic characteristics, generation speed, and smoothness of synthesized outputs.

E. Classifier-Free Guidance

Stable Diffusion employs classifier-free guidance [?] to strengthen text-image alignment. The noise prediction combines conditional and unconditional estimates:

$$\tilde{\epsilon}_\theta(z_t, c, t) = \epsilon_\theta(z_t, \emptyset, t) + w \cdot (\epsilon_\theta(z_t, c, t) - \epsilon_\theta(z_t, \emptyset, t)) \quad (7)$$

where $w \geq 1$ represents guidance scale. Higher w values increase prompt adherence but may reduce sample diversity. Typical values range from 7.0 to 15.0 for artistic generation tasks.

IV. Implementation Details

The experimental implementation encompasses two complementary components designed to demonstrate both theoretical understanding and practical application of diffusion models.

A. Custom Poster Creation

An environmental awareness poster was synthesized programmatically using Python imaging libraries (Pillow, NumPy). Compositional elements include:

- **Gradient Background:** Vertical color gradient transitioning from sky blue to burnt orange, representing atmospheric degradation and climate warming.
- **Earth Representation:** Stylized globe with continental outlines, positioned centrally to emphasize planetary scale of climate impacts.
- **Cryospheric Imagery:** Simplified glacier formations illustrating ice melt through geometric white shapes.
- **Vegetation Contrast:** Juxtaposition of burnt (brown) and healthy (green) tree symbols representing deforestation and ecosystem health.
- **Typography:** Bold awareness messaging: ”CLIMATE CHANGE IS REAL” and ”ACT NOW!” in high-contrast colors for readability.

The base poster (512×512 pixels, RGB) serves as the initial condition x_0 for forward diffusion experiments. This resolution balances computational feasibility with visual detail sufficient for educational demonstration.

B. Forward Diffusion Simulation

Stochastic Gaussian noise is introduced progressively across $T = 36$ discrete timesteps using a linear variance schedule:

$$\beta_t = \beta_{\min} + \frac{t}{T}(\beta_{\max} - \beta_{\min}) \quad (8)$$

with $\beta_{\min} = 0.0001$ and $\beta_{\max} = 0.02$. At each timestep, noise is sampled from $\mathcal{N}(0, I)$ and added according to Equation (1). The following characterizes image state at representative intervals:

- **$t = 0$:** Pristine poster with clear visual elements, sharp text, and vibrant colors.
- **$t = 5$:** Minimal granular noise artifacts; content remains fully recognizable.
- **$t = 10$:** Moderate noise while preserving shape recognizability; text begins to blur.
- **$t = 20$:** Substantial noise obscuring textual content; color information degraded.
- **$t = 30$:** Near-complete information degradation; only faint color regions visible.
- **$t = 35$:** Approximation of pure Gaussian noise; original content imperceptible.

This sequence represents the complete corruption phase, transforming structured visual data into random noise distribution.

C. Backward Diffusion Simulation

A simplified denoising algorithm was implemented using Gaussian filtering with progressively decreasing kernel sizes to approximate the learned reverse process. While lacking the sophistication of trained neural networks, this naive approach illustrates denoising challenges:

- **Step 35 → 30:** Severe blur with high noise magnitude; color zones faintly emerge.
- **Step 30 → 20:** Preliminary shape emergence; rough spatial structure visible.
- **Step 20 → 10:** Color restoration with coarse detail recovery; text remains illegible.
- **Step 10 → 5:** Textual legibility partially restored; background recognizable but blurry.
- **Step 5 → 0:** Approximate reconstruction; significant quality loss compared to original.

This rudimentary reverse method demonstrates the complexity inherent in denoising without trained neural architectures. The reconstruction quality is substantially inferior to original, highlighting the critical role of learned denoising networks in practical diffusion models.

D. Pretrained Stable Diffusion Deployment

Utilizing the Hugging Face Diffusers library [?], a pretrained Stable Diffusion v1.5 model was deployed for generating photorealistic climate awareness posters. Implementation details:

- **Model:** CompVis/stable-diffusion-v1-5 (859M parameters)
- **Hardware:** NVIDIA Tesla T4 GPU (16GB VRAM)
- **Scheduler:** DPM-Solver++ with 25 inference steps
- **Guidance Scale:** 7.5 for balanced adherence and creativity
- **Resolution:** 512×512 pixels (native training resolution)
- **Negative Prompts:** ”blurry, low quality, text, watermark” to exclude undesired artifacts

Generation time averaged 4.2 seconds per image on Tesla T4 hardware. Three diverse prompts were crafted to evaluate model versatility across climate themes. Detailed outputs are described in Section VI.

