**Data Management Plan**

**Data Collection**

[Flickr Faces 70k Thumbnails 128x128 (kaggle.com)](https://www.kaggle.com/datasets/imcr00z/flickr-faces-70k-thumbnails-128x128)

**Overview of dataset.**

The Flickr-Faces-HQ (FFHQ) dataset, released by Nvidia Research in October 2018, comprises high-resolution face images sourced from Flickr, a photo-sharing platform. The dataset's creation involved automated web scraping techniques to gather images tagged with relevant keywords, followed by rigorous curation to ensure quality and relevance. While specific details about the data collection process and ethical considerations are not extensively documented, the dataset serves as a valuable resource for researchers and practitioners in computer vision and machine learning, particularly for tasks related to face synthesis and manipulation.

**Summary of dataset.**

The FFHQ (Flickr-Faces-HQ) dataset is a high-quality collection of 70,000 human face images, each at a resolution of 128x128 pixels in png format, totaling approximately 2 GB. These images feature diverse variations in age, ethnicity, and background, organized in subdirectories.

**Ethical Requirements.**

To ensure ethical use of the FFHQ dataset in compliance with GDPR, University of Helsinki ethical policy, and ethical collection standards, we will verify that the dataset is anonymized and devoid of personally identifiable information (PII), confirming explicit consent was obtained by original data collectors. We will also ensure the original data collection process adhered to ethical standards, documenting the dataset's origin and consent processes.

**Documentation.**

[Abdul-moe/Abdul-moe at my-project (github.com)](https://github.com/Abdul-moe/Abdul-moe/tree/my-project)

**Metadata**

[Abdul-moe/README.md at my-project · Abdul-moe/Abdul-moe (github.com)](https://github.com/Abdul-moe/Abdul-moe/blob/my-project/README.md)

**Security and Storage.**

To ensure security and proper storage of the FFHQ dataset, access will be restricted to authorized personnel using multi-factor authentication, and data will be encrypted in transit and at rest. Regular backups will be performed daily with weekly offsite backups, and versioning will be enabled.