Wish a Pic

By:

Omar Ayman Abdallah	**[IS]**
Moataz Mohamed Abdelmonem	**[IS]**
Nouran Tarek AbdelAzim	**[IS]**
Mohamed Ayman Mohamed	**[IS]**
Hager Ahmed Hassan Alfakharany	**[IS]**
Abdallah Mohamed Shaban	**[IS]**

Under Supervision of:

Dr. Mona AbdelAzim

Assistant Professor,

Information systems Department,

Faculty of Computer and Information Sciences,

Ain Shams University.

T.A. Engy Abdallah

Teaching Assistant,

Information systems Department,

Faculty of Computer and Information Sciences,

Ain Shams University.

July 2025

Acknowledgment

We would like to express our deepest gratitude to

Dr. **Mona Abdelazim** and T.A. **Engy Abdallah** for their continuous support, guidance, and encouragement throughout our graduation project.

Their valuable insights, dedication, and belief in us have been a driving force behind our success. Dr. Mona's sharp guidance and thoughtful direction, along with T.A. Engy's kindness and motivating spirit, helped us overcome challenges and grow both academically and personally.

This project would not have been possible without their support.

We are truly thankful for their time, effort, and belief in our potential.

Thank you, Dr. Mona Abdelazim and T.A. Engy Abdallah, for being such inspiring mentors and for making a meaningful difference in our journey.

Abstract

Wish a Pic is an AI-powered web platform that enables users to generate customized, high-quality visuals such as logos, posters, and advertisements using only simple text or voice prompts. Built to serve individuals with or without design experience, the system combines multiple artificial intelligence models to simplify and accelerate the design process. It integrates **Stable Diffusion XL Base 1.0** for image generation, **NLP-based** prompt enhancement using **spaCy** and **sentiment analysis** via **RoBERTa**, and Arabic-English translation with **Helsinki-NLP**. The platform supports both Arabic and English inputs in text and voice formats, ensuring accessibility across a broader user base. Users can interact with additional features like background removal via **rembg**, on-demand image editing through **SDXL's img2img** capability, and optional similarity detection using **CLIP** to ensure brand safety and originality. A custom-built logo dataset is used for similarity comparison to avoid unintentional copying of existing brand assets. Key functions such as Favorites and History allow users to manage, revisit, and refine their designs. The system prioritizes usability, performance, and accessibility, making it especially valuable for small businesses, freelancers, students, and creators with limited design resources. This thesis outlines the motivation, architecture, AI model integration, and user-focused implementation of the platform. The project demonstrates how AI can lower barriers in visual design by supporting creativity, reducing cost, and saving time. Wish a Pic empowers users to turn abstract ideas into meaningful visuals. The evaluation results showed that the average CLIPScore after prompt enhancement reached **0.4149** compared to **0.3608** before the enhancement, with **87.7%** of the images showing improved alignment with the prompts. Additionally, the survey results indicated that **89.38%** of participants preferred the enhanced images over the original ones, confirming the platform's effectiveness both technically and in terms of user experience.

الملخص (Arabic Abstract)

Wish a Pic هو منصة ويب مدعومة بالذكاء الاصطناعي تتيح للمستخدمين إنشاء تصاميم مرئية مخصصة وعالية الجودة مثل الشعارات، والملصقات، والإعلانات، باستخدام أوامر نصية أو صوتية بسيطة فقط. تم بناء النظام لخدمة الأفراد سواء كانوا يمتلكون خبرة في التصميم أم لا، حيث يجمع بين عدة نماذج من الذكاء الاصطناعي لتبسيط وتسريع عملية التصميم. يدمج النظام لتوليد الصور نموذج **Stable Diffusion XL Base 1.0**، وتحسين الأوامر النصية باستخدام مكتبة **spaCy** لتحليل المكونات اللغوية، ونموذج **RoBERTa** لتحليل المشاعر، بالإضافة للترجمة بين العربية والإنجليزية باستخدام نموذج **-Helsinki NLP**. تدعم المنصة الإدخال الصوتي والنصي باللغتين العربية والإنجليزية، مما يعزز من إمكانية الوصول لمجموعة واسعة من المستخدمين. كما يمكن للمستخدمين الاستفادة من ميزات إضافية مثل إزالة الخلفية باستخدام مكتبة **rembg**، وتحرير الصور عند الطلب من خلال خاصية **img2img** في نموذج **SDXL**، بالإضافة إلى إمكانية التحقق من التشابه باستخدام نموذج **CLIP** لضمان أصالة التصميم وتوافقه مع الهوية البصرية الخاصة بالعلامة التجارية. يتم استخدام مجموعة بيانات مخصصة للشعارات الشهيرة لمقارنة التشابه وتفادي التقليد غير المقصود لتصاميم علامات تجارية معروفة. وتشمل الوظائف الأساسية للمنصة ميزة المفضلات والسجل التي تتيح للمستخدم حفظ الأعمال السابقة وإعادة استخدامها أو تعديلها لاحقًا. يركز النظام على سهولة الاستخدام، والأداء، وقابلية الوصول، مما يجعله مفيدًا بشكل خاص للشركات الصغيرة، والمستقلين، والطلاب، والمبدعين الذين لديهم موارد محدودة في التصميم. تستعرض هذه الرسالة دوافع المشروع، وبنيته، ودمج النماذج الذكية فيه، وتنفيذه بشكل يركز على المستخدم. ويبرهن المشروع على قدرة الذكاء الاصطناعي في تخفيض الحواجز أمام التصميم المرئي من خلال دعم الإبداع، وخفض التكاليف، وتوفير الوقت. **Wish a Pic** يمكّن المستخدمين من تحويل الأفكار المجردة إلى تصاميم مرئية هادفة. وقد أظهرت نتائج التقييم أن متوسط درجة CLIPScore بعد تحسين الأوامر النصية بلغ **0.4149** مقارنة بـ **0.3608** قبل التحسين، حيث شهدت **87.7٪** من الصور تحسنًا في التوافق مع الأوامر. كما أظهرت نتائج الاستبيان أن **89.38٪** من المشاركين فضّلوا الصور الناتجة بعد التحسين مقارنة بالصور الأصلية، مما يؤكد فعالية المنصة من الناحيتين التقنية وتجربة المستخدم.

Table of contents

Table of contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Stands for
APIs	Application Programming Interfaces
CardiffNLP's RoBERTa	Robustly Optimized BERT Approach (by Cardiff NLP)
CLIP	Contrastive Language–Image Pretraining
SDXL	StableDiffusionXL
GAN	Generative Adversarial Network
Helsinki-NLP/opus-mt-ar-en	Open Parallel Corpus Machine Translation (Arabic to English)
LTR	Left-to-Right
RTL	Right-to-Left
NLTK	Natural Language Toolkit
OrderedDict	Ordered Dictionary (Python Collections Module)
Regex	Regular Expressions
rembg	Remove Background (Python library)

Chapter 1: Introduction

1.1 Motivation

In today's world, visuals are one of the most powerful ways to communicate ideas. Logos, social media posts, ads, and posters all help people and businesses share messages, attract attention, and build their identity. But many people especially small business owners, students, and freelancers struggle to turn their ideas into professional designs. Most don't have experience with design software, and hiring designers can be expensive and time-consuming. They often end up using basic templates that don't really reflect their ideas or brand. That's where Wish a Pic comes in. It's designed to make the creative process easy for everyone, no matter their background or skill level. By using AI, the platform helps users turn simple descriptions into high-quality, customized visuals. With just a few words or voice input, users can create designs that look professional without needing to learn complex tools or pay for expensive services. It's fast, simple, and completely free. We believe creativity should be available to everyone, not just design experts. Wish a Pic makes this possible by opening the door for anyone to bring their ideas to life and express themselves visually with confidence.

1.2 Problem Definition

In today's digital world, there's a growing need for creative and attractive visuals. However, many small business owners, freelancers, students, and non-designers face several challenges:

- **Lack of Design Skills:** Many people don't have the expertise needed to create professional designs, so they rely on expensive services or generic templates that don't stand out.

- **High Costs and Time:** Professional design tools and services can be expensive, and not everyone has the time or budget to hire designers or learn complicated software.
- **Limited Creative Tools:** Non-designers often struggle to express their ideas visually, making it hard to create content that truly represents their vision.

These challenges make it difficult for people and businesses to create high-quality designs when they need them especially when working with tight deadlines, limited budgets, or no access to design tools. As a result, many ideas remain unexpressed, or the final visuals fail to reflect the user's true vision. Wish a Pic solves this problem by offering an easy, fast, and free way to generate custom designs based on anyone's ideas, with no design experience required.

1.3 Objective

The main goal of **Wish a Pic** is to help people create high-quality designs easily, quickly, and without spending money. It's made for anyone whether they're business owners, students, freelancers, or just someone with an idea no design experience needed.

Our key objectives are:

- **Simplify Design Process:** Allow users to generate high-quality designs (like logos, posters, and ads) by simply inputting a few descriptive words, removing the need for complex design skills.

- **Save Time and Money:** Wish a Pic is completely free and generates designs in just minutes, helping users avoid expensive software, paid designers, and long waiting times.

- **Support Creativity:** The platform helps users express their ideas visually and gain confidence in their creativity, whether they are professionals or not

1.4 Methodology

To build Wish a Pic, we followed a simple and effective plan using modern web technologies to make the platform fast, easy to use, and accessible to everyone:

- **Understanding User Needs:** We focused on users who don't have design experience but need visuals like small business owners and students. They need something easy, fast, and free.

- **Frontend Development:** We used **Angular** to create a smooth and user-friendly website where users can type their ideas and instantly see AI-generated designs.

- **Backend Development:** We used **.NET** to handle everything behind the scenes like processing user inputs, communicating with the AI model, and sending results back to the user.

- **Database Management:** We used **SQL Server** to store user data, design history, and other important information safely and flexibly.

- **AI Integration:** We connected the system with **Stable Diffusion XL Base 1.0** to turn text prompts into high-quality images. We also added features such as **Feature Extraction** to improve the prompts quality

and match the user's style, **Similarity Detection** to distinguish generated visuals, **Image Editing** and **Background Removal**.

- **Testing & Improvements:** We tested the system with real users to make sure it was easy to use and gave great results. Based on feedback, we improved performance and output quality.

1.5 Time Plan
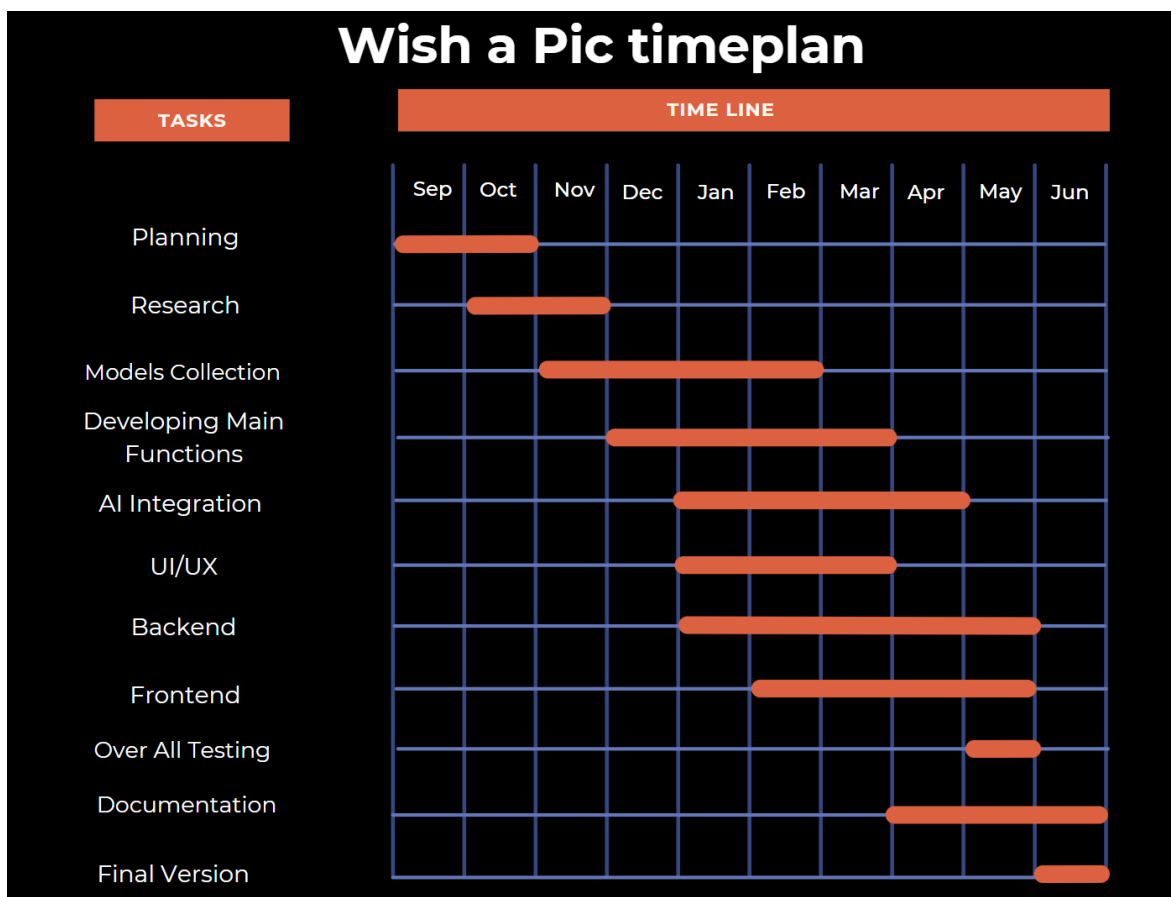
Figure 1.1: Time plan

1.6 Thesis Outline

This document is organized into six main chapters, each addressing a specific aspect of Wish a Pic project:

Chapter 1: Introduction

Provides an overview of the project, including the motivation behind developing "Wish a Pic", the problems it addresses, the objectives, adopted methodology, and the timeline of the project.

Chapter 2: Literature Review

Explores the context and background of AI in graphic design. It discusses the limitations of traditional design tools, the capabilities of advanced AI models, and compares Wish a Pic with existing AI-based design platforms. Key topics include:

- The Emergence of AI in Modern Graphic Design

- Challenges with Conventional Design Tools

- Better Prompt Understanding through Feature Extraction

- Maintaining Brand Identity with Similarity Detection

- Comparative analysis with similar tools

Chapter 3: System Architecture and Methods

Describes the technical structure and workflow of the system. This chapter includes:

- Overall system architecture

- Functional and non-functional requirements

- Explanation of key Functionalities such as:

- User input handling (text/voice)

- Prompt enhancement

- AI image generation (including Arabic language processing)

- Similarity detection

- Image editing

- History and favorites

- Description of system users and their characteristics

- Detailed use case, sequence diagrams

Chapter 4: System Implementation

Details the implementation phase of the project, including:

- Datasets

- Frontend and backend technologies used

- Integration of AI models

- Database setup

Chapter 5: Run the Application

Demonstrates how the system operates in practice, including how user inputs are transformed into output images. This includes screenshots or scenarios to showcase system behavior in real-world-like conditions.

Chapter 6: Conclusion and Future Work

Summarizes the outcomes and contributions of the project. It also addresses the challenges faced during development and provides suggestions for future improvements and possible extensions of the platform.

