Technical Guide – Model 1 (Gologit Model) – Indirect Mapping

I. **Data Preparation**
- Input**:** Dataset containing predictor variables (e.g., 'WB_Faces_2', 'WB_Faces_4') and EQ-5D-Y dimensions as dependent variables.
- Predictors**:** Selected columns depend on the specific predictor set used.
- Outcome**:** Each EQ-5D-Y dimension (e.g., 'Mobility_EQ5D-Y', 'Pain_EQ5D-Y') is modeled as an ordered outcome.

II. **5-Fold Cross-Validation**
- K-Fold Splitting**:** The data is split into 5 folds for training and validation. In each fold:
  - Training Set**:** The GOLOGIT model is fitted using the training data.
  - Validation Set**:** The model predicts the outcome (utility) on the validation data.
- Predicted Utilities**:** Predicted utilities for the validation set are calculated and stored for further analysis.

III. **Model Evaluation**
- Metrics**:** For each fold, calculate:
  - MAE, MSE, and RMSE**:** Prediction accuracy.
- Bootstrapping**:** Confidence intervals for MAE, MSE, and RMSE are computed using bootstrapping (1,000 iterations).
- Coefficient Analysis**:** Coefficients and standard errors for each predictor are calculated and stored for each fold.

IV. **Results Storage**
- Metrics Results**:** Model performance metrics (AIC, BIC, MAE, MSE, RMSE) are saved in an Excel file and averaged across folds.
- Coefficients**:** Coefficients and standard errors for each predictor are saved in a separate Excel file, both per fold and as an average across folds.

Technical Guide – Model 2 (Bayesian Network Model) – Indirect Mapping

I. **Data Preparation**
- Input: Dataset containing EQ-5D-Y dimensions and predictor variables.
- Predictors: Selected predictor columns (e.g., 'WB_2', 'WB_4', 'WB_6'). The specific predictors depend on the predictor set being used.
- Outcome: Each EQ-5D-Y dimension (e.g., 'Mobility', 'Taking_care_of_myself', etc.).

II. **Chi-squared Tests**
- A chi-squared test is performed between each predictor and each EQ-5D-Y dimension.
- Significant Predictors: Predictors with a p-value < 0.05 are considered significant for that dimension.

III. **Model Creation (DAG Structure)**
- Bayesian Network: A Directed Acyclic Graph (DAG) is created for each dimension using the significant predictors as parent nodes.
- Predictor Interactions: Interactions between predictors are added to model dependencies.

IV. **Parameter Estimation**

- CPD Estimation: Conditional Probability Distributions (CPDs) for each predictor-dimension relationship are estimated using Maximum Likelihood Estimation (MLE).
- The CPDs are added to the Bayesian Network model for each EQ-5D-Y dimension.

V. **Predictions**

- Utility Calculation: For each data row, probabilities are calculated for each EQ-5D-Y dimension using the Bayesian Network model and CPDs.
- Utility Formula: A formula is applied to convert the probabilities into a predicted utility score, which is then stored in the dataset.

VI. **Model Evaluation**

- Fit and Accuracy: For each dimension: Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are calculated to assess prediction accuracy.
    - Bootstrapping: Confidence intervals for the accuracy metrics are computed via bootstrapping (1,000 iterations).

VII. **Results Storage**

- Metrics: Model performance metrics are saved in an Excel file for each EQ-5D-Y dimension.
- Utilities: Global utility statistics (mean, max, min predicted utility) are calculated and stored in the same Excel file.

Technical Guide – Model 3 (OLS Regression Model) – Direct Mapping

I. **Data preparation**

- Input: Dataset containing utility measures and predictors.
- Predictors: Selected columns such (e.g. 'WB Faces_2', 'WB Faces_4').
- Outcome: OBSERVED UTILITY

II. **5-fold cross-validation**

- K-Fold splitting: The data is split into 5 folds for training and validation. In each fold:
- Training Set: OLS regression is fitted using the training data.
- Validation Set: The model predicts utility on the validation data.
- Predicted utilities are stored in the dataset for further analysis.

III. **Model evaluation**

- Metrics: For each fold, calculate:
- AIC and BIC (Model fit)
- MAE, MSE, and RMSE (Prediction accuracy)
- Bootstrapping: Confidence intervals for metrics are computed via bootstrapping (1,000 iterations).
- Coefficient Analysis: Coefficients and standard errors for each predictor are calculated and saved per fold.

IV. **Results storage**

- Metrics results: The performance metrics are saved in an Excel file and averaged across folds.

- Coefficients: Coefficients and standard errors are saved separately for each fold and averaged across folds.

Technical Guide – Model 4 (Random Forest Regression Model) – Direct Mapping

I. Data Preparation

- Input: Dataset containing utility measures and predictor variables.
- Predictors: Selected columns depend on the specific predictor set used (e.g., 'WB_Faces_2', 'WB_Faces_4').
- Outcome: The observed utility (e.g., 'OBSERVED UTILITY').

II. 5-Fold Cross-Validation

- The dataset is split into 5 folds for training and validation. For each fold:
    - **Training Set**: The Random Forest model is trained on the training data.
    - **Validation Set**: The model predicts utilities for the validation data.
    - **Predicted Utilities**: Predicted utility values are stored in the dataset for further analysis.

III. Model Evaluation

- Metrics: For each fold, calculate:
    - MAE (Mean Absolute Error): Measures prediction accuracy.
    - MSE (Mean Squared Error) and RMSE (Root Mean Squared Error): Further accuracy measures.
- Bootstrapping: Confidence intervals for MAE, MSE, and RMSE are computed via bootstrapping (1,000 iterations).
- Utility Statistics: Calculate mean, minimum, and maximum predicted utility for each fold.

IV. Results Storage

- Metrics Results: Performance metrics (MAE, MSE, RMSE) and utility statistics (mean, min, max) are saved in an Excel file and averaged across folds.
- Predicted Utilities: The predicted utility values are stored in the original dataset for each fold and saved to Excel.