

---

## *Projektaufgabe zur Vorlesung Explainable AI*

---

Nachfolgend eine detaillierte Strukturierung der Aufgabe mit einer klaren Aufgabenteilung, potenziellen XAI-Verfahren und Hinweisen für die Präsentation:

### **1. Projektstruktur & Teamaufteilung**

Jedes Team bearbeitet folgende Teilaufgaben, die gleichmäßig aufgeteilt werden sollten:

Teamgröße 5-6 Personen (Aufteilungsmethode wird in Vorlesung besprochen)

#### Datensatz-Auswahl & Preprocessing

- Wahl eines geeigneten Datensatzes (z. B. Bilddaten oder tabellarische Daten von Kaggle, UCI, etc.)
- Datenbereinigung und Vorverarbeitung (z. B. Normalisierung, One-Hot-Encoding, Feature Engineering)
- Begründung der Wahl des Datensatzes und Darstellung der Datencharakteristika

#### Training eines KI-Modells

- Auswahl eines geeigneten Modells (z. B. CNN für Bilder, Random Forest für Tabellendaten)
- Modellarchitektur erklären (Hyperparameter, Layers, Aktivierungsfunktionen)
- Training, Validierung, Hyperparameter-Tuning
- Darstellung der Modellgüte mit Metriken (Accuracy, F1-Score, etc.)

#### Implementierung und Anwendung von zwei XAI-Verfahren

- Auswahl von zwei unterschiedlichen XAI-Verfahren
- Detaillierte Erklärung der Funktionsweise
- Anwendung auf das trainierte Modell
- Visualisierung & Interpretation der Ergebnisse

#### Notebook und Ergebnisse vorstellen

- Strukturierung des Notebooks für die Präsentation

- Code-Organisation (gut kommentierter Code)
- Visuelle Darstellungen der XAI-Ergebnisse

### Analyse & Kritische Diskussion

- Vergleich der XAI-Verfahren:
  - Was überzeugt an den Verfahren?
  - Welche Nachteile haben sie?
  - Unterschiede in der Ergebnisinterpretation?
- Verbesserungspotenzial & Ausblick

### Präsentation (35 Minuten pro Team)

- Alle Teammitglieder sollen sich an der Präsentation beteiligen
- Inhaltlich klar strukturiert: Problemstellung → Modell → XAI-Methoden → Ergebnisse & Fazit
- Grafiken, Code-Snippets und Beispielergebnisse sollten anschaulich präsentiert werden
- Diskussion der Ergebnisse mit dem Publikum (inkl. kritischer Reflexion)

## **2. Auswahl möglicher XAI-Verfahren**

Je nach Art des gewählten Modells und Datensatzes können unterschiedliche Verfahren eingesetzt werden:

XAI-Verfahren für Bilddaten (z. B. CNNs)

- Grad-CAM
- Integrated Gradients
- Saliency Maps
- Layer-wise Relevance Propagation (LRP)
- Feature Visualization
- XRAI

XAI-Verfahren für tabellarische Daten (z. B. Random Forest, XGBoost, NN)

- SHAP (Shapley Values)

- Permutation Feature Importance
- Counterfactual Explanations
- Decision Tree Surrogates
- Anchors (Regelbasierte Erklärungen)

### **3. Vergleichskriterien der XAI-Verfahren**

Bei der Präsentation der Ergebnisse sollten die Teams folgende Aspekte diskutieren:

#### **1. Interpretierbarkeit**

- Wie einfach sind die Erklärungen zu verstehen?
- Sind sie für Laien oder nur für Experten verständlich?

#### **2. Modellabhängigkeit**

- Funktioniert das Verfahren nur für bestimmte Modelle (z. B. Grad-CAM nur für CNNs)?
- Ist das Verfahren modellagnostisch (z. B. SHAP)?

#### **3. Genauigkeit & Konsistenz**

- Wie präzise sind die Erklärungen?
- Sind die Ergebnisse stabil bei verschiedenen Input-Werten?

#### **4. Berechnungskosten**

- Wie aufwendig ist die Berechnung?
- Funktioniert das Verfahren für große Datensätze oder benötigt es viel Rechenzeit?

#### **5. Anwendungsszenarien**

- In welchen realen Anwendungsfällen wäre das Verfahren besonders sinnvoll?
- Gibt es Beispiele aus der Industrie oder Forschung?

### **4. Erwartete Ergebnisse & Reflexion**

Jedes Team sollte in der Präsentation folgende Punkte ansprechen:

#### **1. Visualisierung der XAI-Ergebnisse**

- Heatmaps für Bilddaten
- Feature-Wichtigkeitstabellen für tabellarische Daten

- Entscheidungsregeln oder Beispielerklärungen
- 2. Stärken & Schwächen der Verfahren
  - Welche Verfahren haben überzeugende Erklärungen geliefert?
  - Wo waren die Grenzen?
- 3. Verbesserungsvorschläge & zukünftige Arbeit
  - Wie könnten die Erklärungen weiter verbessert werden?
  - Welche weiteren XAI-Methoden könnten ausprobiert werden?

## **5. Praktische Hinweise für die Präsentation**

- Interaktive Elemente: Falls möglich, eine Live-Demo im Notebook durchführen
- Vergleichsgrafiken: Vorher-Nachher-Bilder oder Diagramme nutzen
- Klare Struktur: Jedes Teammitglied sollte eine klare Rolle haben
- Zeitmanagement: 35 Minuten der Präsentation gut aufteilen (10 Min Vorverarbeitung und Modelltraining, Fokusthema 15 Min XAI Verfahren Ergebnisse, 10 Min Diskussion der Ergebnisse)