Advanced Text-to-Image Generation

📘 Project Description

The **Advanced Text-to-Image Generation** project aims to develop a deep learning system capable of generating realistic, high-quality images directly from textual descriptions. The system will utilize **Generative Adversarial Networks (GANs)** and **Diffusion Models**, enhanced with **attention mechanisms** to improve semantic alignment between text and image features.

The project will also integrate **MLOps practices** for model versioning, monitoring, and deployment to ensure scalability, reproducibility, and efficient lifecycle management.

👥 Group Members & Roles

Member Name	Role	Responsibilities
Bishoy Adel Ezzat	**Team Leader / Data Engineer**	Dataset collection, text preprocessing, and ensuring image-text alignment.
Jonathan Gerges Fakhry	**Deep Learning Engineer**	Project coordination, GAN/diffusion model development, integration of attention mechanisms, and overall system design.
Nour Ahmed Mostafa Ibrahim	**MLOps Engineer**	Implement MLflow tracking, model versioning, CI/CD automation.
Fatma Hussein Ali Abdelfattah	**Full Stack Developer**	Develop and integrate the web interface or API for text-to-image generation and deployment.
Mahmoud khaled mahmoud abdulsamiee	**Evaluation & Documentation Specialist**	Evaluate model performance using metrics such as FID, IS, and Precision/Recall. Compare results with baseline or existing models .Write the final technical documentation and academic report

🎯 Objectives

- Develop an AI model that converts textual descriptions into realistic, high-resolution images.
- Compare and optimize different generative architectures (GANs, Diffusion Models).
- Implement attention mechanisms for better text–image alignment.
- Integrate the workflow into a complete MLOps pipeline with deployment readiness.

💼 Tools & Technologies

- **Programming:** Python, PyTorch, TensorFlow
- **NLP Tools:** Hugging Face Transformers, BERT
- **Image Libraries:** OpenCV, PIL
- **MLOps & Deployment:** MLflow, Docker, TensorBoard, FastAPI
- **Data Sources:** COCO Dataset, Flickr30k

📅 Milestones & Deadlines

Milestone	Description	Deadline
1. Data Collection & Preprocessing	Collect and preprocess image–text datasets (COCO, Flickr30k).	Week 1
2. Model Development & Training	Implement GAN and diffusion models for image generation.	Week 2
3. Advanced Techniques & Integration	Add attention mechanisms and integrate end-to-end pipeline.	Week 2 & 3
4. MLOps & Model Management	Set up model tracking, versioning, and deployment workflows.	Week 4
5. Final Report & Demo	Prepare final presentation, report, and live demonstration.	Week 4 & 5

📊 KPIs (Key Performance Indicators)

1. **Data Preparation Quality:**
   Completeness, cleanliness, and accuracy of image–text dataset alignment.
2. **Model Performance & Accuracy:**
   Evaluated by **FID**, **Inception Score**, and human qualitative assessment.
3. **Pipeline Integration & Automation:**
   Degree of automation in data preprocessing, model training, and inference stages.

4. **MLOps & Deployment Readiness:**
   Use of MLflow, Docker, and versioning for continuous integration and scalability.
5. **Output Quality & Usability Score:**
   Realism and relevance of generated images based on textual input.
6. **Documentation & Presentation Quality:**
   Clarity, completeness, and professionalism of project reports and demos.

🧩 Project Planning & Management

Problem Statement

Current text-to-image systems often lack precision and visual fidelity, especially in aligning complex textual descriptions with visual details. This project seeks to enhance text-to-image generation by integrating **attention-based deep generative models** and **robust MLOps practices** for better consistency and scalability.

Objectives

- Build a high-performing generative model capable of producing detailed images from complex text inputs.
- Ensure reproducibility and maintainability through structured MLOps pipelines.

Methodology / Approach

1. **Data Preparation:** Collect COCO/Flickr30k datasets and preprocess both text and images.
2. **Model Development:** Train and fine-tune GAN or Diffusion Models with attention mechanisms.
3. **Pipeline Integration:** Build an automated workflow integrating preprocessing, training, and evaluation.
4. **MLOps Setup:** Implement tracking, deployment, and versioning using MLflow and Docker.
5. **Evaluation:** Assess models with FID, IS, and human judgment for visual quality and relevance.

Expected Outcomes

- A fully functional AI system capable of generating images from text.
- Integration-ready pipeline with reproducible MLOps infrastructure.
- High-quality documentation and live demo for final presentation.

Tools & Technologies

Python, PyTorch, Transformers, TensorFlow, MLflow, Docker, FastAPI, TensorBoard.

🧠 Stakeholder Analysis

Stakeholder	Role	Interest / Contribution
Project Team	Developers, Researchers	System design, development, and integration
Supervisors / Mentors	Evaluators	Oversight, guidance, and technical validation
End Users	Content Creators, Designers	Access to image generation from textual prompts
Institution / Program	Support & Review	Ensures educational and technical standards
AI Community	Researchers	Potential reuse and scalability of the model

🗄 Database Design

The database will support storage, tracking, and management of text prompts, generated images, and model metadata.

Schema Overview

Table	Key Fields	Relationships
Users	user_id (PK), name, email, role	1-to-many with TextPrompts
TextPrompts	prompt_id (PK), user_id (FK), text_content, timestamp	Many-to-1 with Users
GeneratedImages	image_id (PK), prompt_id (FK), image_url, fid_score, created_at	Many-to-1 with TextPrompts
ModelVersions	model_id (PK), version_name, accuracy, loss, created_at	Independent for tracking experiments
Feedback	feedback_id (PK), image_id (FK), rating, comment	Many-to-1 with GeneratedImages