

Document Validation Report for **OCEANAI**



Correspondence: Yes
Percentages: 90.0%

Conclusion:

The codebase substantially implements the methods and system architecture described in the Russian-language documentation. The documentation details preprocessing pipelines (tokenization, bag-of-words, normalization), feature extraction from audio/video/text, and construction of neural models for predicting personality traits using the Big Five framework, all of which are present in the code as modular, well-documented Python classes and scripts. Algorithmic elements such as hyperparameter optimization (referenced as genetic algorithms and surrogate modeling), dynamic graph construction based on feature attention, and model/inference pipelines are covered in the scripts and module structure (with functions for feature extraction, model assembly, loading weights, and multimodal fusion). The scripts support the workflow described, including input via pandas DataFrames, support for specific dataset formats, and interactive evaluation and prediction using pre-trained models. However, there is partial opacity regarding the explicit presence of the genetic algorithm and surrogate-based model search (these might be encapsulated or referenced only indirectly through the provided scripts/libraries). Also, the documentation describes certain evaluation/metric selection and graph-building steps for which the precise low-level code is not exhaustively confirmed in the provided inputs. Still, the core experimental setup, data handling, and outcome reproduction (given pre-trained model weights) is possible. Thus, the code aligns broadly with documentation, with minor potential gaps mainly in internal advanced optimization steps or dynamic graph routines.