Chapter 2: Literature Review

2.1 The Emergence of AI in Modern Graphic Design

In recent years, artificial intelligence (AI) has emerged as a transformative force within the graphic design industry, offering users the ability to create high-quality visuals without the need for advanced design skills or technical expertise. One of the most groundbreaking advancements in this space is text-to-image generation, which allows users to input descriptive text and automatically generate images that match their vision. This technology has significantly reduced the barriers to entry for businesses and individuals who lack the resources or knowledge to create professional designs, enabling them to produce custom visuals quickly and affordably. [5]

2.2 Challenges with Conventional Design Tools

While tools such as **Canva** and **Adobe Express** have made basic design more accessible, they still rely heavily on user input and prior design knowledge. These platforms often require understanding of layout principles, color theory, and typography, and users must manually adjust templates to achieve the desired result. Moreover, most operate under paid subscription models, creating financial barriers for small businesses

or independent creators. As a result, these platforms fall short in delivering fast, affordable, and intuitive solutions for non-designers. [1]

2.3 Strengths and Limitations of Existing AI Models

Advanced AI models like **DALL·E**, **Midjourney**, and **Stable Diffusion** have demonstrated the potential of converting natural language into high-quality images. However, their performance is highly dependent on prompt quality, and they often struggle with capturing specific branding requirements or producing visually cohesive content. Although technically impressive, these models can generate inconsistent results, particularly when users are unfamiliar with how to structure effective prompts or seek outputs aligned with a unique brand identity. [2] [6] [17] [18]

2.4 Wish a Pic: A Context-Aware AI Design Solution

Wish a Pic addresses these limitations through its integration of Stable Diffusion XL Base 1.0, a state-of-the-art open-source model optimized for high-resolution, detailed image generation. This model supports scalable customization and efficiently converts user prompts into refined, brand-consistent visuals. Unlike generic generation tools, Wish a Pic is tailored to provide professional-quality output by understanding context and user intent. [12]

And here is why we choose **Stable Diffusion XL Base 1.0** rather than other **Stable Diffusion** versions which are also open-source and could be used as the core text to image model in Wish a Pic such as:

- **Stable Diffusion 1.5**
- **Stable Diffusion 2.1**
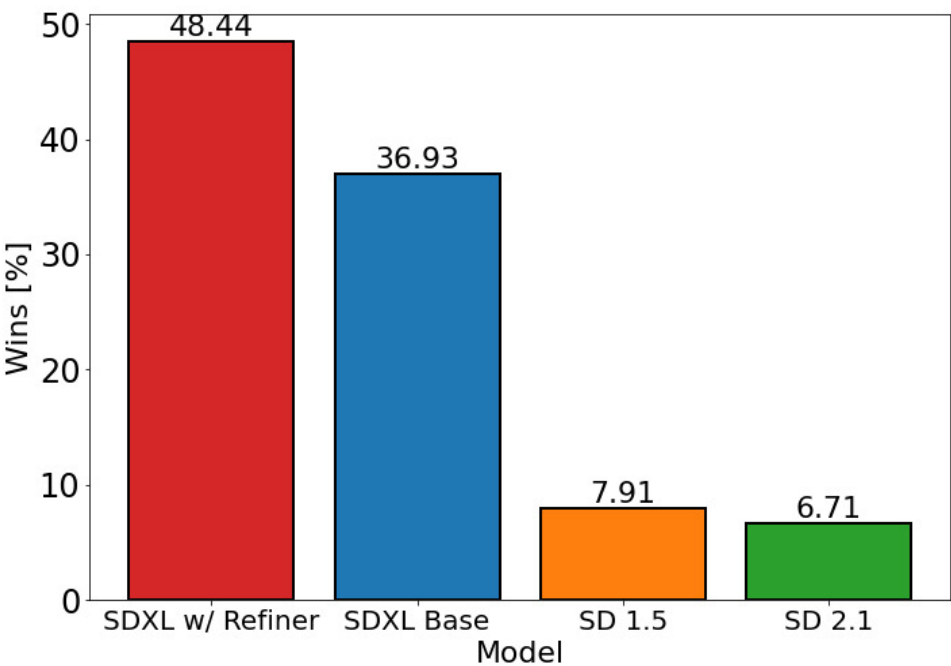
Figure 2.1: User Preference Comparison

To better understand why Stable Diffusion XL Base 1.0 was chosen as the core model for Wish a Pic, it's important to look at how it compares to other versions of the Stable Diffusion model family. The table below presents a technical comparison between **Stable Diffusion 1.5, 2.1, and XL**.

Stable Diffusion XL stands out with significantly higher parameter count and improved context handling, thanks to dual text encoders (OpenCLIP ViT-L and OpenCLIP ViT-bigG) and a broader context dimension limit. These enhancements allow it to interpret longer, more complex prompts with greater accuracy and detail making it the most capable model for high-quality, customized image generation.

Table 2.1: Stable Diffusion Versions Technical Comparison

POC / Version	Stable Diffusion 1.5	Stable Diffusion 2.1	Stable Diffusion XL
# of UNet Parameters	860 M	865 M	2.7 B
Text Encoder	OpenCLIP ViT-L	OpenCLIP ViT-H	OpenCLIP ViT-L & ViT-bigG
Context Dimension	768	1024	2048
Pooled Text Embedding	N/A	N/A	OpenCLIP ViT-bigG

So, by utilizing **Stable Diffusion XL Base 1.0**, Wish a Pic overcomes the limitations of other design tools and AI models, offering a more refined and efficient way to generate images that align with user specific needs.

2.5   Better Prompt Understanding through Feature Extraction

A key differentiator in Wish a Pic is its prompt enhancement pipeline, which utilizes natural language processing (NLP) and feature extraction to improve the clarity and intent of user inputs. Through models like spaCy and RoBERTa, the system identifies important entities, sentiment, and structure in the prompt. This results in outputs that are more aligned with the user's goals. Even with complex or abstract inputs, Wish a Pic ensures that the resulting visuals are coherent, visually compelling, and contextually accurate eliminating the need for users to craft precise prompts manually. [19] [21]

2.6   Maintaining Brand Identity with Similarity Detection

Wish a Pic incorporates a similarity detection system powered by the CLIP model to preserve brand identity and reduce the risk of copyright infringement. This feature automatically compares new designs with existing visual assets, ensuring stylistic consistency across a user's branding materials. It also evaluates similarity scores against a curated dataset of well-known brand logos. If a generated design closely resembles an existing trademarked item, the system alerts the user or prevents the generation altogether. This not only helps users maintain their own brand standards but also provides a proactive safeguard against legal issues. [29]

2.7 Free , Simple and Fast Design Generation for All Users

One of Wish a Pic's most significant contributions is its accessibility. The platform is fully free and removes financial barriers common in other tools. It allows users to generate professional visuals in a matter of minutes, eliminating the need for expensive subscriptions, time-consuming revisions, or third-party designers. This makes it especially beneficial for startups, freelancers, and students working under tight constraints.

2.8 Visual Inspiration and User-Guided Refinement

Wish a Pic serves not only as a generator but also as a source of visual inspiration. For users who lack design experience, the initial output provides a visual reference that helps them better understand their own ideas. To support creative iteration, the platform integrates interactive features such as image editing and background removal. These tools allow users to isolate elements, adjust styles, and fine-tune results encouraging hands-on learning and deeper engagement with the design process. The system thus transforms users from passive consumers into active creators, boosting both confidence and creativity.

2.9   Comparative Analysis with Similar Applications

As AI-generated design tools continue to grow in popularity, several platforms have emerged to help users transform text into visuals. Among the most widely recognized are **DALL·E**, **Deep Fusion GAN**, and **DreamStudio** each offering unique features, capabilities, and approaches to converting user prompts into compelling images.

DALL·E: developed by OpenAI, is known for its ability to generate highly creative and imaginative visuals from text, including surreal or abstract scenes. It supports inpainting and editing features, allowing users to modify parts of an image while preserving context. [2] [3] [4]

Deep Fusion GAN: is a research-focused model that excels in combining multiple data inputs such as text, sketches, or reference images to generate realistic visuals with enhanced texture blending. [10]

DreamStudio: the official interface for Stable Diffusion, offers users a more interactive and user-friendly experience for creating images. It provides fine control over parameters such as resolution, prompt weight, and guidance scale, making it a popular choice for creators who want both speed and customization.

In this section, we compare these platforms with **Wish a Pic** to highlight how our project stands out in addressing the practical needs of everyday users.

This comparison focuses on six key areas:

- **UI Usability:** How easy and intuitive the platform is to use.

- **Content Manipulation:** Features like saving favorites or accessing previous generations.

- **Prompt Enhancement and Preprocessing:** Whether the platform helps users refine or improve vague descriptions.

- **Similarity Detection:** The ability to maintain brand consistency and avoid copyright issues.

- **Arabic Prompt Support:** How well the platform handles Arabic language input.

- **Free Access :** Users can access and use the platform 100% for free.

By analyzing these factors, we aim to show how Wish a Pic not only matches the functionality of popular platforms but also offers thoughtful improvements that make AI design tools inclusive, user-friendly, and supportive of real-world creative needs.

Table 2.2: Related Works Comparison

POC / Title	DALL-E	Deep Fusion GAN	DreamStudio	Proposed System
UI Usability	✓	✗	✓	✓
Content Manipulation	✗	✗	✓	✓
Prompt Enhancement	✗	✗	✗	✓
Similarity Detection	✗	✗	✗	✓
Background Removal	✓	✗	✗	✓
Arabic Prompt Support	✓	✗	✓	✓
Free Access	✗	✓	✗	✓

2.10 Conclusion: Redefining Design with Wish a Pic

In conclusion, **Wish a Pic** represents a significant advancement in the way we approach graphic design. By combining the power of SDXL Base 1.0 with smart prompt enhancement, accurate similarity detection, multiple languages support, image editing services and a user-friendly interface, the platform offers an innovative solution to the challenges faced by traditional design tools and AI-driven models. Unlike other platforms that require users to have design experience or pay expensive subscription fees, Wish a Pic provides an accessible, free, and highly efficient way for anyone whether a beginner or a seasoned designer to create professional-quality visuals. With its focus on customization, brand alignment, and time-saving capabilities, Wish a Pic is uniquely positioned to empower small businesses, content creators, and individuals by eliminating design complexity, reducing costs, and ensuring brand consistency, **Wish a Pic** empowers businesses to create professional visuals effortlessly transforming how they market, communicate and grow.

Chapter 3: System Architecture and Methods

3.1 System Overview

3.1.1 System Architecture
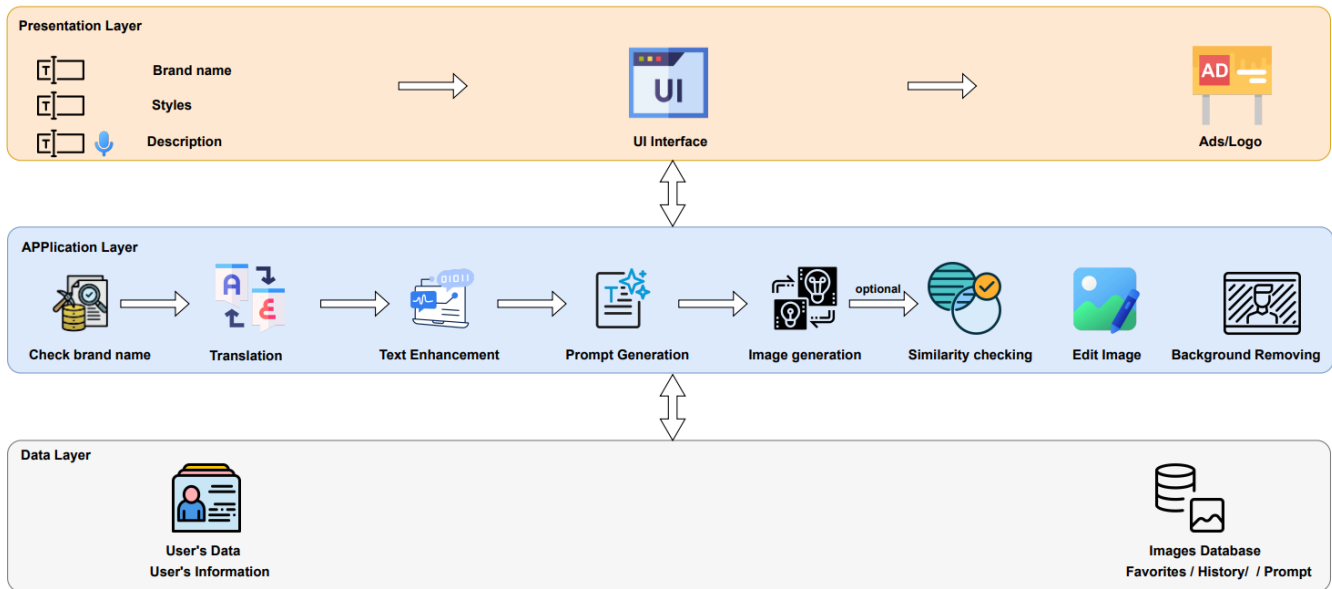
Figure 3.1: Wish a Pic System Architecture

This system architecture diagram presents a clearly layered and modular structure for Wish a Pic platform, divided into the Presentation Layer, Application Layer, and Data Layer. This separation of concerns supports scalability, usability, and maintainability throughout the platform.

- **Presentation Layer**

This top layer handles all user interactions, beginning with login or registration. Users then provide input in the form of brand name, style, and description. The UI guides them through the process of generating visuals,

with optional enhancements such as similarity checking, image editing, and background removal. This layer is designed to provide a simple and intuitive experience for users regardless of their design experience.

- **Application Layer**

This is the core logic layer where various AI models are integrated to process user input and generate tailored visual outputs:

1. **Brand Name Validation:** A custom rule-based system cross-references the input brand name with a curated dataset of real-world brand names to protect brands identity.

2. **Translation:** Utilizes the Helsinki-NLP/opus-mt-ar-en model for translating Arabic prompts into English while preserving meaning and tone.

3. **Text Enhancement:** Applies spaCy (en_core_web_sm) for part-of-speech tagging and CardiffNLP/twitter-roberta-base-sentiment for sentiment analysis. This helps better understand user intent and tone. [19] [21]

4. **Prompt Generation:** Enhanced and structured prompts are produced using NLP pipelines that combine keyword extraction and descriptive pattern generation, ready for image synthesis.

5. **Image Generation:** Employs SDXL Base 1.0, a high-resolution text-to-image model, to generate visuals such as logos, ads, and posters from the enhanced prompt. [12]

6. **Similarity Checking** (optional): Uses CLIP model to compare the generated image embedding with a custom dataset of brand logos using cosine similarity. This ensures the design is original and avoids legal risks. [29]

7. **Image Editing** (optional): Uses the img2img capability of SDXL to modify or regenerate specific parts of an existing image while keeping the overall concept intact. This allows users to refine outputs such as adjusting composition, structure, or style without starting from scratch.

8. **Background Removal** (optional): Relies on the rembg Python library, which is built on the U-2-Net segmentation model, to remove backgrounds from generated images. This makes it easy for users to isolate logos or objects for use in other contexts, like marketing materials or social media designs.

This layer showcases how the platform smartly integrates multiple pretrained AI models to deliver intelligent, context-aware, and brand-safe design outcomes.

- **Data Layer**

The data layer is responsible for storing all user-related data It also holds the database of generated images, saved prompts, generated image history, and favorite designs, this separation ensures secure, scalable, and organized data management, and allows easy expansion into cloud or distributed systems in the future.

3.1.2 Functional Requirements (methods used):

1. User Authentication (Login & Register)

Functionality:

Allows users to create an account or log in to access personalized features like favorites and history.

How it works:

- Backend authentication is handled using secure .NET APIs.

- Registered Users data is stored and managed using SQL server along with their favorites and history.

2. Description Input (Voice or Text)

Functionality:

Allows users to enter their image description either by typing it manually or by using voice input through the microphone.

How it works:

Typing Input: Users can manually type their description into a text box. The system supports both English and Arabic text, automatically sets text direction based on the selected language (RTL for Arabic, LTR for English).

and processes the input directly or sends it for translation if needed.

Voice Input (Optional): Users who prefer not to type can click the microphone icon and describe their image verbally.

- Uses the Web Speech API (SpeechRecognition or webkitSpeechRecognition).
- Displays recognized speech live in the input box.
- Handles unsupported browsers and blocked microphone access with clear error messages. [23] [24]

3. User Input Enhancement and Preprocessing

Functionality:

Improves and adapts user input into a more effective prompt for the image generation model.
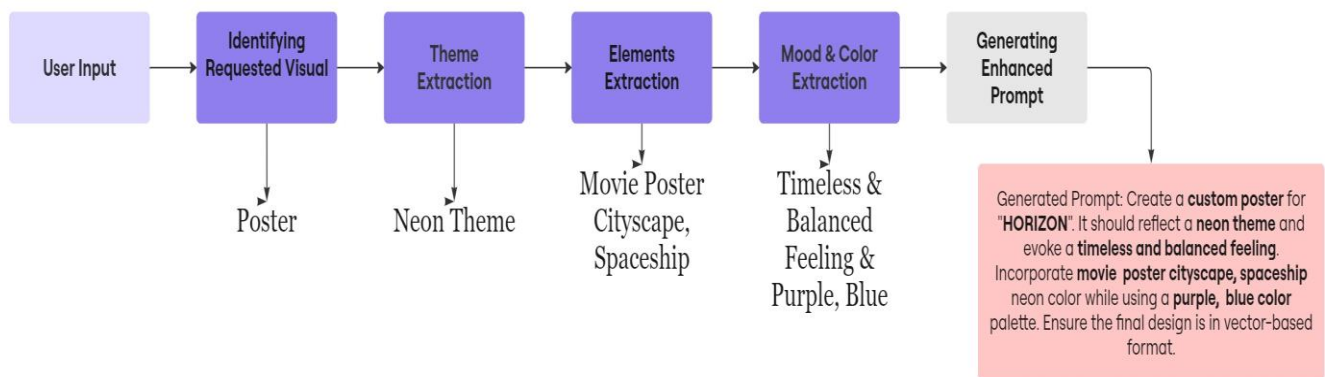
How it works:

- **Feature Extraction:**

  - **spaCy (en_core_web_sm)** for part-of-speech tagging, noun chunking, and lemmatization. [21]

  - **NLTK** for stopwords removal.