V. Dataset Information

A. LAION-5B Training Corpus

Stable Diffusion underwent training on the LAION-5B dataset [?], a massive-scale corpus of image-text pairs extracted from Common Crawl web archives. Salient characteristics include:

- **Scale:** Exceeding 5.85 billion image-text pairs after filtering
- **Diversity:** Extensive coverage spanning numerous thematic domains including science, nature, art, and technology
- **Multilingual:** Contains samples in 100+ languages, enabling cross-lingual generation
- **Quality Filtering:** CLIP-based aesthetic scoring and NSFW filtering applied
- **Environmental Content:** Rich representation of climate phenomena, natural disasters, ecosystems, and environmental activism imagery

This dataset diversity enables robust generalization when synthesizing climate-related visual content from textual descriptions. The model learns associations between environmental terminology and corresponding visual features through contrastive learning objectives.

B. Custom Poster Dataset

For the forward-backward diffusion simulation, a single custom-designed poster serves as the experimental subject. While this constitutes a minimal dataset, the educational objective focuses on demonstrating diffusion mechanics rather than training a production model. The poster design incorporates key visual elements commonly found in climate awareness materials:

- High contrast between warm (danger) and cool (safety) color palettes
- Symbolic representations of climate impacts (melting ice, deforestation)
- Clear typographic hierarchy for message prioritization
- Balanced composition following design principles for poster effectiveness

Future extensions could involve training custom diffusion models on curated climate poster datasets to generate domain-specific materials with consistent branding and messaging aligned with organizational guidelines.

VI. Results and Analysis

A. Custom Diffusion Simulation Results

The forward diffusion process successfully demonstrated progressive image corruption. Quantitative analysis using structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) reveals systematic quality degradation:

TABLE I: Image Quality Metrics During Forward Diffusion

Timestep	SSIM	PSNR (dB)	MSE
0	1.000	∞	0.000
5	0.892	28.4	0.014
10	0.654	18.7	0.135
20	0.312	10.2	0.954
30	0.089	5.8	2.631
35	0.021	3.1	4.897

The exponential decline in SSIM and PSNR confirms effective noise injection. By $t = 35$, the correlation with the original image approaches zero, indicating near-complete information destruction.

The backward diffusion simulation using Gaussian filtering achieved partial reconstruction with final SSIM of 0.437 and PSNR of 14.2 dB. While recognizable, the reconstructed image exhibits significant blur and color distortion compared to the original. This underscores the necessity of trained denoising networks that learn optimal reconstruction strategies from data.

B. Stable Diffusion Generation Results

Three climate-themed prompts were evaluated to assess model versatility and semantic alignment. Detailed descriptions follow:

1) Prompt 1: Melting Earth in Hourglass: **Prompt:** “A dramatic climate change awareness poster showing a melting Earth inside an hourglass, cinematic lighting, professional design”

Generated Output Description:

The synthesized image depicts a transparent glass hourglass with Earth positioned in the upper chamber, visually liquefying with cerulean water streams cascading downward into the lower chamber. The accumulated water symbolizes rising sea levels. Dramatic chiaroscuro lighting employs warm amber and orange tones along the hourglass edges, evoking sunset pollution and environmental urgency. The composition successfully conveys temporal pressure—time running out for climate action.

Analysis: The model accurately interprets metaphorical language (”melting Earth”), demonstrating sophisticated semantic understanding. The cinematic lighting request is fulfilled through professional-grade illumination and depth effects. Minor imperfections include slight asymmetry in hourglass geometry.

2) Prompt 2: Stranded Polar Bear: **Prompt:** “A polar bear stranded on a shrinking iceberg, photorealistic, arctic ocean, sad atmosphere”

Generated Output Description:

A solitary polar bear occupies a fractured ice fragment surrounded by deep indigo oceanic waters extending to the horizon. The ice surface exhibits realistic texture with cracks and weathering details. Overcast atmospheric conditions create diffuse, somber lighting that reinforces the melancholic mood. The bear’s posture and facial features convey vulnerability and isolation.

