

Methodology

This report is generated from a deterministic manifest (phase7-report-manifest-v1) and includes only script-produced artifacts.

- Manifest generated at: 2026-02-11T00:21:36Z
- Dataset artifacts tracked: 1
- Analysis artifacts tracked: 43
- Model artifacts tracked: 28

Core Analyses

Phase-3 and phase-4 outputs are referenced directly from the manifest.

Figures

- analyze_phase3:artifacts_analysis_phase3_core_stats_answer_length_by_label_png ? artifacts/analysis/phase3/core_stats/answer_length_by_label.png
- analyze_phase3:artifacts_analysis_phase3_core_stats_class_distribution_png ? artifacts/analysis/phase3/core_stats/class_distribution.png
- analyze_phase3:artifacts_analysis_phase3_core_stats_question_length_by_label_png ? artifacts/analysis/phase3/core_stats/question_length_by_label.png
- analyze_phase3:artifacts_analysis_phase3_lexical_top_bigrams_evasive_png ? artifacts/analysis/phase3/lexical/top_bigrams_evasive.png
- analyze_phase3:artifacts_analysis_phase3_lexical_top_bigrams_non_evasive_png ? artifacts/analysis/phase3/lexical/top_bigrams_non_evasive.png
- analyze_phase3:artifacts_analysis_phase3_lexical_top_unigrams_evasive_png ? artifacts/analysis/phase3/lexical/top_unigrams_evasive.png
- analyze_phase3:artifacts_analysis_phase3_lexical_top_unigrams_non_evasive_png ? artifacts/analysis/phase3/lexical/top_unigrams_non_evasive.png
- analyze_phase4:artifacts_analysis_phase4_question_behavior_question_behavior_by_label_png ? artifacts/analysis/phase4/question_behavior/question_behavior_by_label.png
- analyze_phase4:artifacts_analysis_phase4_semantic_similarity_semantic_similarity_by_label_png ? artifacts/analysis/phase4/semantic_similarity/semantic_similarity_by_label.png
- analyze_phase4:artifacts_analysis_phase4_topic_modeling_topic_prevalence_by_label_png ? artifacts/analysis/phase4/topic_modeling/topic_prevalence_by_label.png