  - **Regex** to extract colors.

- **Sentiment Analysis:**

  - Uses **CardiffNLP's RoBERTa** model to identify the emotional tone (positive, negative, neutral). [19]

  - Sentiment is used to select matching design moods.

- **Prompt Generator:**

  - Extracts design type (e.g., logo, poster) and relevant features (theme, color, mood).

  - Assembles a refined prompt with clear structure for the model.

- **Redundancy Cleaner:**

  - Uses **regex** and **OrderedDict** to remove repeated words and phrases.

User Input: "generate a sci-fi movie poster with a cityscape, a spaceship, and a blue and purple neon color palette."

Figure 3.2: Prompt Enhancement Flow Chart

4. Image Generation from Arabic Text

Functionality:

Generates images based on user prompts written in Arabic.

How it works:

- Uses the **Helsinki-NLP/opus-mt-ar-en** model to translate Arabic text to English.
- Detects Arabic characters using **regex**.
- Replaces embedded English terms with placeholders before translation to preserve accuracy (e.g., brand names, hashtags).
- Re-inserts original English words after translation.
- Translated prompt is passed to the prompt enhancement then **Stable Diffusion XL Base 1.0** model for image generation.

5. Image Generation (via SDXL)

Functionality:

Converts enhanced text prompts into images.

How it works:

- Uses **Stable Diffusion XL Base 1.0**, a powerful open-source text-to-image model. [12]
- Accepts refined prompts and generates high-resolution visuals.

- Supports different output types (e.g., logo vector, raster image, product ad).

To understand how **Stable Diffusion XL Base 1.0** generates images from user prompts, it's important to look at how text-to-image generation works under the hood. The following diagram illustrates the simplified architecture of a diffusion-based model specifically how a user's text input is converted into a final visual output using components like **CLIP**, **U-Net**, **noise Scheduler** and **image decoder**. [
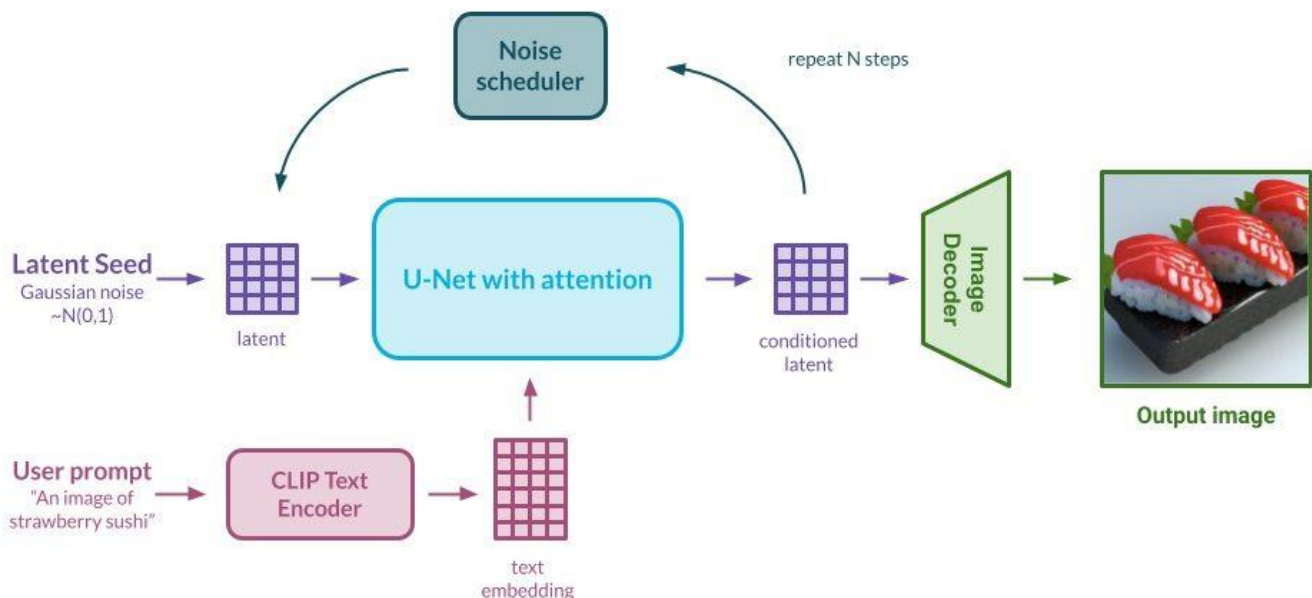
Figure 3.3: Stable Diffusion Architecture

The image generation process in Wish a Pic is based on diffusion models, particularly Stable Diffusion XL. These models convert a user's written prompt into a high-quality image using a series of neural network components. The main steps are:

- **User Prompt Input:** The process starts with the user typing a description, such as "An image of strawberry sushi." This prompt is sent to a CLIP Text Encoder, which transforms the text into a machine-understandable format called a text embedding. [29]

- **Noise Generation (Latent Seed):** At the same time, the system creates a random starting point in the form of Gaussian noise essentially a blurry, meaningless image. This noisy image is represented as a latent vector.

- **U-Net with Attention:** The U-Net model, guided by the text embedding and a noise scheduler, works to gradually remove the noise from the latent image. This process is repeated over several steps, slowly "shaping" the noise into an image that reflects the user's description.

- **Conditioned Latent:** As the model continues, the noisy latent becomes a conditioned latent, meaning it now reflects the structure, colors, and concept described in the prompt.

- **Image Decoder:** The final latent image is passed into an image decoder, which transforms it from encoded values into a visual format a full-resolution image.

- **Output Image:** The result is a realistic image that matches the user's prompt, such as the generated picture of strawberry sushi. This image can then be saved, edited, or reused.

6. Similarity Detection

Functionality:

Protects users from generating images that unintentionally copy or closely resemble existing logos or brand visuals, as it prevents accidental duplication, protect against copyright issues. This helps avoid legal risks and maintains brand originality and uniqueness.

How it works:

Before the image is generated, the system checks whether the brand name the user used matches any existing brand names or logos in a predefined dataset. This early step is important to prevent users from unintentionally recreating protected or copyrighted content.

Then after creating the request design,

The system uses CLIP model for comparison because it understands both the visual style and the concept behind an image. CLIP can detect subtle similarities in design, theme, and symbolism something basic visual models can't do.

The process consists of 2 key steps:

- **Feature Extraction:**

  - Generated images are first resized and normalized.

  - Then, the image is passed through CLIP to extract a feature vector (embedding), which captures the image's visual and semantic information. [29]

  - These embeddings are stored and later used for comparison.

- **Similarity Calculation:**

  - The platform uses cosine similarity to compare the new image's features with those in the existing logo dataset.

  - A similarity score is calculated between using the **cosine similarity**

  - If the score is greater than 0.75, the image is flagged as "too similar" and may be rejected or warned.
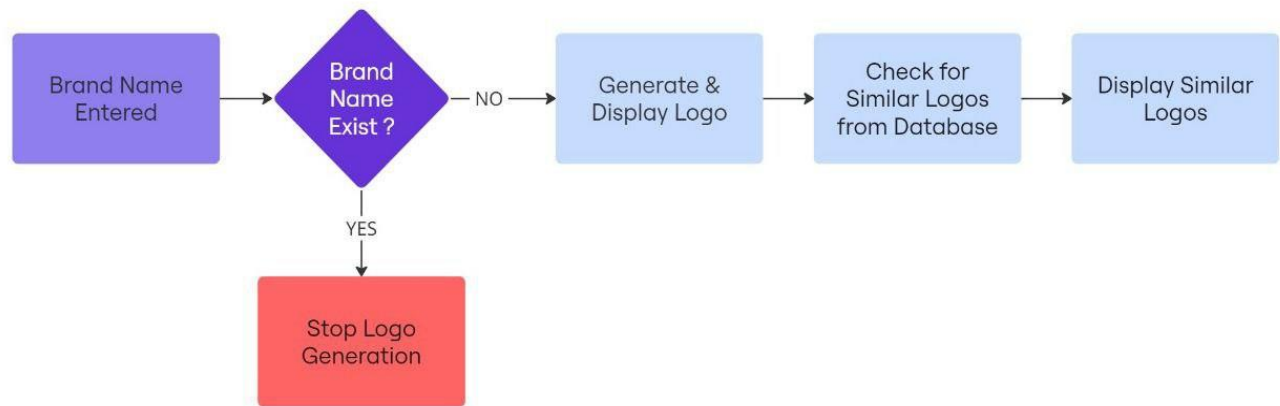
Figure 3.4: Similarity Detection Flow Chart

7. Image Editing (img2img via SDXL)

Functionality:

Allows users to edit or improve existing images using AI.

How it works:

- Uses img2img capability of SDXL.
- Users upload or select an existing image, then apply changes using a text prompt.
- Produces a modified version of the original image while preserving structure.

8. Background Removal

Functionality:

allows users to extract specific items such as logos and icons from images.

How it works:

- Using **.remove**() function from the python library **rembg** which utilizes a deep learning model to identify and isolate the item by removing the background which transform it into **PNG** images.

9. History

Functionality:

Allows users to view their previously generated designs.

How it works:

- Stores metadata (prompt, generation time, image preview) in SQL server.
- Registered users can access their design history for re-use or editing.

10. Favorites

Functionality:

allows users save and organize their favorite generated designs.

How it works:

- Logged-in users can mark any generated image as a favorite.
- These items are saved in the user's account and displayed in the favorites section.

3.1.3 Nonfunctional Requirements

- **Usability:** Wish a Pic is built with a user-friendly interface that allows even non-designers to create stunning visuals. With minimal effort, users can input their prompts and receive designs that align with their specifications, without needing any advanced design skills.

- **Performance:** The platform is designed to process image generation requests quickly and efficiently, enabling users to produce professional-quality visuals in minutes. The system handles multiple users simultaneously without delays, ensuring a smooth experience even during peak usage.

- **Security:** Security is a top priority for Wish a Pic. The platform takes great care to protect users' data and ensure privacy. Additionally, by using features like similarity detection, it minimizes the risk of generating designs that might unintentionally infringe on existing copyrights or trademarks, thus safeguarding users from potential legal issues.

- **Portability:** Wish a Pic is accessible across multiple devices with an internet connection, whether on a computer or mobile. The platform works consistently on different operating systems, ensuring users

can generate designs on the go, without being tied to a specific device or location.

- **Maintainability:** The platform is designed with maintainability in mind, ensuring that it can easily be updated and improved. Regular updates, bug fixes, and new features can be implemented without disrupting the user experience, ensuring that Wish a Pic continues to function optimally and evolve over time.

3.1.4 System Users

Wish a Pic is designed to serve a wide range of users who need fast, creative, and affordable design solutions without requiring professional design experience, the main users of the system include:

- **Small Business Owners:** Individuals who run small businesses and need logos, marketing posts, or promotional materials without hiring a designer.

- **Regular Users and Hobbyists:** Individuals who enjoy art and creativity and want to explore their imagination by generating visuals for personal use, self-expression, or sharing with friends.

- **Content Creators:** Social media influencers, YouTubers, and bloggers who need custom posters, or visuals to enhance their online content.

- **Startups:** Early-stage startups that need high-quality branding visuals, such as pitch decks, app icons, or presentation banners, while keeping costs low.

- **Students:** Users in academic settings who want to create posters, event flyers, or visuals for projects and student activities quickly and easily.

- **Freelancers and Side Hustlers**: People who sell products or online services and want to make attractive designs for ads, logos, or digital content to help their work stand out.

Each of these user types benefits from Wish a Pic's simple interface, powerful AI image generation, and free access to professional-quality designs.

3.2   System Analysis & Design

3.2.1 Use Case Diagram
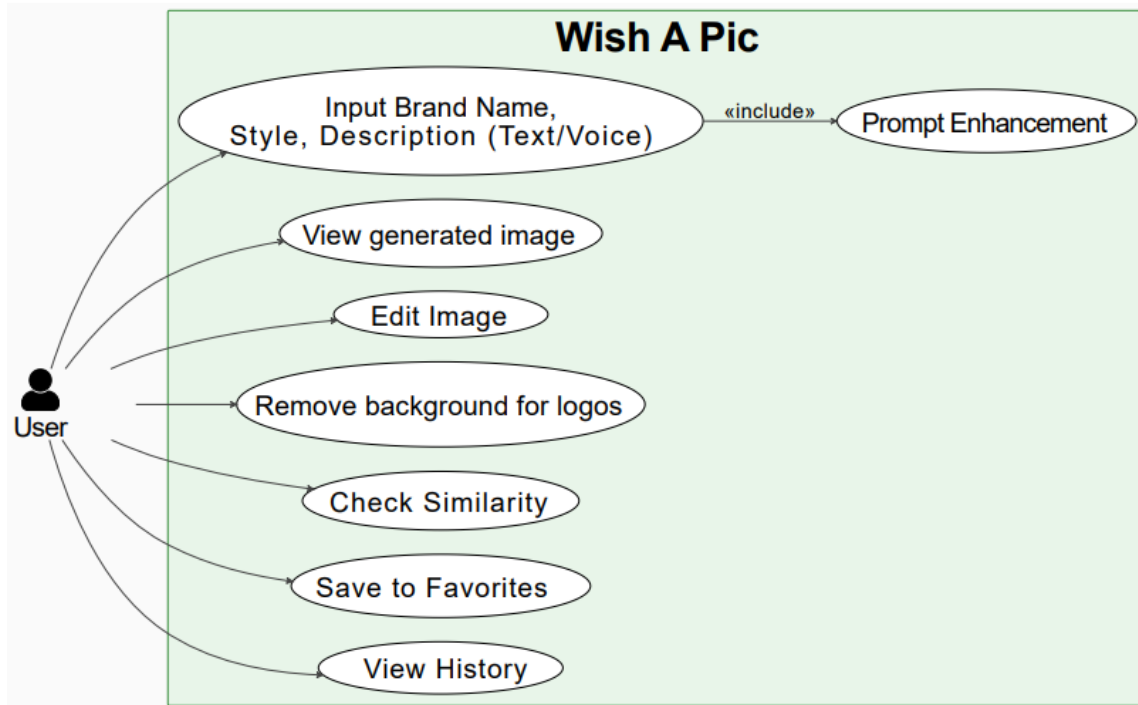
Figure 3.5: Use Case Diagram

This simplified use case diagram effectively captures the core interactions between the user and the Wish a Pic system. It highlights the essential features such as prompt input (via text or voice), prompt enhancement, image generation, editing tools, background removal, and brand similarity checking. Additionally, it includes user-centric functions like saving to favorites and viewing history, which enhance usability and personalization. While simple in structure, the diagram clearly communicates the user flow and reflects the system's goal of making AI-powered design both accessible and interactive for all users.

3.2.2 Sequence Diagrams
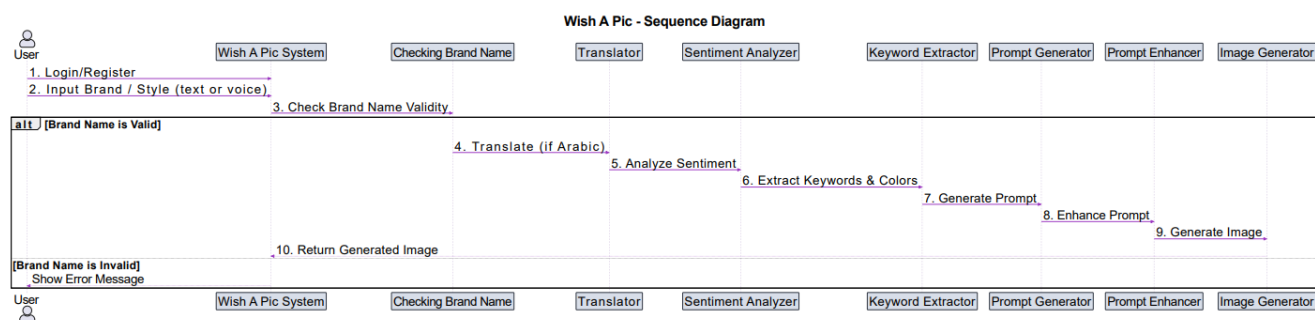
Figure 3.6: Wish a Pic Sequence Diagram

This sequence diagram effectively illustrates the core flow of the Wish a Pic system from user input to image generation while emphasizing modular AI integration. It captures the step-by-step interaction between the user and various backend components, such as brand name validation, translation (if needed), sentiment analysis, keyword extraction, and prompt enhancement. The diagram clearly distinguishes the two branches of logic depending on the validity of the brand name: proceeding to generation if valid, or returning an alert otherwise.

