XAI: Scientific Evaluation Process

1 **Definition of Psychological Target Variables**
e.g. System Trustworthiness, Predictability, Traceability, Oberserveability or Error Recognition Capability

2 **Familiarisation with Data Structures and Identification of Possible Risks**
What types of data are used, what influence can they have? What kind of situation is the person in? Which risks must be excluded or reduced?

3 **Interview-based Selection of Suitable Explanation Approaches Creation of Interface prototypes**
Selection from various, scientifically discussed approaches for generating explanations, e.g. counterfactual explanations

4 **Creation of Interface Prototypes and User tests with selected (controlled) data and errors**
Prototypical implementation based on the given data structure as well as construction of critical application cases (e.g. overfitting data) to test effectiveness

5 **Laboratory Testing (without real data) for Concrete Verification of Target Orientation and Critical Use Cases**
Examination with a larger sample for the general suitability of the approach, here no technical knowledge is required.

6 **Implementation into Partner System**
Evaluation with potential users of the system after connection to a trained model within the partner company

7 **Verification of the System in the Real Application Environment**
Final review and adaptation of the system and systematic evaluation of qualitative and quantitative psychological data