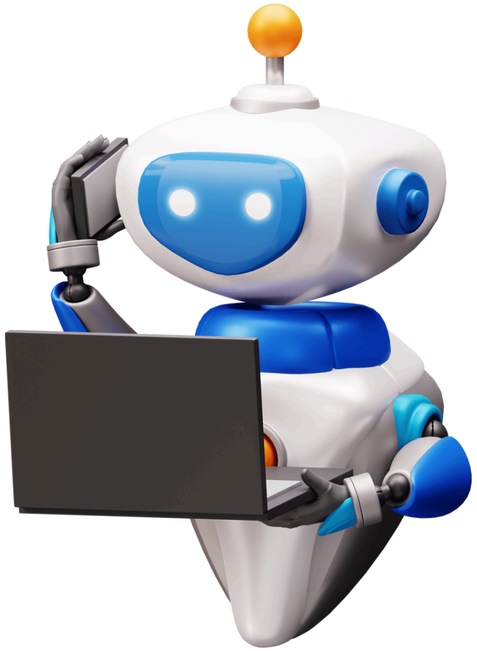
NEURAPIX- AI IMAGE GENERATOR

GEN AI + NLP



Mentor: Grishma Sharma

Date: 15|04|25

Team Members:

Smit Patil-16010122139

Toyash Patil-16010122140

Khushi Poojary-16010122147

INTRODUCTION

AI-based image generation refers to the use of artificial intelligence models to create realistic or artistic images from textual descriptions, sketches, or existing images. This project utilizes Stability AI 2-1 and 3.5 Large Model, which are advanced generative models capable of producing high-quality visuals with fine details and enhanced realism.

1

Creativity Enhancement

AI-powered tools assist artists, designers, and content creators in producing unique visuals with minimal effort.

2

Automation & Efficiency

Reduces the time required to create digital artwork, making design processes more efficient

3

Customization & Personalization

Enables users to generate custom images based on specific inputs, enhancing engagement in gaming, advertising, and social media.

4

Advancements in AI Research

Paves the way for improvements in computer vision, deep learning, and AI-driven creativity

MOTIVATION

Why Do We Need **NeuraPix**-AI image generator?

1

Growing Demand

Industries like advertising, gaming, and entertainment increasingly rely on AI for high-quality image generation

2

Enhancing Image

Overcoming limitations of previous models by improving realism, diversity, and control over generated outputs

3

Bridging the Gap

Empowering artists, designers, and developers with AI-assisted tools for efficient and innovative content creation

4

Efficient Rendering

Optimization in Stable Diffusion 3.5 Large allows for improved performance while maintaining accessibility

SCOPE

Software Development

Generates UI elements and mockups from text to speed up development workflows.

Gaming & Animation

Creates game backgrounds and concept art from prompts to boost creative efficiency

Automation in Content Generation

Automates the creation of visuals for marketing and branding using AI.

OBJECTIVES

AI-Powered Image Generation	Leverage Generative AI and NLP to convert text prompts into high-quality images, enabling effortless visual creation for various domains.
Personalized Visual Design	Allow users to generate custom visuals tailored to their style, theme, or product needs, enhancing marketing and branding impact.
Creative Assistance for Artists & Designers	Provide inspiration and starting points for artists, animators, and designers, reducing creative blocks and speeding up the design process..
Automation of Content Creation	Automate the creation of promotional materials, product mockups, and social media content to improve efficiency and consistency.

OBJECTIVES

Gaming & Animation Asset Generation	Facilitate the creation of characters, environments, and concept art from text inputs, streamlining pre-production pipelines
User-Friendly Interface	Offer an intuitive and responsive interface that makes AI image generation accessible to both technical and non-technical users.
Accessibility for Small Creators	Empower freelancers, small businesses, and solo creators with affordable access to professional-grade visual assets.
Continuous Learning & Improvement	Improve image generation quality by continuously updating the AI model with diverse datasets and user feedback.

METHODOLOGY

Data Collection

- Data Sourcing: Collected high-quality text-to-image datasets like LAION-5B and OpenImages, which include diverse image-text pairs across domains (products, art, gaming, etc.)
- Prompt Templates: Designed structured and creative prompt formats to cover various use cases such as promotional banners, product mockups, game assets, etc.
- API Utilization: Leveraged Stability AI's Stable Diffusion API, which is pretrained on diverse datasets and supports conditional image generation from natural language prompts.

WhatsApp Bot Integration

- Bot Framework: Used platforms like Twilio, 360Dialog, or WhatsApp Business API to build and deploy the bot interface.
- Functionality Flow:
  User Sends Prompt: User sends a message (e.g., "Generate a cyberpunk cityscape at night").
  Server Receives Prompt: A backend (Node.js/Express) receives the message and forwards the prompt to the Stability AI API.
  Image is Generated: The API returns a URL or base64 image.
  Bot Replies with Image: The image is sent back to the user on WhatsApp with optional watermark or caption.
- Backend Stack
  Node.js / Express for handling requests.
  Axios or Fetch for API calls to Stability AI.
  Webhook Integration for Twilio/WhatsApp to receive and respond in real time.

METHODOLOGY

Image Generation Using Stable Diffusion

- Diffusion Model Overview: NeuraPix uses Latent Diffusion Models (LDMs). These models work by learning to denoise random noise into meaningful images, guided by a text prompt in a compressed latent space.
- Workflow:
  Prompt → Text Encoder: The prompt is processed via CLIP (or similar encoder).
  Noise Sampling: The model begins with random noise.
  Denoising Steps: Gradually refines the noise based on prompt semantics.
  Latent → Image: Final latent is decoded into a full-resolution image.
  Output: The generated image is passed back to the bot system.
- Stability AI API: The entire process is handled through Stability AI's cloud-hosted API. Parameters like prompt, style, cfg scale, and seed can be customized.