Each component is shown interacting in a logical, sequential order, making the system architecture transparent and easy to understand. This clarity not only helps demonstrate the intelligent processing behind the scenes but also highlights Wish a Pic's layered AI workflow. It's a solid representation of how the platform transforms raw user input whether in text or voice into a refined, brand-aware visual using smart language and vision models.

Chapter 4: System Implementation and Results

4.1 Datasets

Wish a Pic system leverages a collection of high-quality pre-trained models and curated datasets to power its intelligent image generation and analysis features.

SDXL & SDXL Img2Img: Both models are pretrained on diverse and large-scale image-text dataset with **2 Billion** images, and optimized for high-quality image generation and transformation tasks. They enable the creation of stunning visuals from text prompts and the refinement of existing images with creative flexibility.

Prompt Enhancement and Preprocessing: Uses several models such as:

- **spaCy – en_core_web_sm** which is pretrained with linguistic annotations from sources like OntoNotes, web news, blogs, and social media to apply POS tagging, dependency parsing, noun chunking. [21]

- **CardiffNLP – twitter-roberta-base-sentiment** which is pretrained on Twitter data for sentiment analysis. [22]

- We also used **Color Recognition Dataset** consists of **30,282** colors to extract meaningful color references from text prompts.

Translation Model (Helsinki-NLP/opus-mt-ar-en): Our translation system is fully pretrained as well, allowing it to accurately convert brand prompts across languages while preserving meaning, tone, and embedded design elements.

Similarity Detection: Uses **CLIP** model which is pretrained on **400 million** (image, text) pairs from publicly available datasets gathered from the internet. And to enable accurate similarity comparison, we created a **manually curated custom dataset** of well-known brand logos, organized by brand . This dataset includes:

- Logos from globally recognized companies (e.g., Apple, Nike, Adidas, McDonald's, etc.).

- Multiple logo versions per brand to account for style variations.

- High-resolution, clean logo images for embedding consistency.

These diverse, high-quality pretrained models and curated datasets form the backbone of **Wish a Pic's** intelligence. They ensure the platform delivers creative, accurate, and safe outputs while maintaining language flexibility, brand awareness, and visual precision.

4.2 Frontend and Backend

Frontend: Angular

The Wish a Pic frontend is developed using **Angular 17**, a modern, component-based framework for building single-page web applications (SPAs). Angular was chosen for its strong support for dynamic content rendering, responsive design, and maintainability, the Key aspects of the Wish a Pic frontend include:

- Login and Register interface
- Prompt Input and Design Characteristics Interface.
- Image Generation, Editing and Background Removal Interfaces.
- Favorites & History Interfaces.

Also The design is Responsiveness which makes it adapts across different screen sizes to support desktops, tablets, and mobile browsers.

Backend: .NET

The backend is built using **.NET** Core, a cross-platform, high-performance framework suitable for handling API communication, security, and model integration. It acts as the bridge between the frontend and the AI services. Its responsibilities include:

- **API Endpoints:** Expose RESTful APIs for handling prompt submissions, image generation requests, and user data interactions.

- **Authentication**: Secure user login, registration, and token management.

- **AI Model Coordination:** Although the backend itself is developed in .NET, the AI models (such as SDXL for image generation, CLIP for similarity detection, etc…) are hosted externally on a **Kaggle** Notebook. These models are exposed using a lightweight **Flask** server, and the connection is tunneled to the public web using **NGROK**. This allows real-time interaction with Python-based ML models from the **.NET** backend without the need for costly dedicated infrastructure.

- **Database Communication:** Handles interaction with a **SQL Server database** to store user information, prompt history, favorite images, and generation logs in a structured and secure manner.

This frontend-backend architecture ensures modularity, efficient performance, and a seamless flow between user interactions and AI processing.

4.3 AI Models Integration

To support advanced AI functionalities such as text-to-image generation, image editing, similarity detection, and background removal, the system integrates multiple deep learning models through a modular and scalable approach. Since these models are developed in Python and run on GPUs available through Kaggle, a bridge was built between the **.NET** backend and these models using **Flask** and **NGROK**.

The AI models are hosted on Kaggle Notebooks and exposed via lightweight Flask servers. Each model (or feature endpoint) is deployed independently and accessed through a secure NGROK tunnel, allowing external communication without deploying a permanent cloud server.

To enable this setup, the following key components are installed and configured inside the Kaggle environment. Each Flask application is initialized with Cross-Origin Resource Sharing enabled (**CORS**) to allow communication with the frontend

Image Generation: Uses the **SDXL** model to generate high-quality visuals from enhanced user prompts. [12]

Image Editing: Applies **SDXL Img2Img** to modify or refine existing visuals using an updated prompt, preserving composition while adapting the style or content.

Translation Into Arabic: Uses **Helsinki-NLP/opus-mt-ar-en** model which helps convert Arabic prompts into fluent English before feeding them to the prompt enhancement process and image generation.

Similarity Detection: Uses a **CLIP-based model** to compare generated visuals against known brand logos and flag outputs with high visual similarity. [29]

Background Removal: Processes generated images to remove their background using segmentation-based models utilized by the **rembg** python library.

We also used some python libraries to helps performing some major tasks such as:

- **spCay:** Used for linguistic analysis (tokenization, POS tagging, noun chunking, etc.) with the en_core_web_sm model. [21]
- **nltk:** For stopwords filtering (stopwords.words("english")).
- **re:** Used for regex-based parsing (e.g. extracting colors , removing redundancy).
- **torch:** For running inference on the sentiment model.
- **transformers:** Loads the cardiffnlp/twitter-roberta-base-sentiment sentiment analysis model. [22]
- **rembg:** to apply the background removal functionality

4.4 Database Setup

The database structure of the Wish a Pic platform is designed to support two core functions: user authentication and image management. To ensure both scalability and security, the system leverages the built-in **ASP.NET** Identity framework alongside custom relational tables to store and manage generated image data.

The database consists of two primary tables:

1. AspNetUsers Table

This table is responsible for storing user information and managing login credentials. It is based on the ASP.NET Identity schema but has been simplified to include only the fields relevant to this project. These include:

- **Id:** A unique string identifier for each user (Primary Key).
- **FullName**: The full name of the user.
- **UserName and Email:** Used for login and identity.
- **PasswordHash:** A hashed version of the user's password for secure authentication.
- **RefreshToken and RefreshTokenExpirationDateTime:** Used for session management and token renewal.

ASP.NET Identity ensures that authentication, password encryption, and security token handling are managed efficiently, reducing development overhead and enhancing protection.

2. Images Table

This table stores metadata related to generated images and links each image to its respective user:

- **ImageId:** Unique identifier for the image (Primary Key).
- **UserId:** Foreign key referencing AspNetUsers.Id, establishing a one-to-many relationship (one user can generate multiple images).
- **Prompt:** The textual prompt provided by the user to generate the image.
- **Image:** a base64 string that represents the generated image.
- **IsFavorite:** A Boolean flag indicating if the user marked the image as a favorite.

This schema allows users to save and organize their generated content and later retrieve it via the Favorites and History features.

4.5 Results

This section presents the outcomes of evaluating the Wish a Pic system, focusing on the effectiveness of the prompt enhancement process. Both automated scoring and user feedback were used to measure improvements in visual relevance and user satisfaction, this section also introduces some output samples to evaluate the image generation, similarity detection, background removal and image editing visually.

- **Prompt Enhancement Evaluation**

To evaluate the performance of the system, we used **CLIPScore** to measure the alignment between input text prompts for 100 generated images before and after the prompt enhancement.

The used version of CLIP is **Vit-H-14** which uses Cosine similarity to calculate CLIP-Score. The Cosine Similarity ranges from **-1** to **+1**, but the CLIPScore range **starts from 0 and rarely goes above 0.5** for realworld/generated content.

CLIPScore equation:

$$\textbf{CLIPScore(I , C)=max(100×cos(EI,EC),0)} \qquad \textbf{(4.1)}$$

where EI and EC are the image and text embeddings, respectively.

This means that a cosine similarity of 0.35 corresponds to a CLIPScore of ~35,

with a bound between 0 and 100

Detailed CLIPScore ranges:

- **0.00 to 0.25** means the image is a very poor matched to the prompt.

- **0.26 to 0.35** shows a partial match, with a minimal relevance to the prompt.

- **0.36 to 0.40** shows a good match, where the image reasonably represents the prompt.

- **0.41 to 0.**45 reflects a strong alignment, showing that the image closely matches the text.

- Scores above **0.45 are rare** and indicate an excellent or near-perfect match between the image and the prompt.

Using Stable Diffusion XL Base 1.0, we recorded an average CLIPScore of **0.3608** across a **100** prompt sample before applying the prompt enhancement system. After integrating the prompt preprocessing and enhancement, the average CLIPScore increased to **0.4149**.

This represents a measurable improvement in image-text alignment and overall generation quality. Moreover, **87.7%** of the evaluated prompts achieved a higher CLIPScore after enhancement, confirming the consistency of improvement. The remaining **12.3%** of prompts showed no significant change, which may be attributed to either overly simple descriptions or cases where the base model already performed well without preprocessing.

We also conducted a user feedback survey for **175** people that showed a significant appreciation for the results where **89.38%** preferred the generated images from enhanced prompts rather than those before the enhancement process, **7.23%** liked the images before prompt enhancement and the remaining **3.39%** found both attractive.
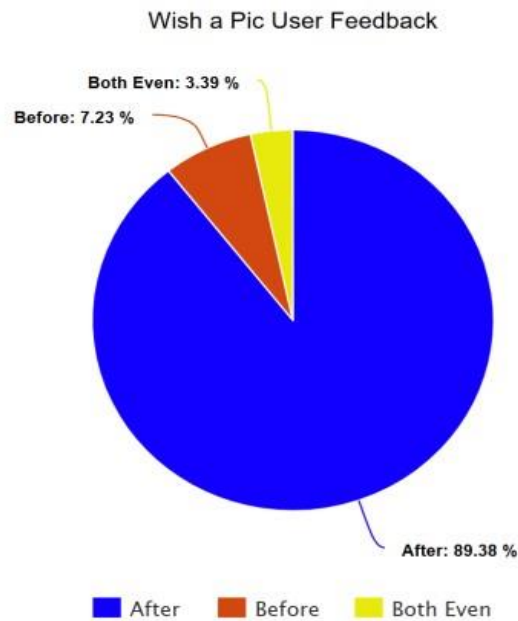
Figure 4.1: User Feedback

- **Image Samples**

These are some samples of Wish a Pic generated images using enhanced Prompts compared to generated images directly from SDXL base 1.0 without prompt enhancement, followed by a similarity detection, background removal and also image editing examples.

Figure 4.2: Cheese shop logo Before and After Prompt Enhancement

User Input: Cheese shop logo with stylized cheese wedge, yellow and rustic tones, flat vector style branding, vintage-modern design, yellow background

CLIPScore = 0.3051

Enhanced Prompt: Logo for "Cheese Shop" with (Vintage & Retro style). It should reflect a rustic flat modern theme and evoke timeless balanced feeling. Incorporate cheese shop logo, cheese, wedge, yellow tones, vector deli-style branding, vintage-modern design, white background, while using white, cheese, yellow, color palette. The final design must be vector-based, scalable design.

CLIPScore = 0.4845

Figure 4.3: Progress Brand Poster Before and After Prompt Enhancement

User Input: Create a sleek poster for Progress brand that features a modern electric car with silhouette and lighting effects using blue green and silver colors.

CLIPScore = 0.2544

Enhanced Prompt: poster for "Progress" with (futuristic & high-tech style), It should reflect electric sleek modern high digital theme and evoke bold energetic feeling. Incorporate A poster, car, brand, silhouette, lighting effects while using green, blue, silver The final design must be vector-based, scalable design.

CLIPScore = 0.4827

Figure 4.4: Toy Land Logo Before and After Prompt Enhancement

User Input: شعار لمتجر ألعاب الأطفال ، يحتوي على تصميم دب دمية مرح على طائرة، بألوان أساسية زاهية، وخط ممتع.

CLIPScore = 0.2504

Enhanced Prompt: Logo for "Toys Land" with (Cartoonish Style). It should reflect a flat design and evoke fun energetic feeling. Incorporate a teddy bear on a plane, bright primary colors, child-friendly branding, playful design, using red, blue, yellow, and soft illustrative palette. The final design must be vector-based, scalable design.

CLIPScore = 0.4777

Figure 4.5: Bake Haven Logo Before and After Prompt Enhancement

User Input A logo for a bakery shop features a bread and a chef hat on top of it using coffee, beige and white colors.

CLIPScore = 0.2683

Enhanced Prompt: Logo for "Bake Haven" bakery Shop with (Cozy & Warm style). It should reflect cozy theme and evoke timeless balanced feeling. Incorporate bakery shop Logo features bread, chef hat on top, using coffee, beige white color palette. The final design must be in vector-based, scalable format.

CLIPScore = 0.4521

- **Similarity Detection**

Figure 4.6: Similarity Detection Example

The System detects the similar image with similarity score equals **91.2%**

And the score increases as the common features between images increase.

- **Background Removal**

Figure 4.7: Cheese Shop Logo Before and After Background Removal

- **Image Editing**

Figure 4.8: John Wick Poster Before and After Image Editing

Edit prompt: make the poster more realistic, add neon lighting and rain.

Figure 4.9: Bird Image Before and After Image Editing

Edit prompt: change the bird color to green and the leaves color to red.

Chapter 5: Run The Application

In this section we'll see how **Wish a Pic** platform, helps you to make high-quality design accessible to everyone. Whether you're a small business owner, a student, or simply someone with a creative idea, Wish a Pic allows you to turn your words into stunning, brand-safe visuals with just a few simple steps.

The following sections include screenshots from our working system to illustrate the user interface, core features, and AI interactions in real time.

