Dataset Source & Justification

In this project, we focused on **evaluating and extending the performance of a pre- trained face restoration model (GFPGAN)** rather than training a new model from scratch. As such, the objective was not to build a large-scale training dataset, but to **carefully select a diverse and representative evaluation set** to assess real-world generalization and visual quality.

© Dataset Description

We curated a dataset of **30 short GIFs** sourced from publicly available, open-source platforms such as **Pexels**, **Pixabay**, **and GIPHY**. These GIFs were chosen to reflect a **broad range of realistic degradations** commonly found in user-generated media, including:

- Compression artifacts (typical of internet GIFs)
- Blur (motion or focus-related)
- Low resolution
- Noise (digital grain or poor lighting conditions)

Each GIF was frame-extracted and standardized for evaluation. The dataset includes:

Category	Count
Standard GIFs	20
Blurred GIFs	3
Noisy GIFs	3
Low-resolution GIFs	4
Total	30

Why a Smaller Dataset Works Here

- Model Evaluation, Not Training: Our goal was to assess and enhance the
 outputs of an existing, pre-trained model (GFPGAN v1.3). This does not require
 large-scale data for learning but rather focused, varied inputs for qualitative and
 quantitative benchmarking.
- **Diversity Over Volume**: The selected samples span multiple types of degradation, which allowed us to **stress-test the model's generalization** without requiring a large number of similar examples.

- Depth Over Breadth: Each sample was thoroughly analyzed using objective metrics (PSNR, SSIM, LPIPS, Temporal Loss) and manual human evaluation, offering a detailed performance profile that complements the model's original capabilities.
- Educational & Feasibility Scope: As part of an academic semester project, the dataset size was chosen to balance depth of analysis with time and resource constraints, while still delivering meaningful insights.

By focusing on **quality, diversity, and evaluation depth**, this dataset enabled us to effectively demonstrate how the base model can be extended and improved — without requiring training on large-scale data.