AJEENKYA D Y PATIL UNIVERSITY | THE INNOVATION UNIVERSITY

A
MINI PROJECT REPORT ON

"AI Ghibli Art Trends: A Social Media Data Analysis"

FOR

Term Work Examination

Bachelor of Computer Application in Artificial Intelligence & Machine Learning (BCA - AIML)
Year 2025-2026

Ajeenkya DY Patil University, Pune

-Submitted By
Ms. Divya Sutar

Under the guidance of
Prof. Vivek More

Ajeenkya DY Patil University

D Y Patil Knowledge City,
Charholi Bk. Via Lohegaon,
Pune - 412105
Maharashtra (India)

Date: 14 / 04 / 2025

CERTIFICATE

This is to certified that_Divya sutar_____
A student of **BCA(AIML) Sem-IV** URN No 2023-B-08022004A
has Successfully Completed the Dashboard Report On

"AI Ghibli Art Trends: A Social Media Data Analysis"

As per the requirement of
Ajeenkya DY Patil University, Pune was carried out under my supervision.
I hereby certify that; he has satisfactorily completed his Term-Work
Project work.

Place: Pune

Introduction :

"AI Ghibli Trend Dataset v2", explores the growing trend of AI-generated art inspired by the unique visual style of *Studio Ghibli*. With 500 records, it captures a snapshot of how users create, share, and engage with Ghibli-style images using AI tools across social platforms.

Each entry includes:

- **Prompts** used to generate the artwork (e.g., *"Magical Ghibli forest with floating lanterns"*)
- **Engagement metrics**: likes, shares, and comments
- **Platform** data (e.g., Reddit, TikTok, Instagram)
- **Technical aspects**: generation time, GPU usage, image resolution, and file size
- **Quality & authenticity**: a style accuracy score, whether the image was hand-edited, and if ethical concerns were flagged
- **Social feedback**: top user comments

This dataset is ideal for analyzing:

- Popular themes and prompt patterns
- User engagement across platforms
- Technical performance vs. visual quality
- Ethical issues in AI-generated content
- The cultural impact of AI in creative domains

Objective:

1. Trend Analysis

   Identify which types of **prompts** and **themes** (e.g., "enchanted ruins," "Ghibli-style sunset") are most popular.

   Observe how **user preferences** evolve over time and across social media platforms.

2. Engagement Insights

   Understand how different kinds of AI-generated Ghibli content perform in terms of:

   - Likes
   - Shares
   - Comments

   Compare engagement **across platforms** like Reddit, TikTok, and Instagram.

3. Technical Performance Evaluation

   Analyze the **efficiency** of AI models based on:

   - Image generation time
   - GPU usage
   - File size and resolution

   Correlate these factors with **visual quality** (style accuracy score) and **engagement levels**.

4. Human Involvement and Authenticity

   Explore the impact of **hand-editing** on engagement and perception.

   Assess how authenticity (AI vs. hand-edited) influences user reactions.

5. Ethical and Social Considerations

   Track instances where **ethical concerns** are flagged (e.g., potential copyright infringement, misleading content).

   Examine public sentiment through **top user comments** to gauge reception and concerns.

Methodology & Approach:

1. Data Collection

Source Platforms: Data was gathered from popular social media platforms such as **Reddit**, **Instagram**, and **TikTok**, where users frequently share AI-generated artwork.

Selection Criteria: Posts were included based on the presence of Ghibli-style prompts or tags, high engagement, or being explicitly labeled as AI-generated.

Time Frame: The dataset includes images created between **early 2025 and March 2025** (based on `creation_date` values).

2. Data Generation & Labeling

Prompt Extraction: Each image is associated with a **text prompt** that was used to generate the artwork (e.g., "Studio Ghibli-inspired ocean with giant fish").

Technical Metrics: Data such as **generation time**, **GPU usage**, **resolution**, and **file size** were recorded from the AI tools used during the image creation process.

Visual Quality Assessment: A **style accuracy score** (0–100) was assigned, reflecting how closely the image resembles authentic Ghibli aesthetics—this may have been assessed using a trained model or human raters.