Let's begin exploring Wish a Pic in action:

- **Register & Login**
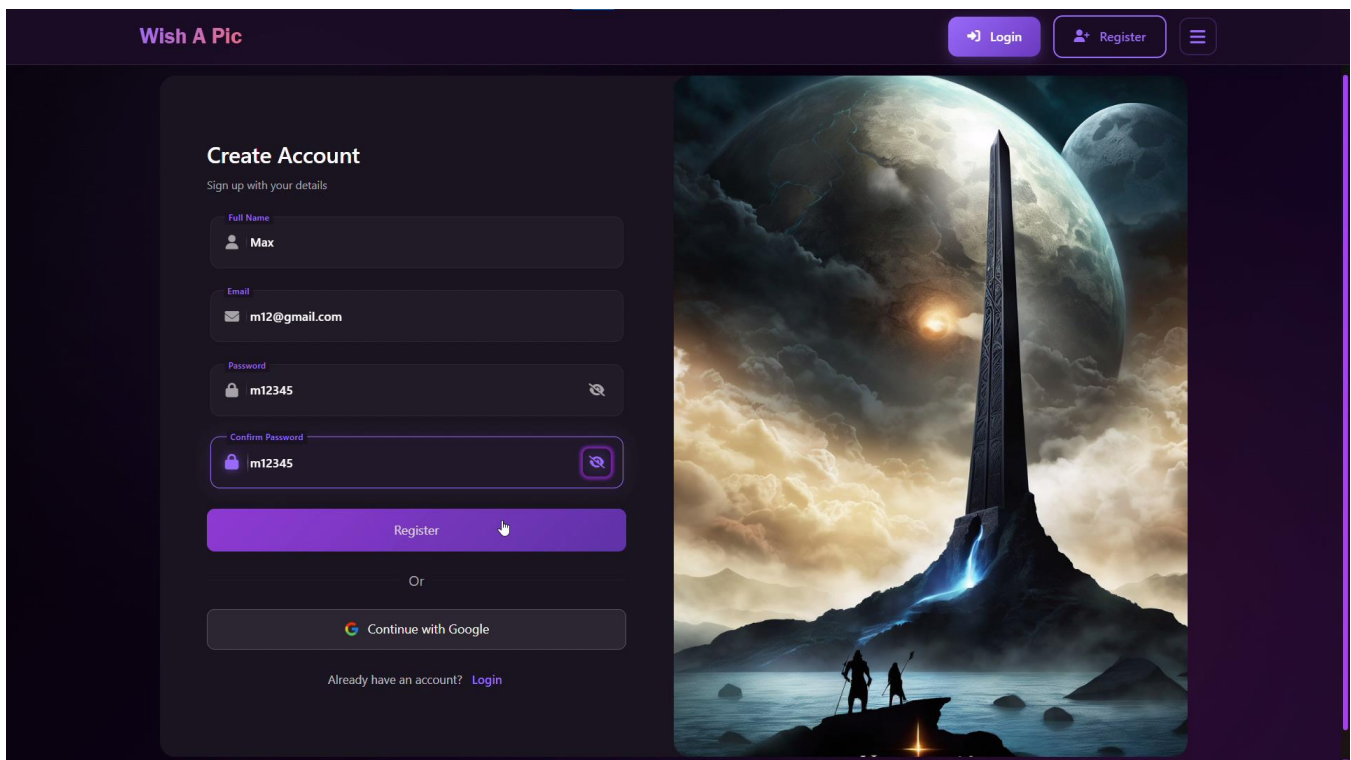
Figure 5.1: Register Page
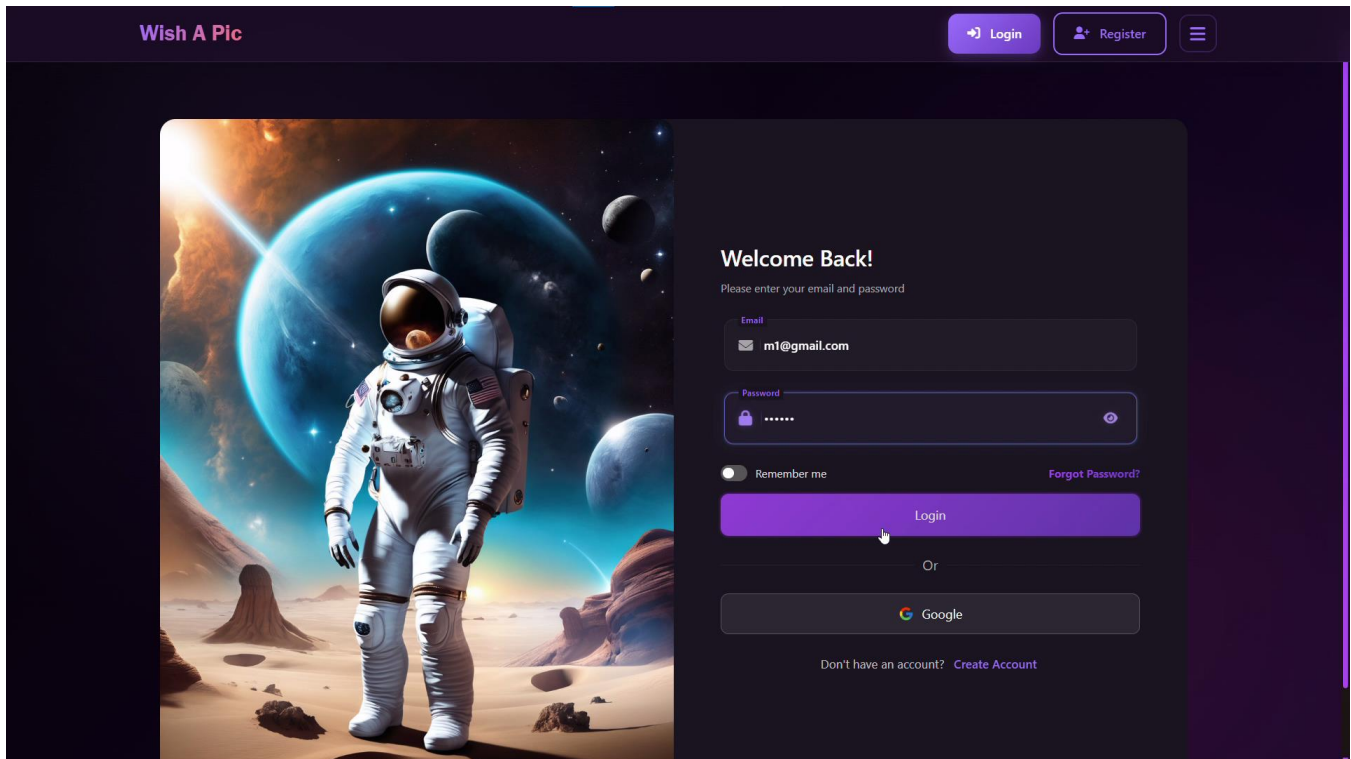
Figure 5.2: Login Page

To get started with Wish a Pic, users need to create an account by filling out a simple registration form. The steps are as follows:

1. Navigate to the Register page.

2. Enter your Full Name.

3. Provide a valid Email Address.

4. Create a Password that meets the platform's security criteria.

5. Confirm your password by re-entering it in the Password Confirmation field.

6. Click the Register button.

If all fields are valid and passwords match, the user account is created successfully, and the user is redirected to the login page.

Once registered, users can log in to their account:

1. Navigate to the Login page.
2. Enter your Email Address.
3. Enter your Password.
4. Click the Login button.

Upon successful authentication, the user is taken to the main interface, where they can begin generating visual content using prompts.

- **Home Page**
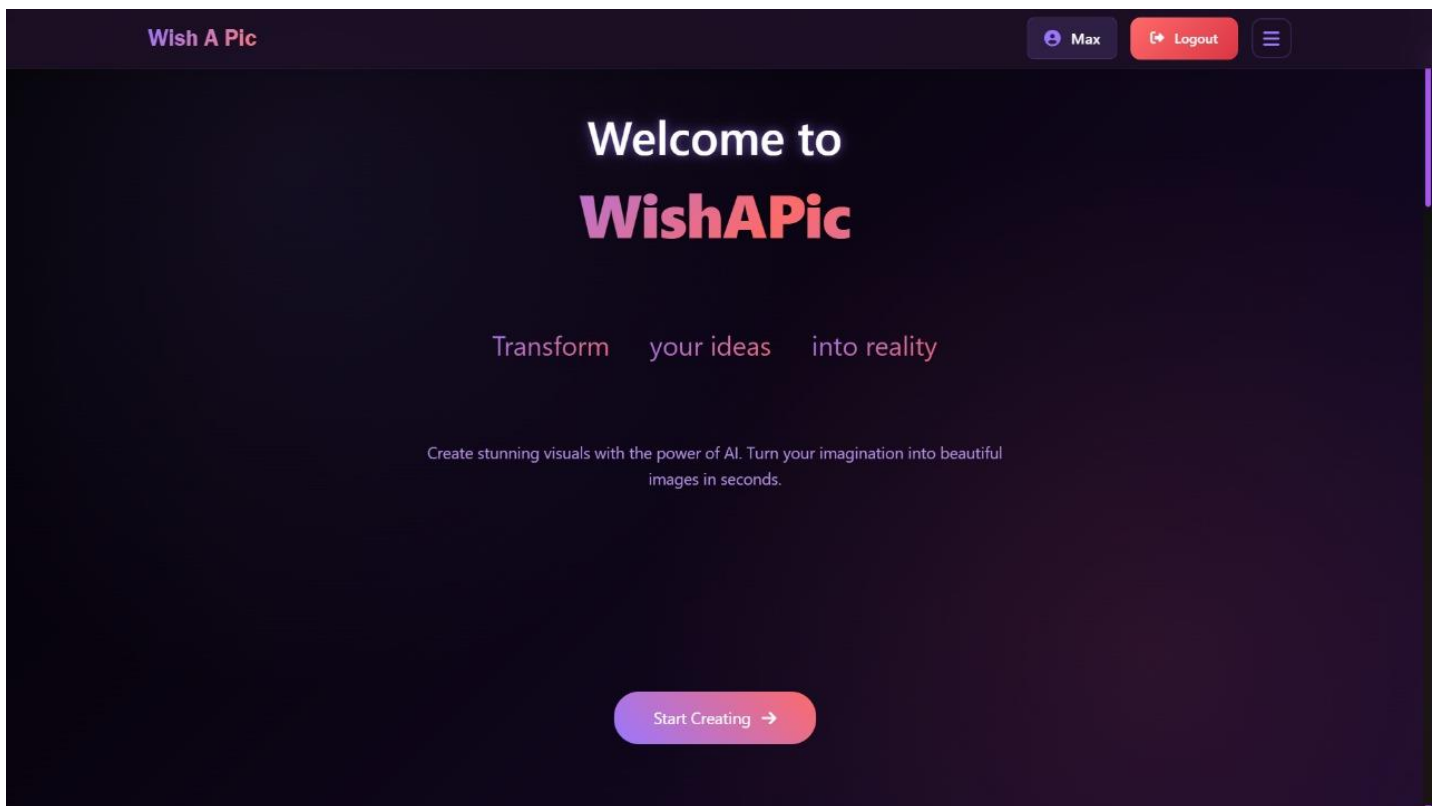
Figure 5.3: Home Page
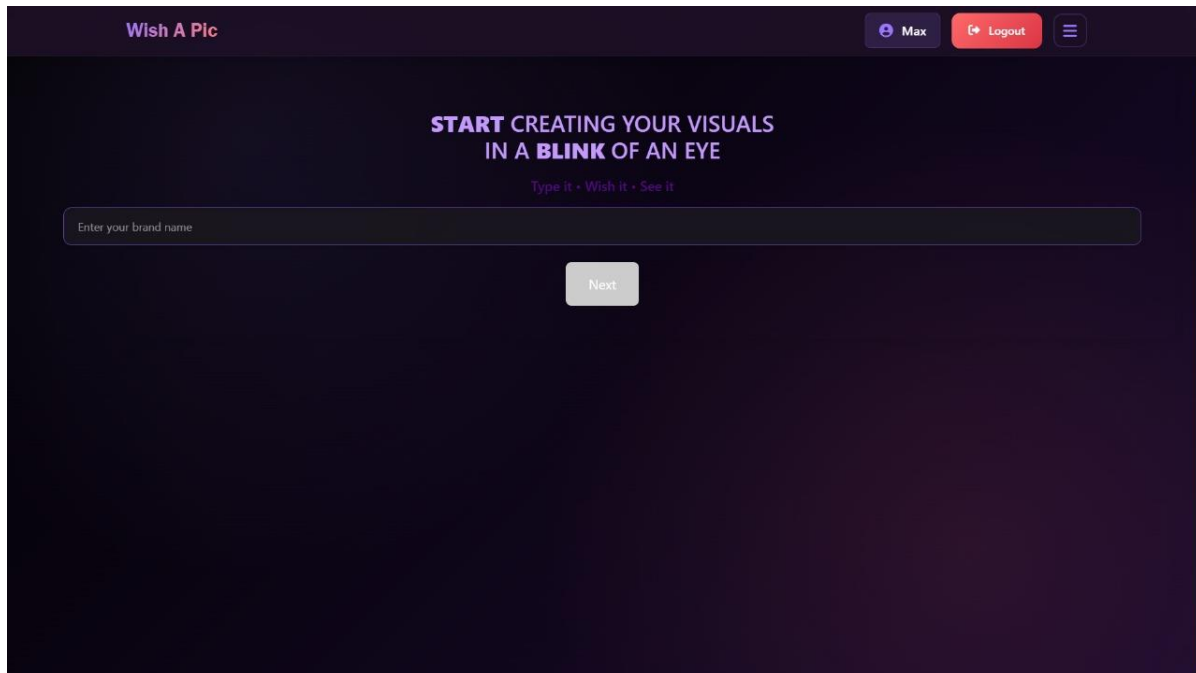
After logging in, users are directed to the Home Page, which serves as the central hub for accessing all main features of Wish a Pic. The layout is designed to be clean, intuitive, and accessible for users with or without design experience.

From the Home Page, users can:

1. Navigate to the Prompt Input Page to start generating visuals by entering text or voice prompts.

2. Open the Favorites section to revisit and manage previously saved designs.

3. Access the History page to explore their design activity and revisit past results.

4. Apply Similarity Detection for the generated designs.

5. Use the Image Editing and Background Removal features to refine or modify generated designs.

6. Log Out securely when finished using the platform.

The Home Page provides a smooth and efficient entry point into the creative process, giving users access to all essential features in one place.
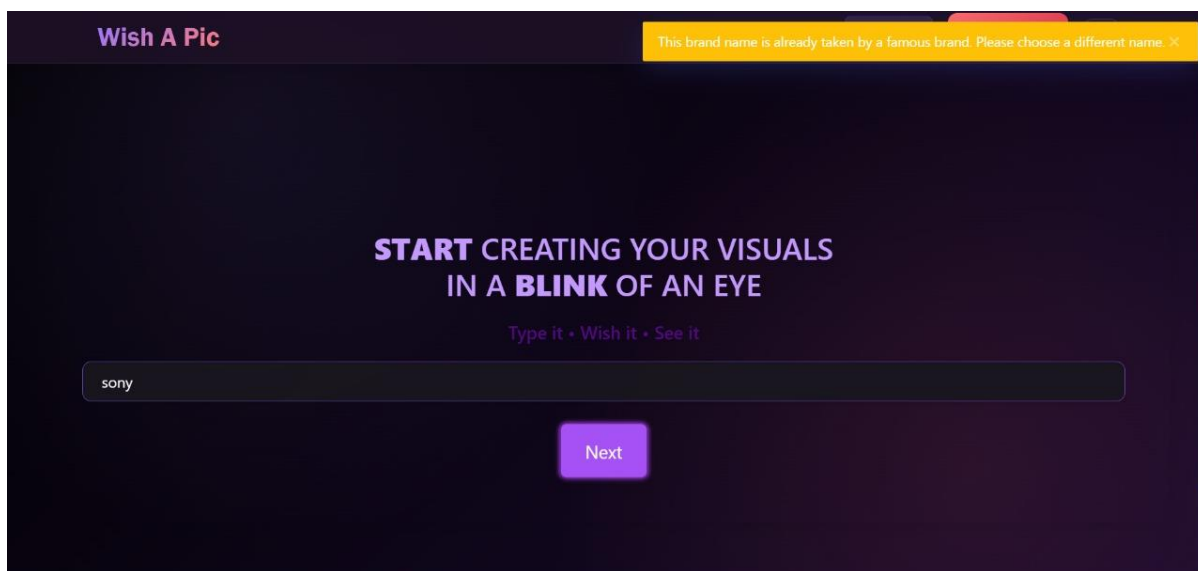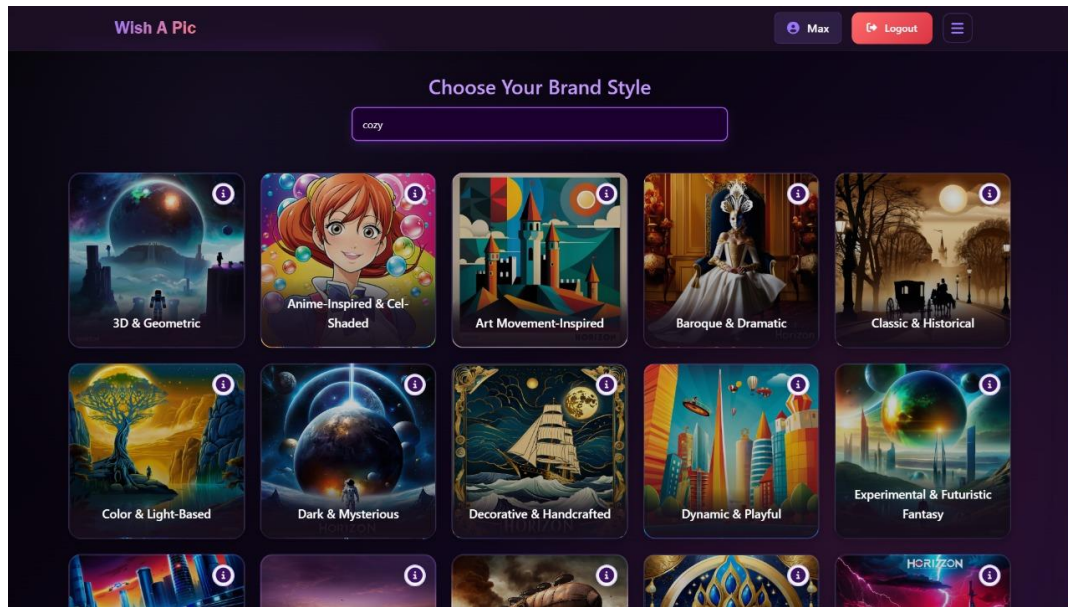
- **Image Generation**

Figure 5.4: Brand Name Input Page

The first step of the Image Generation process is that users enter their brand names then it got checked by the system to insure brand identity protection.