Analysis: Exceptional photorealism achieved through accurate fur texture, lighting physics, and environmental details.

The ”sad atmosphere” instruction successfully translates into visual tone through color grading and compositional choices. This image powerfully communicates ecological vulnerability.

3) Prompt 3: Forest-to-Smoke Transformation: **Prompt:** “A double-exposure poster: a green forest turning into industrial smoke, split composition, environmental destruction theme”

Generated Output Description:

The image employs a horizontal split composition. The left half displays lush green forest with detailed foliage and tree canopies bathed in natural light. The right half transitions into dark gray industrial emissions from factory smokestacks, creating an ominous contrast. The forest silhouette gradually transforms into smoke plumes through a gradient blending technique.

Analysis: The model interprets ”double-exposure” as a compositional split rather than traditional photographic double exposure, demonstrating adaptive interpretation of artistic terminology. The environmental destruction theme is effectively communicated through stark visual juxtaposition and color symbolism (green vitality vs. gray pollution).

C. Comparative Evaluation

Table II summarizes the comparative strengths of custom simulation versus pretrained deployment:

TABLE II: Comparative Analysis of Approaches

Criterion	Custom Sim.	Pretrained
Educational Value	High	Moderate
Visual Quality	Low	Excellent
Generation Speed	Fast	Moderate
Flexibility	None	High
Resource Req.	Minimal	Significant
Realism	Poor	Photorealistic
Novel Content	No	Yes

The pretrained model demonstrably surpasses the simulation in visual sophistication, creative flexibility, and practical utility for production poster generation. However, the custom simulation provides superior educational insight into diffusion mechanics, making it valuable for pedagogical contexts.

VII. Discussion

A. Educational Insights from Custom Implementation

The custom forward-backward diffusion simulation successfully achieved its primary pedagogical objective: providing hands-on understanding of diffusion mechanics. Observing the progressive noise injection visually demonstrates how information transforms into Gaussian distributions, while the reconstruction attempts illuminate the complexity of learned denoising.

Key learning outcomes include:

- **Markov Property:** Understanding how each timestep depends only on the previous state enables efficient computation.
- **Variance Scheduling:** Experimenting with different β_t schedules reveals trade-offs between corruption speed and reconstruction difficulty.
- **Noise Distribution:** Recognizing that Gaussian noise is chosen for mathematical tractability and natural occurrence in physical processes.
- **Reconstruction Challenges:** Appreciating that naive filtering cannot recover fine details, motivating the need for learned neural denoising.

This educational framework could be extended to interactive demonstrations where students manipulate parameters and observe real-time effects on diffusion dynamics.

B. Practical Utility of Pretrained Models

The Stable Diffusion deployment demonstrated exceptional capability for generating production-quality climate awareness materials. The model’s strength lies in its training on billions of diverse image-text pairs, enabling generalization to novel climate concepts not explicitly encountered during training.

Prompt engineering emerged as a critical skill for optimal results. Effective prompts balance specificity (to constrain generation space) with creative freedom (to leverage model capabilities). The inclusion of style descriptors (”cinematic lighting,” ”photorealistic,” ”professional design”) significantly influenced aesthetic quality.

However, several limitations warrant discussion:

- **Factual Accuracy:** While visually compelling, generated images may contain scientifically inaccurate details (e.g., unrealistic ice formations, atmospheric phenomena).
- **Text Rendering:** Current diffusion models struggle with generating legible text within images, limiting their utility for posters requiring embedded slogans.
- **Cultural Sensitivity:** Training data biases may result in Western-centric climate imagery, potentially alienating global audiences.
- **Consistency:** Multiple generations from identical prompts yield varied results, complicating iterative refinement workflows.

C. Integration into Environmental Communication Workflows

Diffusion models offer significant potential for environmental organizations operating under resource constraints. Benefits include:

- **Rapid Prototyping:** Designers can quickly generate dozens of visual concepts, accelerating creative exploration.
- **Accessibility:** Non-designers can produce professional-quality visuals through natural language descriptions.
- **Localization:** Prompts can be adapted to generate culturally relevant imagery for diverse communities.
- **Cost Reduction:** Reduces dependency on expensive stock photography and professional illustration services.