Manual Review Tags:

is_hand_edited: Whether the image was post-processed by a human.

ethical_concerns_flag: Whether the content raised potential ethical issues (e.g., artistic plagiarism, misleading realism).

3. Engagement Analysis

User Interaction Metrics: The dataset includes:

- `likes`
- `shares`
- `comments`

Top Comment: Extracted from each post to help analyze viewer sentiment and reactions.

4. Data Structuring

All entries are organized into a structured CSV format with 16 columns.

Each row represents a **unique AI-generated image**, identified by a `image_id` and associated with a `user_id`.

5. Analytical Approach (Suggested Use)

Researchers or analysts may apply:

- **Descriptive statistics** to evaluate distribution of style scores, engagement, or GPU usage.
- **Trend analysis** to identify popular prompts or rising visual themes.
- **Correlation analysis** to explore relationships between technical aspects and engagement.
- **Sentiment analysis** on top comments for public opinion insights.
- **Ethical review metrics** to study frequency and context of flagged content.

Implementation & Code:

Step 1 : import pandas

theory :

Loads Pandas to work with datasets (like reading CSVs, filtering data, etc.).
Imports Matplotlib for creating visualizations like line plots, bar charts, etc.
Imports Seaborn, a library for making beautiful and easy statistical plots (like box plots, heatmaps, etc.).

code :

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Step 2 : Data Loading and Exploratory Data Analysis (EDA)

theory :

Purpose: Loads the CSV file into a **Pandas DataFrame** called `df`

This is the **data loading step** where you import your dataset to start analysis.

Shows **basic structure and metadata** of the DataFrame:

- Number of rows and columns
- Column names and data types
- Number of non-null (non-missing) entries

Helps identify missing values or data types that need conversion.

Displays the **first 5 rows** of the dataset.

Gives a quick look at what kind of data you're working with (columns and values).

Shows **summary statistics** (count, mean, std, min, max, quartiles) for all numerical columns.

Helps understand **data distribution**, **central tendency**, and **spread** of each variable.

code :

```
# Load the dataset
df = pd.read_csv("ai_ghibli_trend_dataset_v2.csv")

# Display basic info
print("Basic Info:")
print(df.info())
print("\nSample Data:")
print(df.head())

# Summary statistics
print("\nSummary Statistics:")
print(df.describe())
```

