AI Project Stages and the Role of an AI Engineer/Researcher

Artificial Intelligence (AI) projects require a systematic workflow to ensure accuracy, scalability, and practical usefulness. One of the most widely used frameworks to structure AI projects is the **OSEMN Framework**—Obtain, Scrub, Explore, Model, and Interpret. Each stage plays a critical role in taking the project from problem definition to real-world deployment.
This article explains the stages of an AI project along with an example use case to illustrate the process.
Use Case Chosen: *Sentiment Analysis on Customer Reviews for an E-Commerce Platform*.

O – Obtain (Data Acquisition)

Description

The first stage focuses on understanding the business problem and identifying the data needed to solve it.

1. Understanding the Business Problem

A clear understanding of the problem helps the AI team design the correct approach.

Key aspects include:

- **Type of Problem:** Classification/regression/clustering/forecasting
- **Desired Outcome:** What insights or actions are expected?
- **Business Value:** What benefit will the solution bring?

For example, in a sentiment analysis project, the goal may be *to classify customer reviews into positive, negative, or neutral* to improve customer satisfaction strategies.

2. Collecting Relevant Data

Data is collected from internal or external sources.

The data acquisition activities include:

- Identifying data sources
  (databases, APIs, social media, logs)
- Specifying requirements
    - Format (CSV, JSON, Text, etc.)
    - Size/volume

    - Frequency (daily/real-time)
- Ensuring basic data availability and accessibility

Role of an AI Engineer/Researcher

- Understand the problem in depth (scope, timeline, limitations)
- Identify data requirements (types, formats, labels, frequency)
- Collaborate with business teams and data providers
- Validate whether collected data is usable for modeling
  Engineers are **not directly responsible for data collection**, but they verify quality and ensure it fits project needs.

S – Scrub (Data Cleaning and Preprocessing)

Description

Raw data usually contains noise, missing information, or irrelevant content. Cleaning prepares the dataset for meaningful analysis.

Activities include:

- Handling missing values
- Removing duplicates
- Text preprocessing (tokenization, stop-word removal, stemming/lemmatization)
- Selecting relevant features
- Converting data formats
- Labeling data (if required)

In sentiment analysis, cleaning involves removing special symbols, URLs, emojis, and correcting inconsistent spellings.

Role of an AI Engineer/Researcher

- Ensure the dataset is clean, consistent, and structured
- Design preprocessing pipelines for automation
- Apply NLP preprocessing techniques
- Document all cleaning steps for reproducibility
  The engineer ensures high-quality data enters the modeling phase.

E – Explore (Exploratory Data Analysis – EDA)

Description

Exploration helps understand patterns, correlations, and hidden insights in the dataset.

Activities include:

- Statistical summary generation
- Data visualization (graphs, histograms, word clouds)
- Identifying patterns and outliers
- Understanding class distribution (imbalanced vs. balanced dataset)

Example:
In sentiment analysis, EDA might reveal that 70% of reviews are positive, meaning the classes are imbalanced and need resampling.

Role of an AI Engineer/Researcher

- Perform EDA to uncover meaningful insights
- Identify potential issues (bias, imbalance, noise)
- Provide recommendations to adjust data collection or processing
- Decide the right features needed for modeling

EDA guides the direction for model selection and training.

M – Model (Modeling and Experimentation)

Description

This is the core stage of building and training machine learning or deep learning models.

Activities include:

- Selecting appropriate algorithms
  (Naive Bayes, SVM, LSTM, Transformers for sentiment analysis)
- Splitting data into training/validation/test sets
- Training the model
- Hyperparameter tuning
- Evaluating performance (accuracy, F1-score, confusion matrix)

Example:
For sentiment analysis, the engineer may try:

- Logistic Regression
- LSTM-based RNN
- BERT Transformer-based models

Role of an AI Engineer/Researcher

- Choose and justify suitable models
- Build, train, and tune models
- Implement ML pipelines
- Run multiple experiments to optimize performance
- Use version control for models and datasets
- Ensure model fairness, accuracy, and efficiency

This is the most technically intensive phase of the AI project.

N – Interpret (Interpretation and Deployment)

Description

The final stage deals with explaining results, measuring impact, and preparing the AI system for real-world use.

Activities include:

- Interpreting predictions and explaining decision boundaries
- Validating results with business teams
- Converting the model into deployable formats
- Deploying via APIs, cloud platforms, or integrated systems
- Monitoring model performance after deployment
- Updating and retraining when needed (MLOps)

Example:
Sentiment analysis results might be integrated into dashboards for customer service teams.

Role of an AI Engineer/Researcher

- Explain model behavior clearly to non-technical stakeholders
- Collaborate with DevOps for deployment
- Design monitoring pipelines (logging, drift detection)
- Create documentation, reports, and dashboards
- Ensure ethical, fair, and safe use of the model

Interpretation ensures the AI solution is understandable, reliable, and practically usable.

Conclusion

AI projects require a structured and disciplined approach to move from raw data to meaningful, deployable intelligence. Using frameworks like **OSEMN**, AI Engineers and Researchers ensure that each stage—Obtain, Scrub, Explore, Model, and Interpret—is carefully executed.
Their responsibilities span across understanding the business objective, analyzing and cleaning data, conducting experiments, deploying models, and ensuring the solution is trusted, scalable, and ethical.