# Final Project Time-line

# Matjar_Assist: E-commerce Product Description Generator


**Prepared By:**

1. Shadwa Salah Sayed Soliman 21090739
2. Asmaa Hassan Hassan Soliman Shrief 21107488
3. Joy Ishak Fathy Asaad 21085542
4. Mohamad Gamal Ibrahim Fahim 21086192
5. Abdelrahman Ahmed Mohamed Elmohamady 21093667
6. Mina Boshra Asham Soliman 21039201
7. Ahmed Abdelsalam Abdelshafy Mansour 21072450

## Milestone 1: Data Collection and Preprocessing

| Task | Subtasks | Responsible | Deadline |
|---|---|---|---|
| Data Collection | Collect a relevant dataset for content generation (e.g., news articles, blog posts, product descriptions, etc.). Choose a domain-specific dataset based on your area of interest. | | 30 Sep.2025 |
| | Use publicly available datasets such as Common Crawl, Kaggle Text Datasets, or custom datasets scraped from websites using web scraping techniques | | |
| Data Preprocessing | Clean the dataset by removing any irrelevant content, special characters, and duplicates. | | |
| | Tokenize the text and normalize it (lowercasing, stemming, lemmatization). | | |
| | Perform additional preprocessing tasks such as removing stop words, punctuations, and handling numerical data. | | |
| | Split the dataset into training and validation sets. | | |
| Deliverables | Cleaned and Preprocessed Text Dataset | | |
| | Data Preprocessing Notebook | | |

## Milestone 2: Model Development and Training

| Task | Subtasks | Responsible | Deadline |
|---|---|---|---|
| Text Generation Mode | Choose a suitable architecture for text generation (e.g., GPT, GPT-2, or BERT based models from the Hugging Face Transformers library). | | 30 Oct. 2025 |
| | Fine-tune the model on the preprocessed dataset for content generation tasks. | | |
| Generative AI (GAN) Implementation | Develop a Generative Adversarial Network (GAN) for content generation. This GAN will consist of a Generator (creating synthetic text) and a Discriminator (evaluating the quality of the generated text). | | |
| | Train the GAN model on the dataset and iteratively improve the quality of generated text. | | |
| Training and Evaluation | Train both the text generation model and the GAN using powerful computing resources. Use AWS EC2 or Google Colab if necessary. | | |
| | Evaluate the performance of the models by generating sample content and assessing quality using Perplexity, BLEU Score, or human evaluations. | | |
| Deliverables | Trained Text Generation Model | | |
| | GAN Implementation | | |

## Milestone 3: Advanced Techniques and Pipeline Integration

| Task | Subtasks | Responsible | Deadline |
|---|---|---|---|
| Attention Mechanisms | Integrate Attention Mechanisms into the Transformer model to improve the quality of generated text. This can include techniques like Self-Attention in the context of sequence-to-sequence models. ) | | 15 Nov. 2025 |
| | Fine-tune the model to handle longer texts and better contextual understanding by implementing multi-head attention layers. | | |
| Generative AI Pipeline | Create a pipeline that automates data preprocessing, model training, and content generation. | | |
| | Integrate the preprocessing steps, model inference, and post-processing of generated content into a single automated system. | | |
| | Use tools like Apache Airflow or Kubeflow to orchestrate the pipeline and handle scheduling and execution. | | |
| Deliverables | Enhanced Text Generation Model with Attention Mechanisms | | |
| | Automated Generative AI Pipeline | | |

## Milestone 4: MLOps and Model Management

| Task | Subtasks | Responsible | Deadline |
|---|---|---|---|
| MLOps Integration | Use MLflow for managing the model lifecycle, including tracking experiments, monitoring model performance, and storing model versions. | | 30 Nov. 2025 |
| | Implement Model Versioning to keep track of different iterations of the model, ensuring reproducibility and traceability of content generation models. | | |
| Model Deployment and Monitoring | Deploy the text generation model to a cloud environment (e.g., AWS Lambda, AWS SageMaker, or Google Cloud AI Platform) for real-time content generation. | | |
| | Set up model monitoring and logging to track the performance and quality of generated content over time. | | |
| Automation and CI/CD | Integrate the pipeline into a Continuous Integration/Continuous Deployment (CI/CD) workflow for model retraining, deployment, and scaling. | | |
| | Use GitHub Actions, Jenkins, or CircleCI for automating the model deployment process. | | |
| Deliverables | Deployed Generative Content System | | |
| | MLOps Tracking with MLflow | | |

# Milestone 5: Final Report, Presentation, and Demonstration

| Task | Subtasks | Responsible | Deadline |
|---|---|---|---|
| Final Report | Data Collection and Preprocessing: Overview of the dataset and preprocessing steps. | | 7 Dec. 2025 |
| | Model Development: Explanation of the models used (GPT, Transformer, GAN), and the evaluation process. | | |
| | Pipeline Integration: Description of the generative AI pipeline and its automation. | | |
| | MLOps Practices: How MLflow and model management tools were implemented for versioning and monitoring. | | |
| | Business Impact: Discussion of the value this automated content generation system can provide, such as time-saving and scalability. | | |
| Final Presentation | Prepare a presentation highlighting key project achievements and challenges. | | |
| | Provide a live demonstration of the deployed generative content system in action. | | |
| Demonstration | Show how the system can generate content in real-time using the deployed model. | | |
| | Showcase the quality of the generated content, and the responsiveness of the system. | | |

**Thank You**