output :

```
Basic Info:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 500 entries, 0 to 499
Data columns (total 16 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   image_id              500 non-null    object
 1   user_id               500 non-null    object
 2   prompt                500 non-null    object
 3   likes                 500 non-null    int64
 4   shares                500 non-null    int64
 5   comments              500 non-null    int64
 6   platform              500 non-null    object
 7   generation_time       500 non-null    float64
 8   gpu_usage             500 non-null    int64
 9   file_size_kb          500 non-null    int64
 10  resolution            500 non-null    object
 11  style_accuracy_score  500 non-null    int64
 12  is_hand_edited        500 non-null    object
 13  ethical_concerns_flag 500 non-null    object
 14  creation_date         500 non-null    object
 15  top_comment           500 non-null    object
dtypes: float64(1), int64(6), object(9)
memory usage: 62.6+ KB
None

Sample Data:
                              image_id   user_id  \
0  77ce5c72-eb45-4651-bcb1-c0677c0fceaf  6a7adf3d
1  7d66c67f-0d11-4ef9-895c-d865ef11fe40  523b8706
2  d7978afd-3932-4cce-9a21-5f9bf2bc1f64  0e02592a
3  cb34636a-a15c-4b15-999c-759dbb8896fe  9ed78a42
4  7511fbb8-db05-4584-a3a4-e8bb525ed58b  69ec8f02

                                     prompt  likes  shares  comments  \
0  Studio Ghibli-inspired ocean with giant fish    916     410       555
1               Ghibli-style village at sunset   2965    1361       417
2   A lone traveler exploring an enchanted ruin   4727     655       785
3    Spirited Away-style bustling market street   1629    1954       212
4  Magical Ghibli forest with floating lanterns   2573    1281       913

    platform  generation_time  gpu_usage  file_size_kb resolution  \
0     Reddit             4.80         49          1684  1024x1024
1     Reddit            11.11         81          2808  1024x1024
2  Instagram             5.56         41          1800  2048x2048
3     TikTok            12.45         88           479  2048x2048
4     TikTok             4.80         64          1789    512x512

   style_accuracy_score is_hand_edited ethical_concerns_flag creation_date  \
0                    89            Yes                   Yes    2025-03-11
1                    92            Yes                    No    2025-03-11
2                    61             No                    No    2025-03-06
3                    76             No                    No    2025-03-23
4                    58             No                   Yes    2025-03-06

                                    top_comment
0  So nostalgic, feels like childhood memories. 🎥...
1     Absolutely stunning! Love the details. 🎨 #5729
2      Is this AI or hand-painted? Incredible! #8001
3      Is this AI or hand-painted? Incredible! #5620
4  This looks straight out of a Ghibli movie! 🌟 #...

Summary Statistics:
             likes       shares    comments  generation_time   gpu_usage  \
count   500.000000   500.000000  500.000000       500.000000  500.000000
mean   2601.262000  1040.182000  506.872000         8.317780   61.124000
std    1429.433498   562.668738  283.384066         3.903103   18.151131
min     105.000000    13.000000    5.000000         1.540000   30.000000
25%    1343.500000   587.750000  276.750000         5.027500   45.000000
50%    2566.500000  1092.000000  518.000000         8.380000   63.000000
75%    3913.250000  1502.000000  744.250000        11.540000   77.000000
max    4944.000000  1999.000000  998.000000        14.990000   90.000000
```

```
       file_size_kb  style_accuracy_score
count    500.000000            500.000000
mean    2511.822000             74.626000
std     1390.178578             14.679001
min      101.000000             50.000000
25%     1374.750000             62.000000
50%     2498.000000             74.000000
75%     3729.000000             87.250000
max     4973.000000            100.000000
```

Results & Visualization:

Step 1 : Data Transformation and Visualization

theory :

1. Date Conversion
   - Converts the `'creation_date'` column to **datetime format**.
   - Useful for **time-based analysis**, sorting, or filtering.

2. Extract Top 5 Most Liked Prompts
   - Sorts data by the 'likes' column in descending order.
   - Retrieves the top 5 prompts that received the most likes
   - Helps identify high-performing content.

3. Box Plot: Likes by Platform

Creates a **box plot** to show the **distribution of likes** for each **platform**.
Helps visualize:

- Median likes
- Variability (spread)
- Outliers
- Comparison between platforms

Adds a clear title to the plot for better understanding.

code :

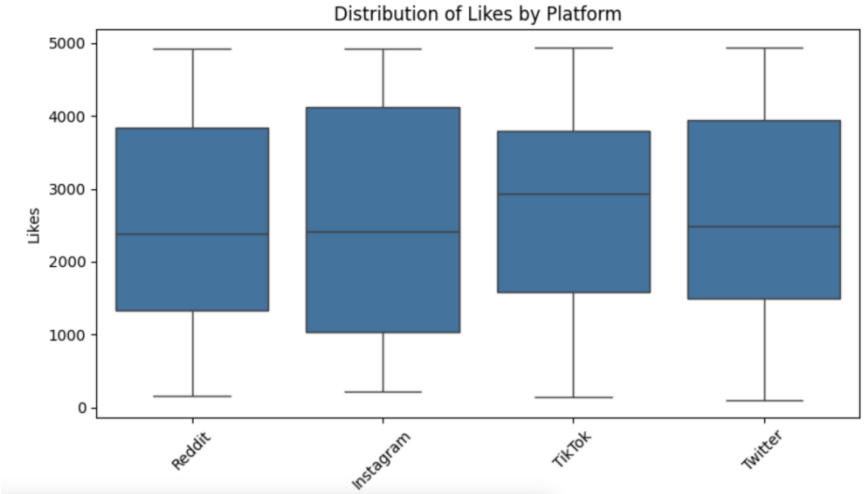
```
# Convert creation_date to datetime
df['creation_date'] = pd.to_datetime(df['creation_date'])

# Top 5 most liked images
top_liked = df.sort_values(by="likes", ascending=False).head(5)
print("\nTop 5 Most Liked Prompts:")
print(top_liked[['prompt', 'likes', 'platform']])

# Plot 1: Likes by Platform
plt.figure(figsize=(8, 5))
sns.boxplot(data=df, x='platform', y='likes')
plt.title("Distribution of Likes by Platform")
plt.ylabel("Likes")
plt.xlabel("Platform")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

output :