Data Preprocessing

- Text Prompt Sanitization: Cleaned user-entered prompts via NLP preprocessing (removing irrelevant tokens, typos, and normalizing inputs) for more accurate image generation.
- Prompt Optimization: Enhanced vague prompts using NLP enrichment (like adjective injection, scene detailing) to produce higher-quality image outputs.

METHODOLOGY

Automation & Intelligence Features

- Prompt Guidance: The bot can suggest prompt ideas using NLP and predefined categories (e.g., “Type 1 for branding, 2 for anime art”).
- Multi-language Support (Optional)

1. Supports prompts in regional languages using translation models before passing to the image generator.
2. Rate Limiting & Queue Handling
3. Implemented queuing and throttling to handle multiple user requests efficiently without hitting API rate limits.

Continuous Model Enhancement

- User Feedback Loop: Collects user ratings (e.g., “👍 / 👎”) on image quality to analyze and improve prompt formats.
- Adaptive Prompt Tuning: Incorporates real-world feedback to improve how vague or short prompts are expanded automatically for better visuals.
- Monitoring & Logging: Logs prompts, generation time, success/failure status for system improvement and debugging.

LITERATURE SURVEY

Paper Title (Including Author Details, Year of publication, Conference/Journal	Methodology	Dataset Used	Observation of proposed methodology	Pros	Cons	Findings
"High-resolution Image Synthesis with Latent Diffusion Models" – *Rombach et al., CVPR 2022*	Used latent diffusion in compressed space to generate images efficiently	COCO, OpenImages	Reduced compute while maintaining high quality	Low memory, high resolution	May blur fine details	Enabled scalable image generation with fewer resources
Hierarchical Text-Conditional Image Generation with CLIP Latents" – *Ramesh et al., arXiv 2022*	Used CLIP and transformers to improve prompt-image alignment	Internal OpenAI datasets	Highly accurate text-to-image results	Excellent semantic alignment	Requires heavy compute	Foundation of DALL·E 2
Diffusion Models Beat GANs on Image Synthesis" – *Dhariwal & Nichol, NeurIPS 2021*	Improved denoising diffusion model with classifier guidance	CIFAR-10, ImageNet	Outperformed GANs in image quality	Stable, interpretable	Slower than GANs	Proved diffusion as a superior generative method
Photorealistic Text-to-Image Diffusion with Deep Language Understanding" – *Saharia et al., arXiv 2022*	Combined image diffusion with deep NLP models	LAION-400M	High realism and better prompt adherence	Rich language understanding	Potential bias in training	Strong baseline for photorealistic generation
Denoising Diffusion Probabilistic Models" – *Ho et al., NeurIPS 2020*	Introduced basic DDPM framework for generative tasks	CIFAR-10	Foundation for all diffusion-based models	Simple, effective	High training cost	Core model behind newer diffusion models
"ControlNet: Adding Conditional Control to Text-to-Image Diffusion Models" – *Lvmin Zhang et al., 2023*	Enabled edge, pose, depth-based control in generation	MS-COCO + custom controls	Greatly improved control over output images	Fine control, multi modal	Complexity in control input	Advanced interactive image creation
"GLIDE: Towards Photorealistic Image Generation and Editing with Text-guided Diffusion Models" – *Nichol et al., 2022*	Introduced editing + generation using diffusion + guidance	Public image-text pairs	Enabled image manipulation via prompts	Text-based editing	Limited resolution	Allowed creative flexibility
"Imagen: Photorealistic Text-to-Image Diffusion Models with Large Language Models" – *Saharia et al., 2022*	Combined LLMs with diffusion for detailed generation	Internal datasets	Higher photorealism than DALL·E	Deep NLP + image quality	Not open-source	Major milestone in realistic image generation
"Versatile Diffusion: Text, Images and Beyond" – *Kim et al., 2023*	Unified multi-modal inputs (text, sketches) for generation	MS-COCO, LAION	Handles multiple input types	Flexible interface	Still under research	Useful for creative tools
"Stable Diffusion: High-resolution Image Synthesis using Latent Text-to-Image Diffusion" – *Stability AI, 2022*	Combines U-Net + CLIP + latent space diffusion	LAION-5B	Open-source, highly customizable	Fast inference, public availability	May lack high semantic depth	Enabled community-driven generative tools

IMPLEMENTATION

CONCLUSION

- NeuraPix harnesses the power of Generative AI and NLP to simplify and accelerate the image creation process.
- Empowers users—especially artists, marketers, and designers—to generate high-quality, customized visuals instantly through a simple WhatsApp chat.
- Eliminates the need for complex tools or technical skills, making content generation more accessible and efficient.
- Supports a wide range of use cases including branding, gaming assets, product design, and promotional material creation.
- By integrating with WhatsApp, NeuraPix ensures ease of use, real-time generation, and instant sharing—bridging the gap between creativity and technology in a conversational interface.

REFERENCES

Research Papers & Articles

- Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

- Ramesh, Aditya, et al. "Hierarchical text-conditional image generation with CLIP latents." arXiv preprint arXiv:2204.06125, 2022.

- Saharia, Chitwan, et al. "Photorealistic text-to-image diffusion models with deep language understanding." arXiv preprint arXiv:2205.11487, 2022.

- Dhariwal, Prafulla, and Nichol, Alex.
  "Diffusion Models Beat GANs on Image Synthesis."
  Advances in Neural Information Processing Systems (NeurIPS), 2021.

- Ho, Jonathan, Jain, Ajay, and Abbeel, Pieter.
  "Denoising Diffusion Probabilistic Models."
  Advances in Neural Information Processing Systems (NeurIPS), 2020.