Figure 5.5: Brand Name Checking

Figure 5.6: Style Selection or Input

Second step is that users get to choose the style of their requested design each style contains set of images also generated by our system to help users understand the style vibe and tone, to insure flexibility and user satisfaction Wish a Pic supports the ability of entering styles based on user desire **even if it's not offered by the system**.
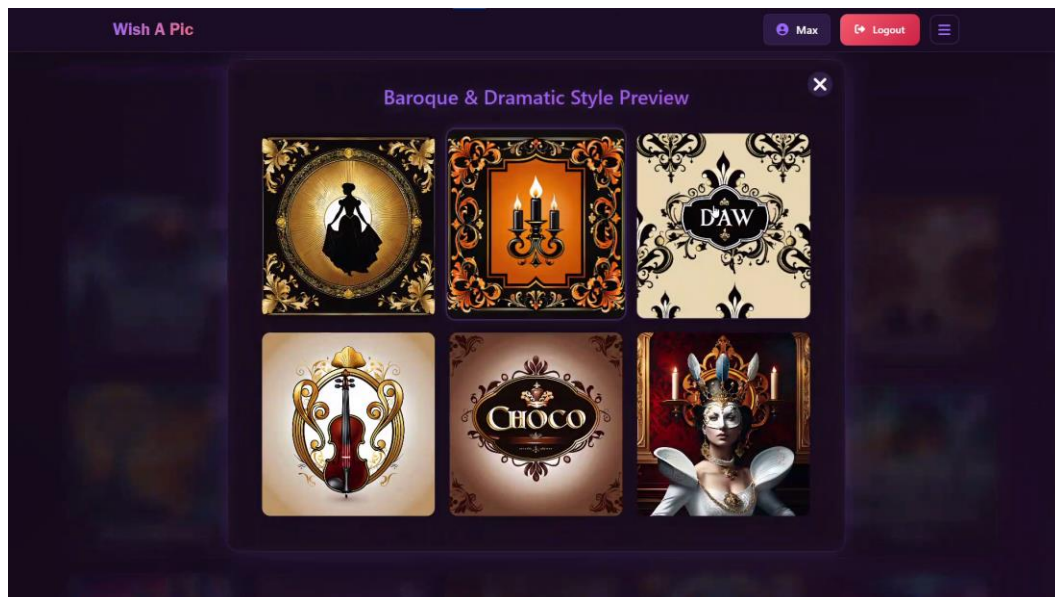
Figure 5.7: Style Example

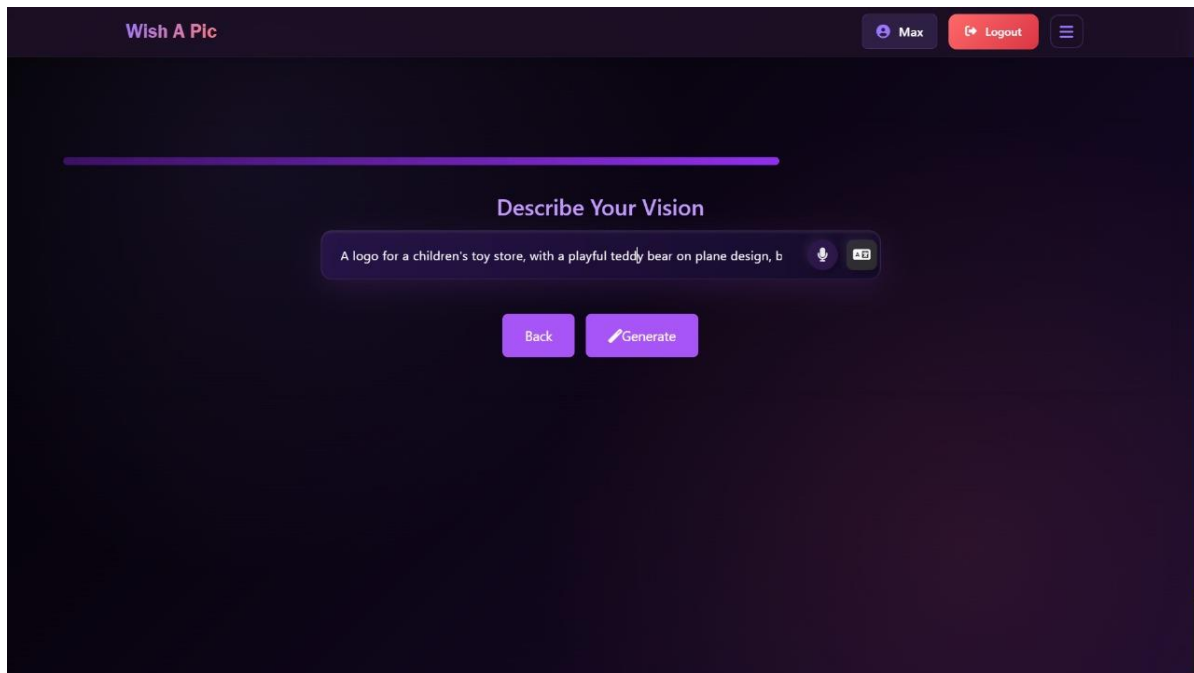Third step is to input requested design description.
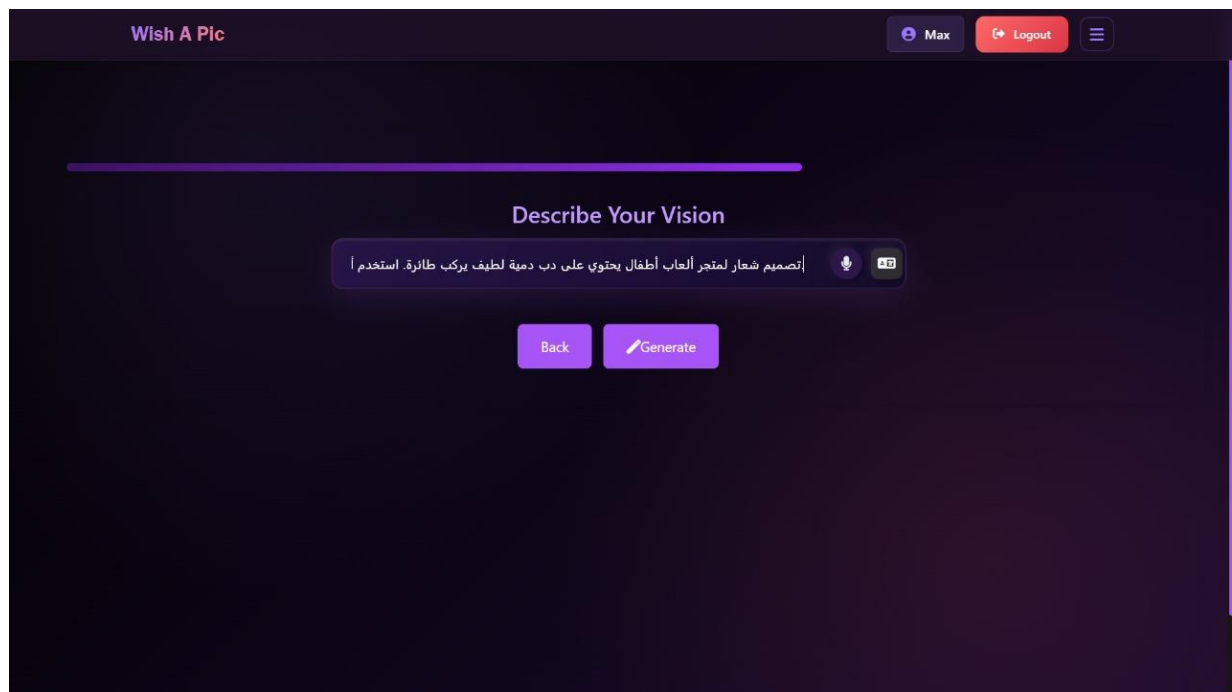
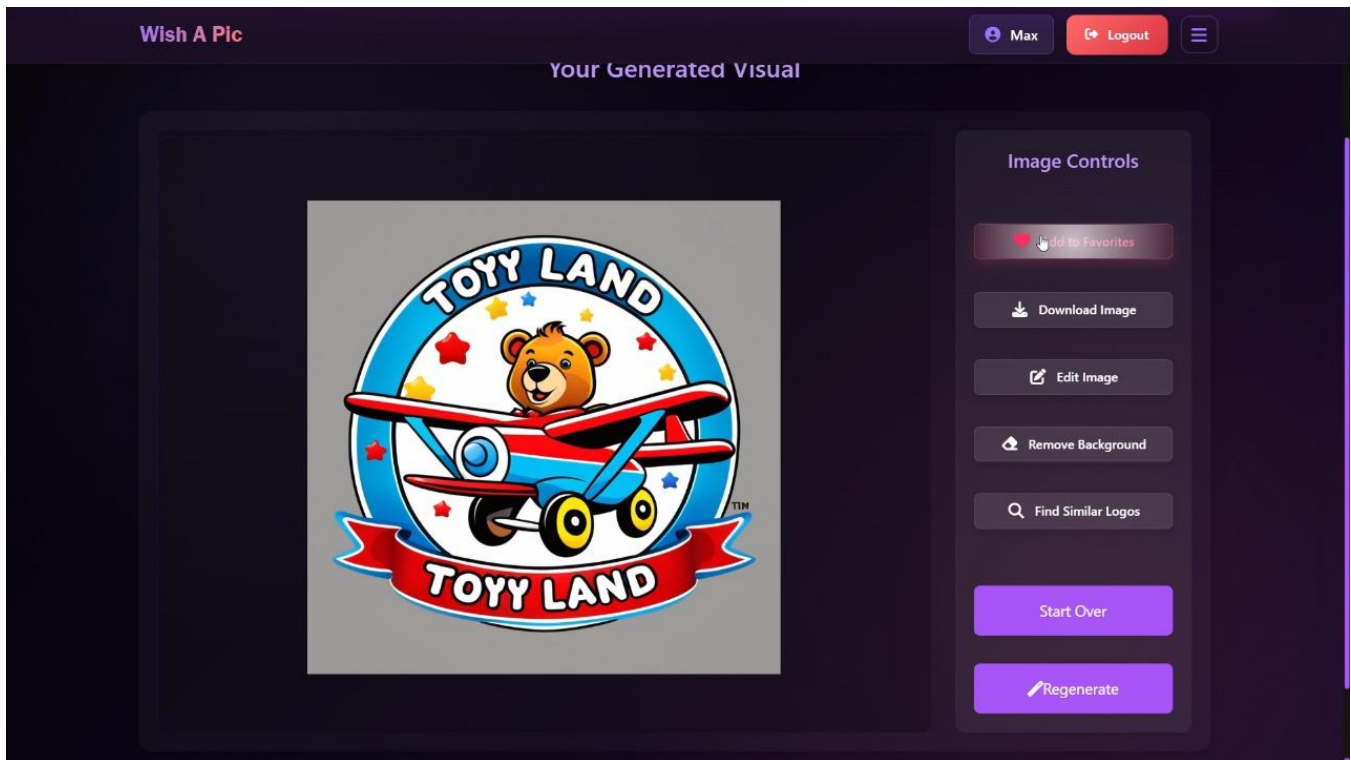

Figure 5.8: English Text Description Input

Figure 5.9: Arabic Text Description Input

Then after describing the requested design just click on the Generate button and wait for Wish a Pic to finish the job.

Figure 5.10: Output Logo From Text Description

Wish a Pic system supports English and Arabic text description input which applies more usability and flexibility to design process helping people from different cultures and backgrounds to use Wish a Pic easily without any complications or language obligation.