```
Top 5 Most Liked Prompts:
                                              prompt  likes   platform
485        Ghibli-style mountain with floating islands   4944     TikTok
155        A lone traveler exploring an enchanted ruin   4940    Twitter
234  Anime-style train passing through a fantasy world   4931    Twitter
418  Anime-style train passing through a fantasy world   4930  Instagram
215           Ghibli-style night sky with glowing stars   4928     Reddit
```



Step 2 : Style Accuracy Score Distribution

theory :

❖ Histogram with KDE (Kernel Density Estimation)

- df['style_accuracy_score']: Uses the column representing how well the image matches a specific artistic style (e.g., Studio Ghibli).
- bins=20: Divides the score range into 20 intervals (bars).
- kde=True: Adds a smooth curve showing the probability distribution of the scores.
- color="skyblue": Sets the color of the histogram bars.

❖ Plot Labels and Layout

- Adds a **title** and axis labels for clarity.
- `tight_layout()` makes sure labels/titles don't get cut off.
- `show()` displays the final plot.

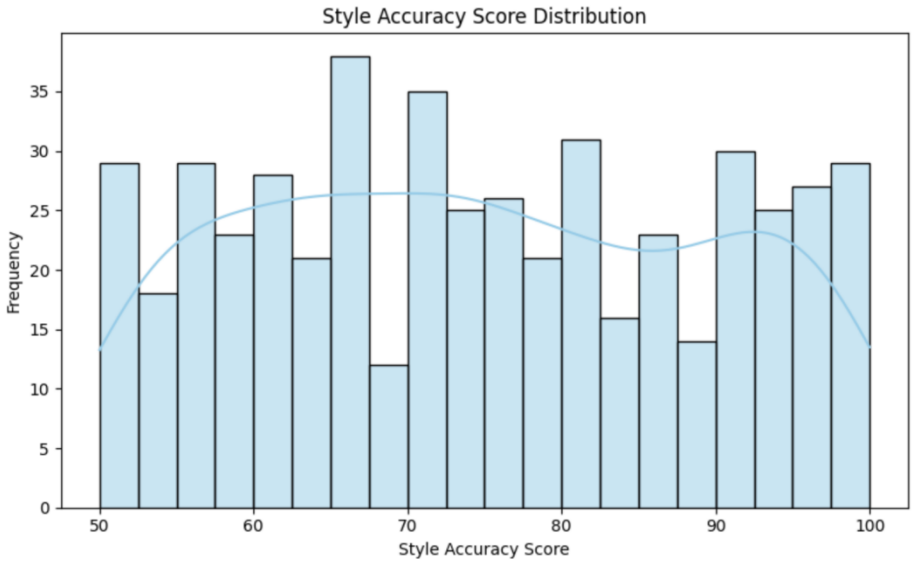
❖ Purpose of This Plot:

- Helps **understand the overall quality** of the generated images in terms of matching style.
- Identifies if most images score high, low, or somewhere in between.
- Useful for **quality assessment** or model improvement.

code :