Tables

- analyze_phase3:artifacts_analysis_phase3_artifact_index_json ? artifacts/analysis/phase3/artifact_index.json (stage: analyze)
- analyze_phase3:artifacts_analysis_phase3_core_stats_class_distribution_csv ? artifacts/analysis/phase3/core_stats/class_distribution.csv
- analyze_phase3:artifacts_analysis_phase3_core_stats_class_distribution_json ? artifacts/analysis/phase3/core_stats/class_distribution.json
- analyze_phase3:artifacts_analysis_phase3_core_stats_length_summary_csv ? artifacts/analysis/phase3/core_stats/length_summary.csv
- analyze_phase3:artifacts_analysis_phase3_core_stats_length_summary_json ? artifacts/analysis/phase3/core_stats/length_summary.json
- analyze_phase3:artifacts_analysis_phase3_core_stats_length_tests_json ? artifacts/analysis/phase3/core_stats/length_tests.json
- analyze_phase3:artifacts_analysis_phase3_core_stats_quality_metrics_csv ? artifacts/analysis/phase3/core_stats/quality_metrics.csv
- analyze_phase3:artifacts_analysis_phase3_core_stats_quality_metrics_json ? artifacts/analysis/phase3/core_stats/quality_metrics.json
- analyze_phase3:artifacts_analysis_phase3_lexical_lexical_summary_csv ? artifacts/analysis/phase3/lexical/lexical_summary.csv
- analyze_phase3:artifacts_analysis_phase3_lexical_lexical_summary_json ? artifacts/analysis/phase3/lexical/lexical_summary.json
- analyze_phase3:artifacts_analysis_phase3_lexical_top_bigrams_evasive_csv ? artifacts/analysis/phase3/lexical/top_bigrams_evasive.csv
- analyze_phase3:artifacts_analysis_phase3_lexical_top_bigrams_non_evasive_csv ? artifacts/analysis/phase3/lexical/top_bigrams_non_evasive.csv
- analyze_phase3:artifacts_analysis_phase3_lexical_top_ngrams_json ? artifacts/analysis/phase3/lexical/top_ngrams.json (stage: analyze)
- analyze_phase3:artifacts_analysis_phase3_lexical_top_unigrams_evasive_csv ? artifacts/analysis/phase3/lexical/top_unigrams_evasive.csv
- analyze_phase3:artifacts_analysis_phase3_lexical_top_unigrams_non_evasive_csv ? artifacts/analysis/phase3/lexical/top_unigrams_non_evasive.csv
- analyze_phase4:artifacts_analysis_phase4_artifact_index_json ? artifacts/analysis/phase4/artifact_index.json (stage: analyze)
- analyze_phase4:artifacts_analysis_phase4_question_behavior_question_behavior_comparison_csv ? artifacts/analysis/phase4/question_behavior/question_behavior_comparison.csv
- analyze_phase4:artifacts_analysis_phase4_question_behavior_question_behavior_metrics_csv ? artifacts/analysis/phase4/question_behavior/question_behavior_metrics.csv
- analyze_phase4:artifacts_analysis_phase4_question_behavior_question_behavior_summary_json ? artifacts/analysis/phase4/question_behavior/question_behavior_summary.json
- analyze_phase4:artifacts_analysis_phase4_question_behavior_taxonomy_metadata_json ? artifacts/analysis/phase4/question_behavior/taxonomy_metadata.json
- analyze_phase4:artifacts_analysis_phase4_semantic_similarity_hypothesis_summary_json ? artifacts/analysis/phase4/semantic_similarity/hypothesis_summary.json
- analyze_phase4:artifacts_analysis_phase4_semantic_similarity_qa_similarity_rows_parquet ? artifacts/analysis/phase4/semantic_similarity/qa_similarity_rows.parquet
- analyze_phase4:artifacts_analysis_phase4_semantic_similarity_semantic_similarity_by_label_csv ? artifacts/analysis/phase4/semantic_similarity/semantic_similarity_by_label.csv
- analyze_phase4:artifacts_analysis_phase4_semantic_similarity_semantic_similarity_by_label_json ? artifacts/analysis/phase4/semantic_similarity/semantic_similarity_by_label.json
- analyze_phase4:artifacts_analysis_phase4_topic_modeling_document_topics_parquet ? artifacts/analysis/phase4/topic_modeling/document_topics.parquet
- analyze_phase4:artifacts_analysis_phase4_topic_modeling_model_config_json ? artifacts/analysis/phase4/topic_modeling/model_config.json
- analyze_phase4:artifacts_analysis_phase4_topic_modeling_topic_prevalence_by_label_csv ? artifacts/analysis/phase4/topic_modeling/topic_prevalence_by_label.csv
- analyze_phase4:artifacts_analysis_phase4_topic_modeling_topic_summary_json ? artifacts/analysis/phase4/topic_modeling/topic_summary.json
- analyze_phase4:artifacts_analysis_phase4_topic_modeling_topic_top_terms_csv ? artifacts/analysis/phase4/topic_modeling/topic_top_terms.csv

Modeling

- Best classical family (macro F1): boosting
- Classical metrics snapshot: {'accuracy': 1.0, 'f1_macro': 1.0, 'precision_macro': 1.0, 'recall_macro': 1.0}
- Transformer metrics snapshot: {'accuracy': 0.5, 'f1_macro': 0.3333333333333333, 'precision_macro': 0.25, 'recall_macro': 0.5}

Model artifacts used:

- phase5_classical_baselines:artifacts_models_phase5_boosting_classification_report_json ? artifacts/models/phase5/boosting/classification_report.json
- phase5_classical_baselines:artifacts_models_phase5_boosting_confusion_matrix_json ? artifacts/models/phase5/boosting/confusion_matrix.json
- phase5_classical_baselines:artifacts_models_phase5_boosting_metrics_json ? artifacts/models/phase5/boosting/metrics.json
- phase5_classical_baselines:artifacts_models_phase5_boosting_model_bundle_pk1 ? artifacts/models/phase5/boosting/model_bundle_pk1
- phase5_classical_baselines:artifacts_models_phase5_boosting_run_metadata_json ? artifacts/models/phase5/boosting/run_metadata.json
- phase5_classical_baselines:artifacts_models_phase5_logreg_classification_report_json ? artifacts/models/phase5/logreg/classification_report.json
- phase5_classical_baselines:artifacts_models_phase5_logreg_confusion_matrix_json ? artifacts/models/phase5/logreg/confusion_matrix.json
- phase5_classical_baselines:artifacts_models_phase5_logreg_metrics_json ? artifacts/models/phase5/logreg/metrics.json
- phase5_classical_baselines:artifacts_models_phase5_logreg_model_pk1 ? artifacts/models/phase5/logreg/model_pk1
- phase5_classical_baselines:artifacts_models_phase5_logreg_run_metadata_json ? artifacts/models/phase5/logreg/run_metadata.json