And to ease the process even more, Wish a Pic also supports Voice Input feature also for both English and Arabic which also empowers usability and makes users feel more comfortable and satisfied with the experience of our platform.
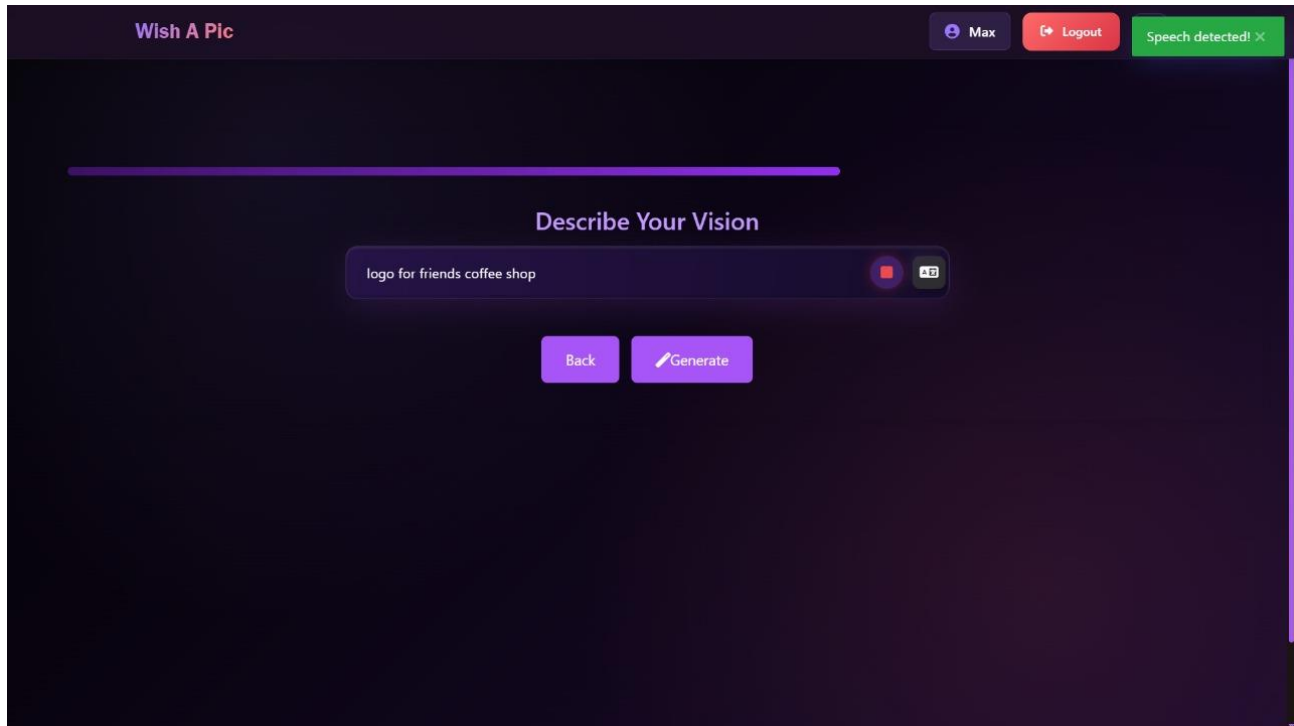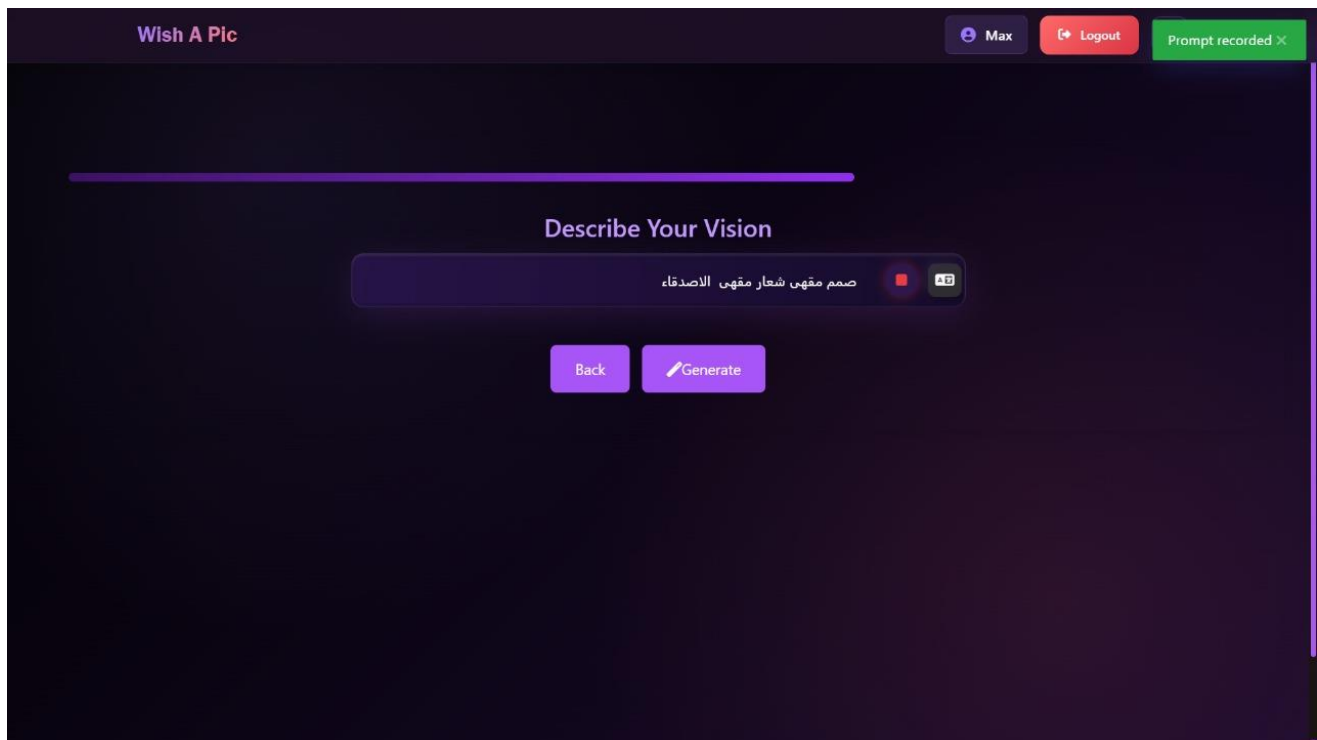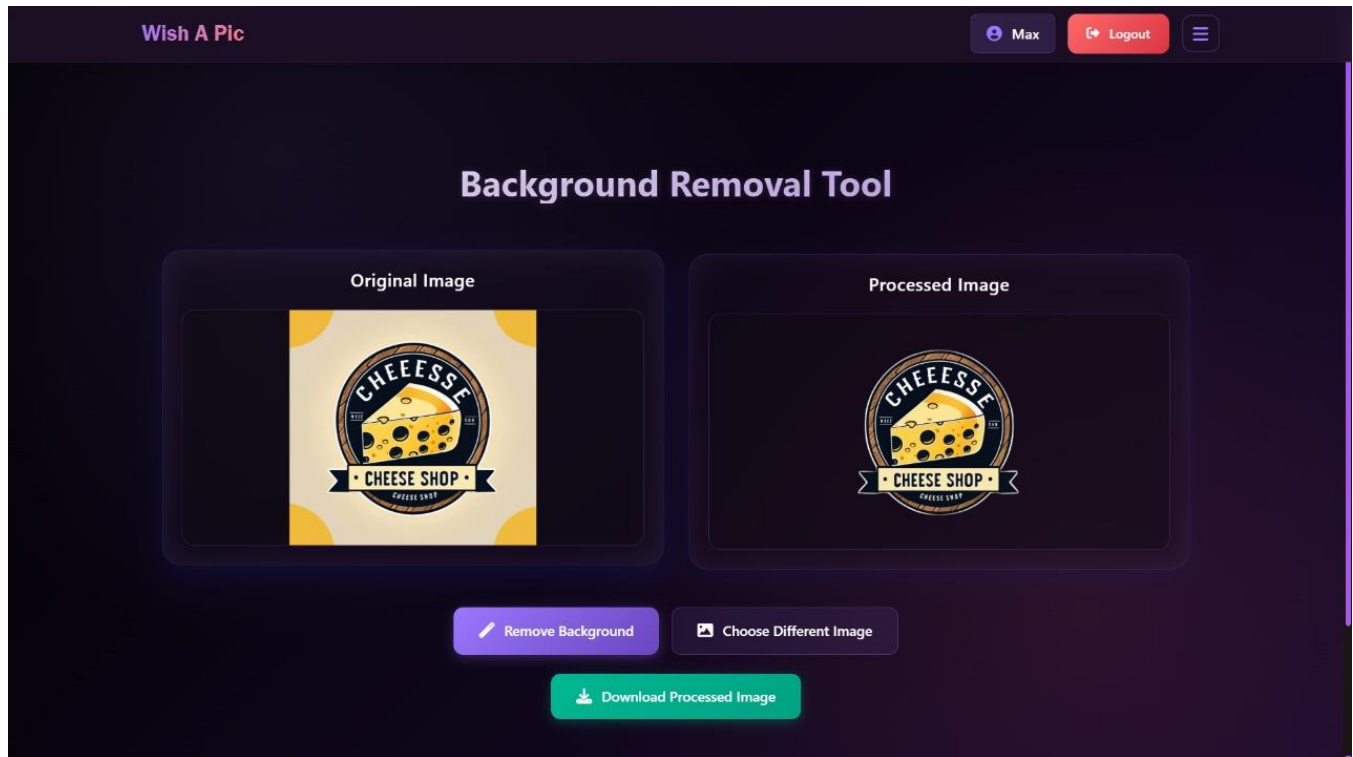
Figure 5.11: English Voice Input

Figure 5.12: Arabic Voice Input

- **Background Removal**

Figure 5.13: Background Removal Page

Wish a Pic provides the Background Removal services which helps to extract specific items from generated images such as logos, badges, icons or even characters and download it just by the following steps:

1. Click on the Remove Background button after generating an image so the generated image goes directly to the Background Removal Page.

2. You can also select any image from your device by clicking on the choose different image button.

3. Then click on the Remove Background button.

- **Similarity Detection**

Wish a Pic offers an optional Similarity Detection feature that empowers users to check whether their generated designs closely resemble any well-known brand logos.

After generating an image, the user can click the "Find Similar Logos" button to activate the similarity detection process.
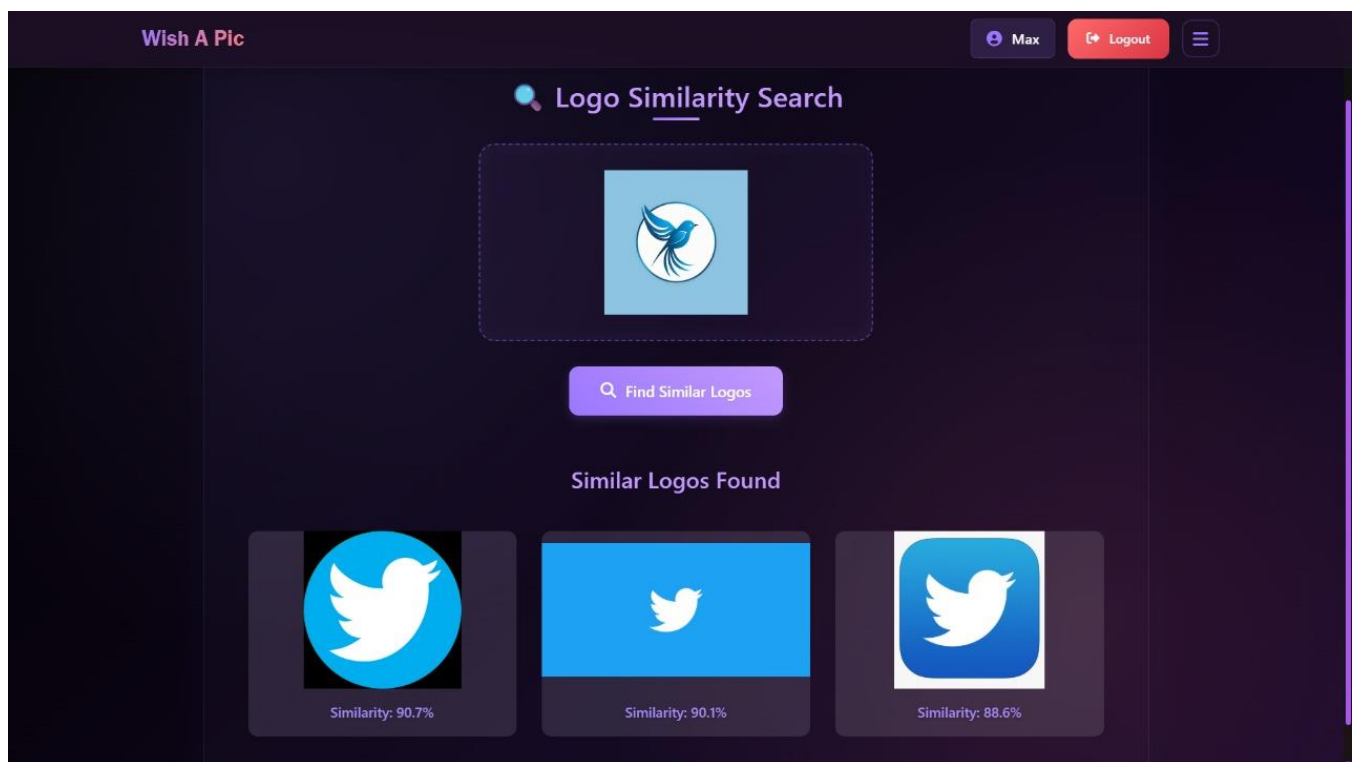
Figure 5.14: Similarity Detection Page

This tool is especially helpful for users who plan to use their visuals in commercial or branding contexts and want to avoid legal issues and accidental infringement.

- **Image Editing**

As another optional service Wish a Pic supports Image editing by simple steps:

1. Click on the Edit Image button after generating an image.
2. Enter description for the edit details.
3. Then click generate so Wish a Pic starts working.

After that Wish a Pic system starts editing the image based on the provided input.
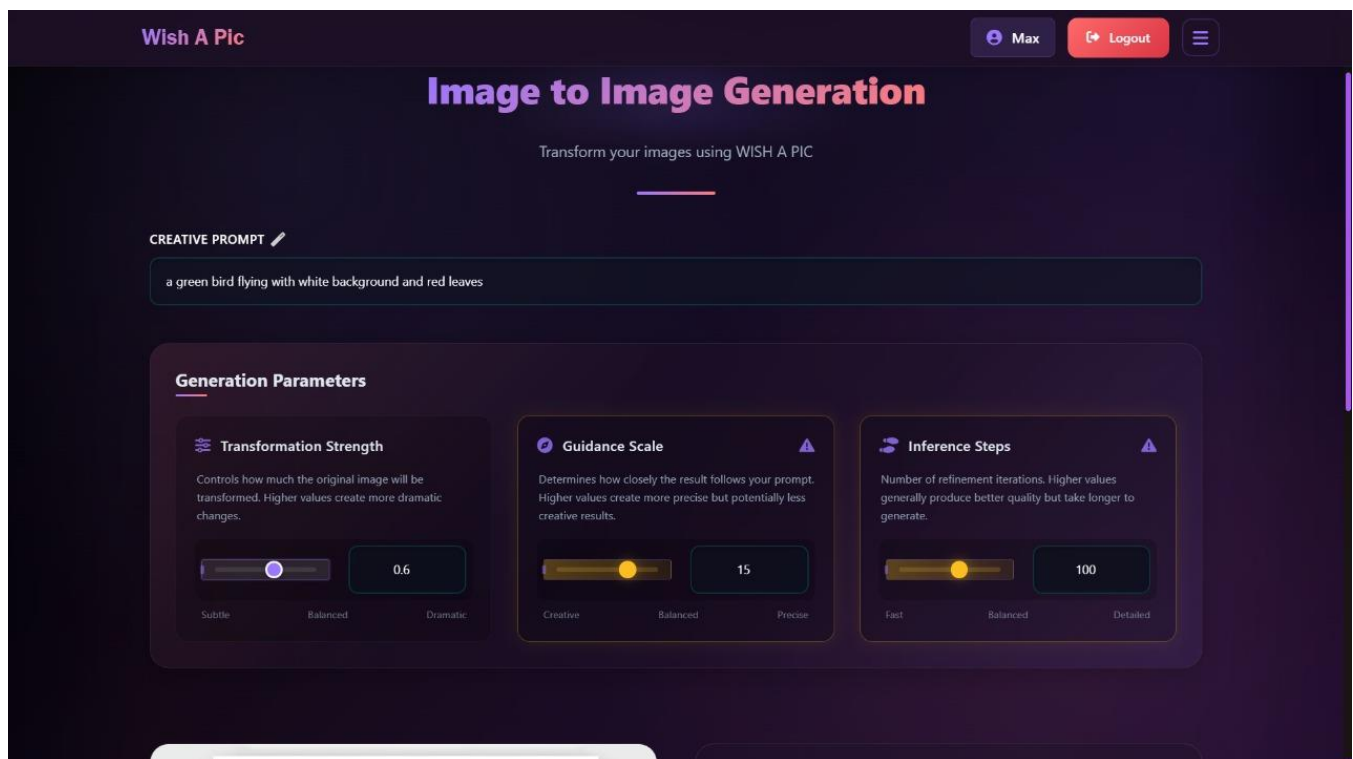
Figure 5.15: Edit Description Input
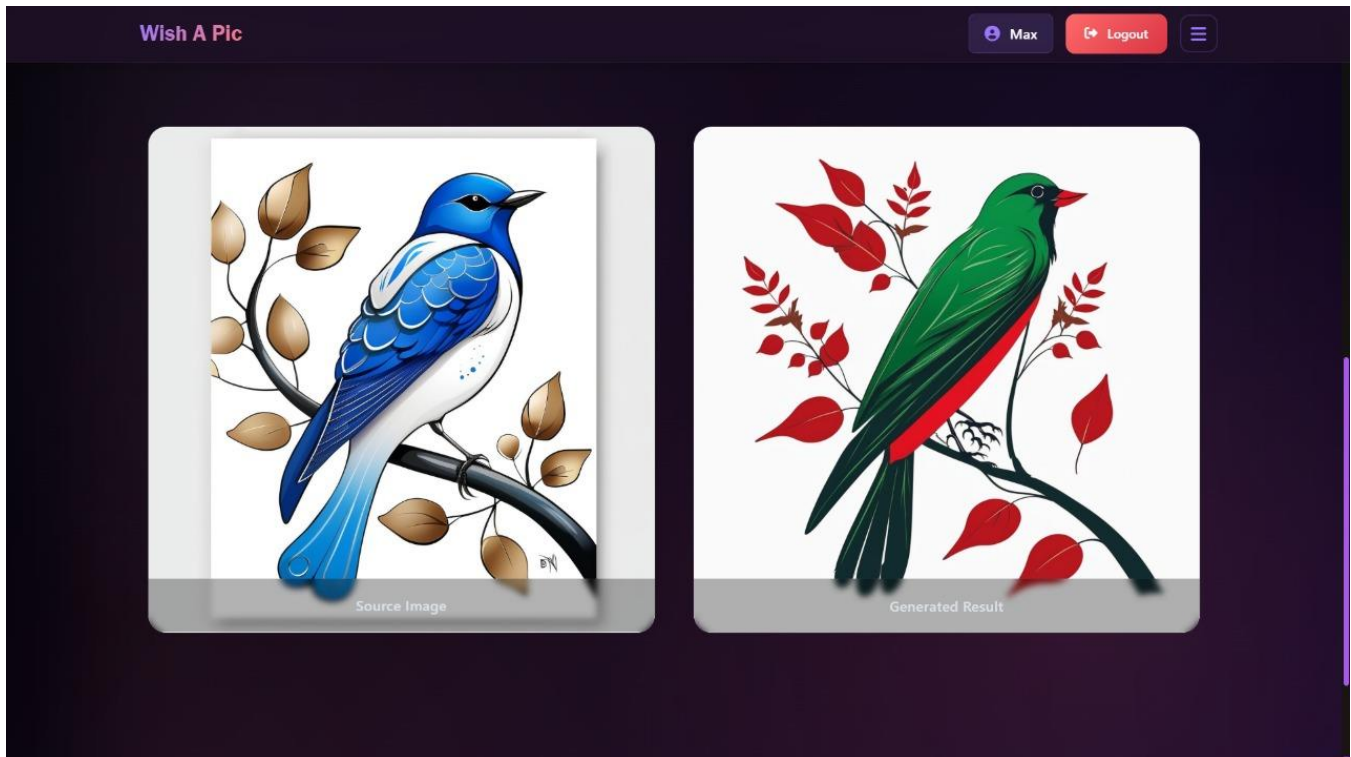
Figure 5.16: Edit Image output

• Favorites & History

To enhance the user experience, Wish a Pic includes two key features:

Favorites and History. These screens allow users to manage and revisit their previously generated visuals with ease.