```
# Plot 2: Style Accuracy Score Distribution
plt.figure(figsize=(8, 5))
sns.histplot(df['style_accuracy_score'], bins=20, kde=True, color="skyblue")
plt.title("Style Accuracy Score Distribution")
plt.xlabel("Style Accuracy Score")
plt.ylabel("Frequency")
plt.tight_layout()
plt.show()
```

output :



Step 3 : Daily Trend of AI Ghibli Posts

theory :

❖ Step-by-Step Breakdown

- value_counts(): Counts how many posts were created on each date.
- sort_index(): Sorts the dates in chronological order (since value_counts() returns them unordered)
- The result: a time series of post counts per day.

❖ Plotting the Bar Chart

This plots a **bar chart** with:

- **X-axis** = Dates
- **Y-axis** = Number of posts on that data

❖ Final Formatting

- Adds a title and axis labels for clarity.
- Rotates date labels for better readability.
- Ensures proper layout with `tight_layout()`.

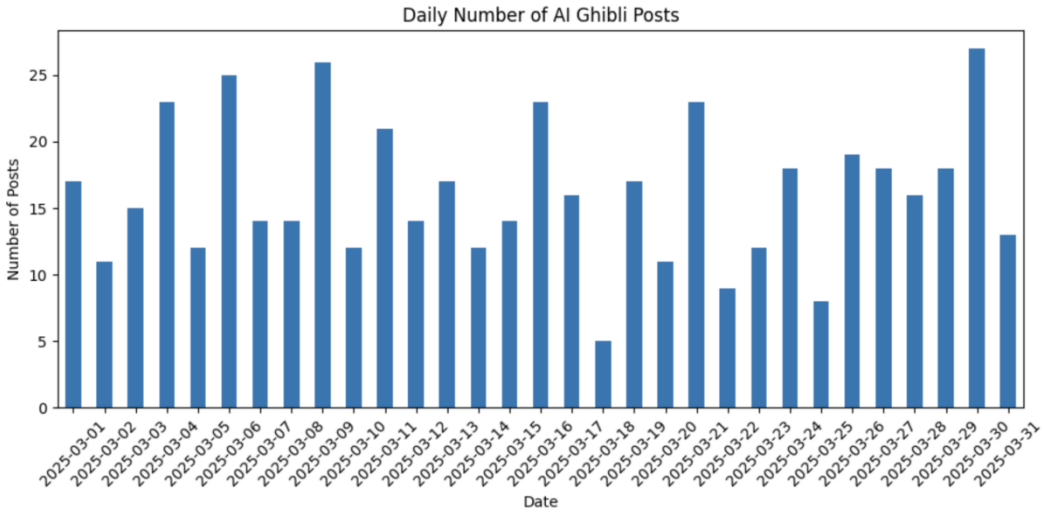
❖ Purpose of This Plot:

- Helps identify **posting frequency trends** over time.
- Reveals **peak activity days**, viral trends, or drops in interest.
- Useful for understanding **user engagement patterns**.

code :

```
# Plot 3: Daily Trend of Posts
daily_counts = df['creation_date'].value_counts().sort_index()
plt.figure(figsize=(10, 5))
daily_counts.plot(kind='bar')
plt.title("Daily Number of AI Ghibli Posts")
plt.xlabel("Date")
plt.ylabel("Number of Posts")
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

output :



Conclusion & Future Scope :

❖ Future Scope :

- **Sentiment Analysis**: Incorporate sentiment from user comments to measure emotional response toward different prompts.
- **Time Series Forecasting**: Predict future engagement or posting trends using models like ARIMA or Prophet.
- **Platform Comparison**: Deep dive into engagement metrics (likes, shares, comments) across platforms.
- **Image Analysis**: Use computer vision to analyze visual features of the art and correlate with engagement.
- **User Behavior**: Explore user demographics and behaviors, if available, to better personalize content.
- **Content Recommendation**: Build a model to recommend prompt styles based on past successful ones.
- **AI Model Improvement**: Use the feedback loop of likes and style accuracy to fine-tune generative AI models.

❖ Conclusion :

This project explored and visualized trends in AI-generated Studio Ghibli-style art using social media data. Key insights include:

- **Popularity by Platform**: Boxplot analysis revealed which platforms received the most engagement in terms of likes.
- **Top Performing Prompts**: We identified the most liked prompts and their platforms, helping understand user preferences.
- **Style Accuracy**: Histogram analysis showed the distribution of style accuracy scores, reflecting how well the AI-generated content matched the Ghibli aesthetic.
- **Posting Trends**: The daily trend plot helped visualize content creation patterns, showing spikes in activity and potential virality.

These insights offer a strong understanding of user engagement, content quality, and temporal trends in AI-generated art content.