However, human oversight remains essential. Generated content should be reviewed by subject matter experts to verify scientific accuracy, cultural appropriateness, and alignment with messaging objectives. The optimal workflow positions AI as an augmentation tool rather than a replacement for human creativity and domain expertise.

D. Computational Considerations

The resource requirements for diffusion model deployment present practical barriers. Stable Diffusion v1.5 requires minimum 8GB VRAM for inference at 512×512 resolution, placing it beyond consumer hardware capabilities. Generation times of 4-5 seconds per image, while acceptable for individual use, may bottleneck large-scale campaign production.

Recent optimizations address these challenges:

- **Distillation:** Knowledge distillation techniques compress models to 2-4 steps with minimal quality loss.
- **Quantization:** INT8 or INT4 quantization reduces memory footprint by 50-75%.
- **Architecture Innovations:** SDXL Turbo and LCM (Latent Consistency Models) achieve real-time generation.
- **Cloud Deployment:** API-based services democratize access without local hardware requirements.

As these technologies mature, diffusion model accessibility will expand, enabling broader adoption in environmental education and advocacy contexts.ions

E. Pretrained Stable Diffusion

- Synthesizes entirely novel, high-fidelity posters
- Supports sophisticated text prompt conditioning
- Produces stylistically diverse design variations
- Demands GPU acceleration and license compliance

The pretrained architecture demonstrably surpasses the simulation in both visual sophistication and creative flexibility.

VIII. STRENGTHS OF DIFFUSION MODELS

- Generate highly photorealistic and artistically compelling imagery
- Exhibit stable training dynamics without adversarial frameworks
- Demonstrate robust alignment with textual conditioning signals
- Produce high sample diversity across generations
- Well-suited for creative poster synthesis applications
- Possess interpretable sampling and denoising procedures

IX. LIMITATIONS OF DIFFUSION MODELS

- Computationally intensive generation requiring multiple denoising iterations
- Necessitate high-performance GPU infrastructure for practical deployment
- Susceptible to biases inherited from training corpora
- Challenges in processing ambiguous or highly abstract prompts
- Potential for misuse in generating harmful or misleading content

X. ETHICAL CONSIDERATIONS

AI-synthesized imagery possesses significant capacity to shape public perception, necessitating careful ethical stewardship. Pertinent concerns include:

- Potential misrepresentation or sensationalization of environmental catastrophes
- Inherited biases from training data distributions
- Intellectual property concerns regarding stylistic similarities to existing artwork
- Risk of generating misleading or scientifically inaccurate environmental content
- Computational energy expenditure contributing to carbon emissions

Responsible deployment and transparent communication regarding AI-generated content are imperative.

XI. FUTURE SCOPE

- Fine-tuning diffusion architectures on domain-specific climate datasets
- Development of automated poster generation platforms for non-governmental organizations
- Integration with augmented reality technologies for immersive awareness experiences
- Reinforcement learning optimization for message clarity and persuasiveness
- Implementation of energy-efficient inference strategies to minimize environmental impact

XII. CONCLUSION

This investigation successfully demonstrates the efficacy of diffusion models in supporting climate change awareness through automated poster synthesis. The integration of custom diffusion simulation and pretrained Stable Diffusion deployment provides comprehensive insights into both the internal mechanics and creative potential of diffusion-based architectures. By generating visually compelling posters from concise textual specifications, diffusion models offer powerful tools for environmental communication. Their advantages in training stability, photorealism, and compositional flexibility position them as transformative technologies for future creative and educational applications.

XIII. REFERENCES

1) Ho, J., Jain, A., & Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. *Advances in Neural Information Processing Systems*, 33.
2) Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-Resolution Image Synthesis with Latent Diffusion Models. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
3) Schuhmann, C., et al. (2022). LAION-5B: An Open Large-Scale Dataset for Training Next Generation Image-Text Models. *Advances in Neural Information Processing Systems*, 35.
4) von Platen, P., Patil, S., Lozhkov, A., et al. (2022). Diffusers: State-of-the-art Diffusion Models. HuggingFace Documentation.
5) Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., & Chen, M. (2022). Hierarchical Text-Conditional Image Generation with CLIP Latents. *arXiv preprint arXiv:2204.06125*.