The **Favorites screen** lets users save designs they find useful, visually appealing, or worth refining later. This allows for easy access to top picks without needing to regenerate or search through past outputs.

The **History screen** provides a chronological view of all images generated during a user's sessions. Each entry includes the original prompt and its corresponding result, helping users track their creative progress and revisit ideas for future development.

Figure 5.17: Favorites Page

Figure 5.18: History Page
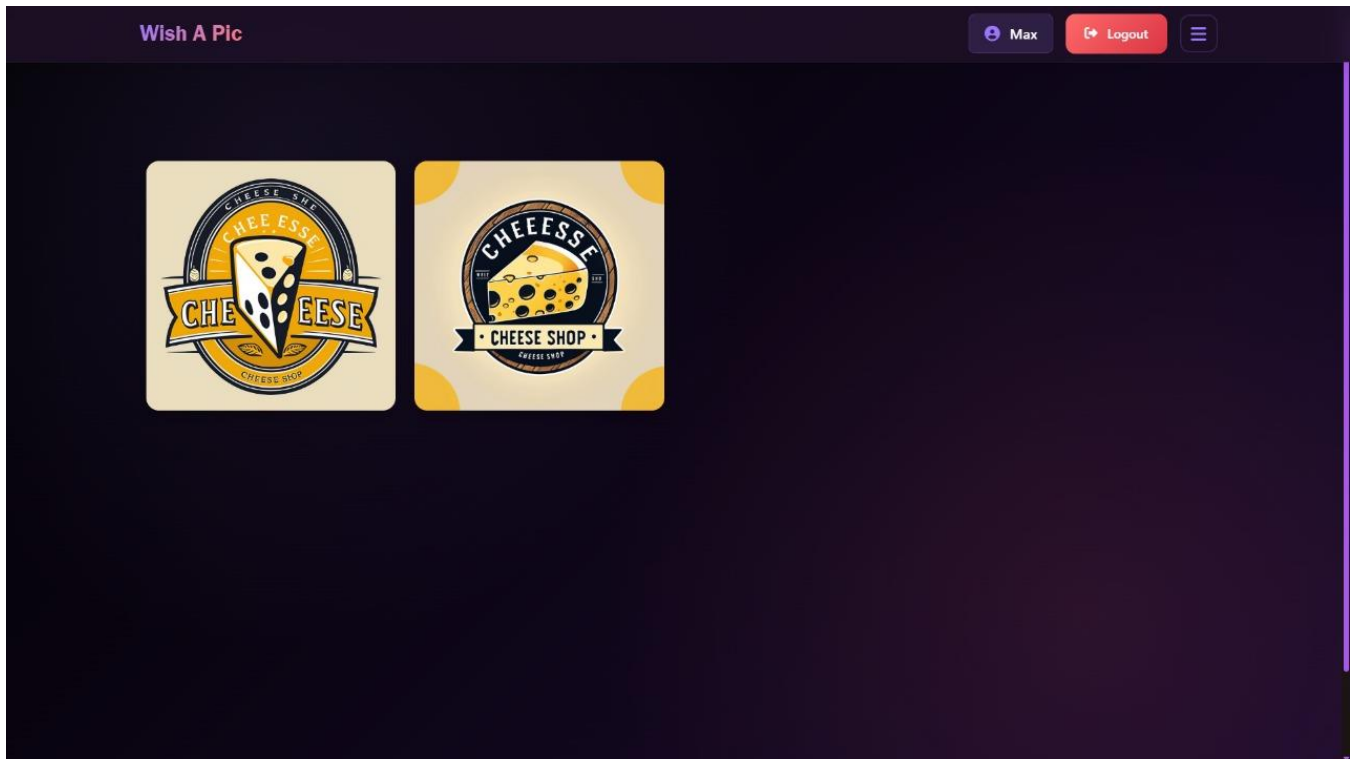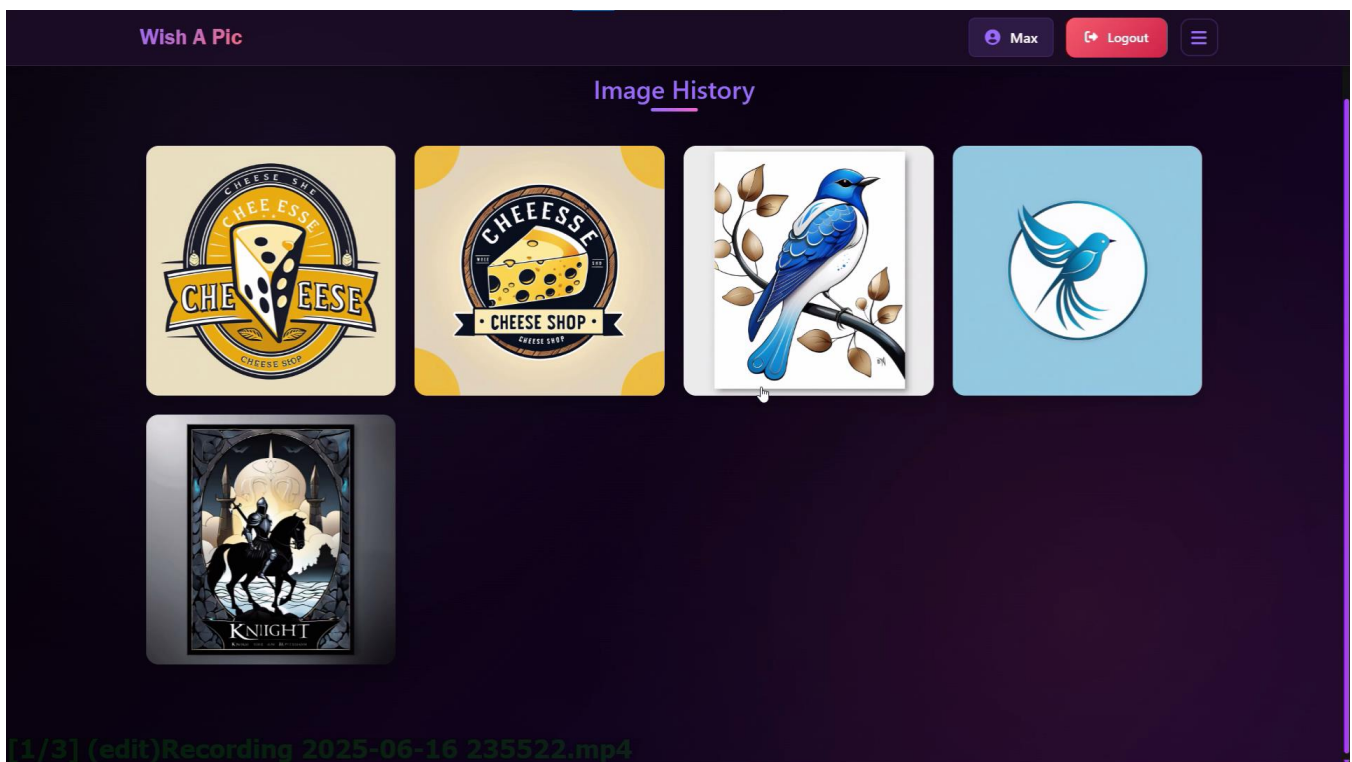
Chapter 6: Conclusion and Future Work

6.1 Conclusion

This project introduced **Wish a Pic**, an AI-powered platform designed to simplify visual content creation for users of all skill levels. By allowing users to generate logos, posters, and other branded visuals using simple text descriptions whether typed or spoken Wish a Pic bridges the gap between creative vision and technical execution. The platform supports both English and Arabic prompts, enhances input with smart prompt enhancement and preprocessing, produces professional-quality designs using Stable Diffusion XL Base 1.0, image editing to edit designs, background removal for extracting logos and icons, and maintains the brand identity and consistency using similarity detection.

One of the key contributions of this project is the focus on usability and accessibility. Unlike many existing tools that require either design knowledge or advanced prompt crafting, Wish a Pic empowers everyday users including small business owners, students, and hobbyists to create appealing visuals without cost or complexity. Features like prompt enhancement and preprocessing, similarity detection, content history, and favorites management make the platform more powerful, while maintaining a simple user experience.

6.2 Challenges Faced

- **Building a Seamless User Interface:** One of the main challenges was creating a user interface that is simple, clear, and responsive for users of all skill levels. The design had to work smoothly across various devices (PC, tablet, mobile) while still supporting All features like prompt input, image display, and content saving all without overwhelming new users.

- **Integrating AI Models:** Integrating multiple AI models such as Stable Diffusion XL Base 1.0 for image generation, NLP models for prompt enhancement, and CLIP for similarity detection required precise coordination between frontend and backend. Each model had different requirements and outputs, and ensuring they worked together in real time was technically demanding.

- **Arabic Description Support:** Supporting Arabic input wasn't as simple as translating it. The system needed to correctly detect Arabic prompts, translate them meaningfully into English without losing their intent, and then reformat them for the image generation model. Handling mixed-language prompts and Arabic structure was a key complexity.

- **Voice Input Handling:** Implementing voice input came with several challenges, including browser compatibility issues, microphone permission handling, and live speech recognition.

- **Balancing Accuracy with Creativity:** A major challenge was ensuring that the generated images matched user intent while still allowing for creative and visually impressive outputs. Too much structure limited creativity, while too much flexibility led to unrelated or messy designs. Fine-tuning this balance was a continuous process.

6.3 Future Work

- **User Feedback & Rating System:** Let users rate their generated designs or provide feedback on results. This data can help improve prompt processing and recommend better prompt formats to new users.

- **Input Restrictions and Prompt Validation:** the platform will introduce input restrictions that detect and block prompts containing inappropriate, harmful, or explicit language.

- **Smart Style Suggestions:** The platform can suggest design styles such as minimal, modern, vintage, or playful based on the user's

prompt. These suggestions will help guide non-expert users in shaping a more cohesive and professional-looking output, aligned with their intended tone or branding.

- **Enhancing Image Editing:** Currently, image generation is mostly automatic. In the future, adding interactive editing features (like adjusting elements, applying filters, or re-generating specific parts) will give users more control over the final visual without starting from scratch.

In conclusion, Wish a Pic successfully demonstrates how AI can make high-quality designs more accessible, efficient, and inclusive. With future enhancements, the platform has strong potential to become a go-to creative tool for users around the world.

References

[1] Pedroso, J. E., Sulleza, R. S., Francisco, K. H. M. C., Noman, A. J. O., & Martinez, C. A. V. (2023). Unlocking the power of Canva: Students' views on using the all-in-one tool for creativity and collaboration. Journal of Digital Learning and Distance Education, 2(2), 443-461.

[2] Marcus, G., Davis, E., & Aaronson, S. (2022). A very preliminary analysis of DALL-E 2. arXiv preprint arXiv:2204.13807.

[3] Reddy, M. D. M., Basha, M. S. M., Hari, M. M. C., & Penchalaiah, M. N. (2021). Dall-e: Creating images from text. UGC Care Group I Journal, 8(14), 71-75.

[4] Tamilkodi, R., Suryakala, K., Vamsi, N., Arvind Reddy, M., Nithin Kumar, M., & Venkat, M. (2024, February). Transforming Text into Art: Exploring DALL-E's Text-to-Image Generation Capabilities. In International Conference on Smart Data Intelligence (pp. 413-421). Singapore: Springer Nature Singapore.

[5] Ko, H. K., Park, G., Jeon, H., Jo, J., Kim, J., & Seo, J. (2023, March). Large-scale text-to-image generation models for visual artists' creative works. In Proceedings of the 28th international conference on intelligent user interfaces (pp. 919-933).

[6] Hanna, D. M. (2023). The use of artificial intelligence art generator "midjourney" in artistic and advertising creativity. Journal of design sciences and applied arts, 4(2), 42-58.

[7] Tsidylo, I. M., & Sena, C. E. (2023). Artificial intelligence as a methodological innovation in the training of future designers: Midjourney tools. Information Technologies and Learning Tools, 97(5), 203.

[8] Ruiz, N., Li, Y., Jampani, V., Pritch, Y., Rubinstein, M., & Aberman, K. (2023). Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 22500-22510).

[9] Baldridge, J., Bauer, J., Bhutani, M., Brichtova, N., Bunner, A., Castrejon, L., ... & Malik, C. (2024). Imagen 3. arXiv preprint arXiv:2408.07009.

[10] Tao, M., Tang, H., Wu, F., Jing, X. Y., Bao, B. K., & Xu, C. (2022). Df-gan: A simple and effective baseline for text-to-image synthesis. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16515-16525).

[11] Bahani, M., El Ouaazizi, A., & Maalmi, K. (2022). AraBERT and DF-GAN fusion for Arabic text-to-image generation. Array, 16, 100260.

[12] Podell, D., English, Z., Lacey, K., Blattmann, A., Dockhorn, T., Müller, J., ... & Rombach, R. (2023). Sdxl: Improving latent diffusion models for high-resolution image synthesis. arXiv preprint arXiv:2307.01952.

[13] Balaji, Y., Nah, S., Huang, X., Vahdat, A., Song, J., Zhang, Q., ... & Liu, M. Y. (2022). ediff-i: Text-to-image diffusion models with an ensemble of expert denoisers. arXiv preprint arXiv:2211.01324.

[14] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10684-10695).

[15] Salimans, T., & Ho, J. (2022). Progressive distillation for fast sampling of diffusion models. arXiv preprint arXiv:2202.00512.

[16] Esser, P., Kulal, S., Blattmann, A., Entezari, R., Müller, J., Saini, H., ... & Rombach, R. Scaling rectified flow transformers for high-resolution image synthesis, 2024. URL https://arxiv. org/abs/2403.03206, 2.

[17] Borji, A. (2022). Generated faces in the wild: Quantitative comparison of stable diffusion, midjourney and dall-e 2. arXiv preprint arXiv:2210.00586.

[18] Dehouche, N., & Dehouche, K. (2023). What's in a text-to-image prompt? The potential of stable diffusion in visual arts education. Heliyon, 9(6).

[19] Tan, K. L., Lee, C. P., Anbananthen, K. S. M., & Lim, K. M. (2022). RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network. IEEE Access, 10, 21517-21525.

[20] Wang, Y., Sun, Y., Fu, Y., Zhu, D., & Tian, Z. (2024). Spectrum-BERT: pre-training of deep bidirectional transformers for spectral classification of Chinese liquors. IEEE Transactions on Instrumentation and Measurement.

[21] Altinok, D. (2021). Mastering spaCy: An end-to-end practical guide to implementing NLP applications using the Python ecosystem. Packt Publishing Ltd.

[22] Loureiro, D., Barbieri, F., Neves, L., Anke, L. E., & Camacho-Collados, J. (2022). TimeLMs: Diachronic language models from Twitter. arXiv preprint arXiv:2202.03829.

[23] Borsos, Z., Sharifi, M., Vincent, D., Kharitonov, E., Zeghidour, N., & Tagliasacchi, M. (2023). Soundstorm: Efficient parallel audio generation. arXiv preprint arXiv:2305.09636.

[24] Li, J. (2022). Recent advances in end-to-end automatic speech recognition. APSIPA Transactions on Signal and Information Processing, 11(1).

[25] Jeon, J., Lee, S., & Choe, H. (2023). Beyond ChatGPT: A conceptual framework and systematic review of speech-recognition chatbots for language learning. Computers & Education, 206, 104898.

[26] Li, J., Tang, T., Zhao, W. X., Nie, J. Y., & Wen, J. R. (2024). Pre-trained language models for text generation: A survey. ACM Computing Surveys, 56(9), 1-39.

[27] Koonce, B. (2021). ResNet 50. In Convolutional neural networks with swift for tensorflow: image recognition and dataset categorization (pp. 63-72). Berkeley, CA: Apress.

[28] DeepSense.ai. (2023). Example of the Stable Diffusion inference process. DeepSense.ai.

[29] Bhat, A., & Jain, S. (2023). Face recognition in the age of clip & billion image datasets. arXiv preprint arXiv:2301.07315.

[30] Marvin, G., Hellen, N., Jjingo, D., & Nakatumba-Nabende, J. (2023, June). Prompt engineering in large language models. In International conference

on data intelligence and cognitive informatics (pp. 387-402). Singapore: Springer Nature Singapore.

[31] Ekin, S. (2023). Prompt engineering for ChatGPT: a quick guide to techniques, tips, and best practices. Authorea Preprints.