SUGGESTED WORKFLOW [SUBJECT TO CHANGE]

1. Data Preparation and Preprocessing:
   1. Clean and preprocess the data, handling missing values and any inconsistencies.
   2. Join session and product data as needed to create a unified dataset for further processing.
   3. Apply techniques like data sampling or data partitioning to work with large datasets more efficiently.
2. Feature Engineering:
   1. Extract relevant features from session and product data, such as session length, product categories, or user behavior patterns.
   2. Create additional features that might be helpful for the recommendation models, such as aggregated user preferences, temporal features, or product similarity measures.
   3. Normalize or scale features as needed.
   4. Utilize embeddings from pre-trained language models for text-based features like product titles and descriptions.
3. Validation Strategy:
   1. Split the enriched training data into a new training set and validation set, maintaining a similar distribution of session and product characteristics.
   2. Consider using cross-validation or other resampling techniques to maximize the use of available data.
   3. Evaluate the model's performance using a custom evaluation function that simulates the phase 1 test set conditions.
4. Model Development:
   1. Explore different types of recommendation models, including collaborative filtering, matrix factorization, and sequence-based models (e.g., RNNs, Transformers).
   2. Train these models on the new training set obtained from the validation strategy.
   3. Generate recommendations for the validation set using each model.
5. Model Selection and Optimization:
   1. Calculate an approximation of MRR for each model using the validation set, considering the model's top-ranked recommendations for each session.
   2. Select the best-performing model(s) based on the approximated MRR.
   3. Refine the selected model(s) further by tuning hyperparameters, adjusting features, or employing regularization techniques.
   4. Iterate steps 4 and 5 as needed to optimize model performance.
6. Model Evaluation on the Phase 1 Test Set:
   1. Apply the best-performing model(s) to the phase 1 test set without the ground truth.
   2. Generate the top 100 recommended product IDs (ASINs) for each session in the test set, ranked in decreasing order of confidence.
   3. Submit the recommendations to the competition organizers in the required format.
7. Monitoring and Updating the Model:
   1. Monitor the competition's public leaderboard and feedback from the organizers to assess the model's performance.
   2. Update the model as needed, incorporating new insights or techniques, and resubmit the recommendations.