MODEL CARD

Version 6.0 | CORAL Ordinal Regression Models

Impacter® Pathway

Overview & Scope

This model suite scores K-12 student open-ended responses on eight human-skills competencies using a fine-tuned DeBERTa-v3-base architecture. Models generate rubric-aligned ordinal scores calibrated to human expert scoring of student reasoning, reflection, and narrative evidence.

Models are trained on annotated student responses from multiple school district implementations and evaluated using standard educational measurement agreement statistics (QWK, adjacent agreement, SMD).

Intended Use

- Scoring student reflections for Portrait of a Graduate / soft-skills competencies
- Behavioral health and social-emotional reflection scoring
- Workforce and career readiness / CTE assessments
- Short-form constructed-response tasks

Out of Scope

- High-stakes summative academic content scoring
- Individual diagnostic or clinical decisions without human review
- Any use outside of rubric-anchored prompts and approved task types

Training & Evaluation Data

The model is trained on district-provided student responses of written and transcribed speech paired with hand-scored human evaluations on a 0-4 scale. These examples come from authorized educational partners and represent authentic answers to open-ended prompts. Scoring is performed on newly collected student responses using held-out validation sets.

Model Configuration

Models are trained using CORAL ordinal regression and evaluated on held-out data. New versions are registered in MLflow and promoted to Production only if QWK improves relative to the current Production model.

Once promoted, Production models are frozen in the registry for deterministic scoring and full auditability. All inference uses frozen Production models with deterministic tokenization and scoring, ensuring identical responses receive identical scores across administrations.

Model releases incorporate SMD and subgroup fairness checks aligned to educational standards (QWK ≥ 0.70, degradation ≤ 0.10, |SMD| < 0.15, $|SMD|_S$< 0.10).

Model Configuration

- **Architecture:** Microsoft/ DeBERTa-v3-base (184M parameters)
- **Version:** IMPACTER Model Suite v6.0
- **Loss Function:** CORAL Ordinal Regression
- **Platform:** Databricks with MLflow Model Registry

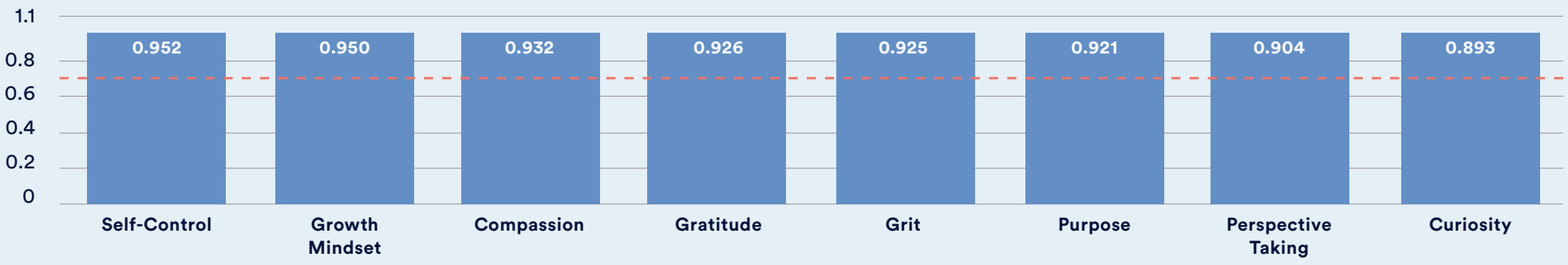
Summary Statistics

- **QWK (Quadratic Weighted Kappa):** Measures human-ML scoring agreement with higher weight on larger disagreements. Range: -1 to 1, where 1 = perfect agreement.
- **Accuracy:** Percentage of ML scores that exactly match human scores. Measures precise scoring alignment.
- **Adjacent Accuracy:** Percentage of ML scores within ±1 point of human scores. Typical human-human rater agreement benchmark.

AVG QWK	AVG ACCURACY	AVG ADJACENT ACC
0.918	77.3%	98.0%

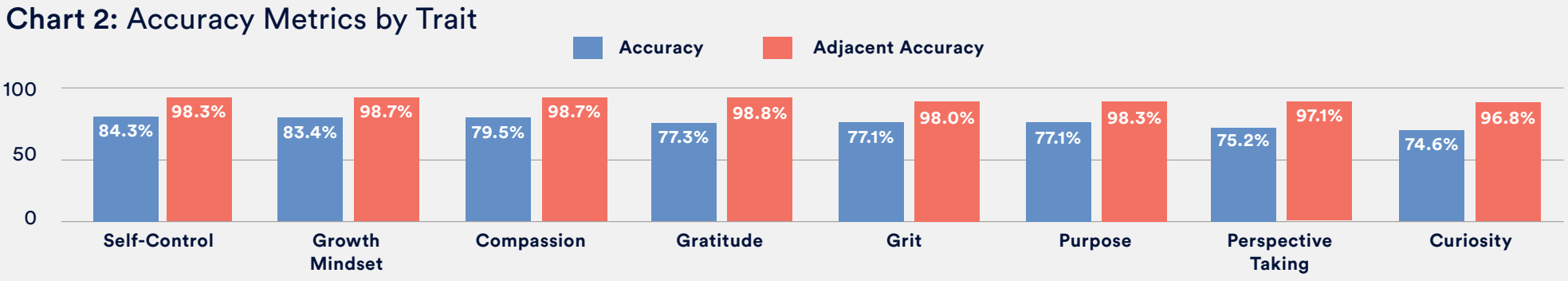
Data Visualization

Chart 1: QWK Performance by Trait



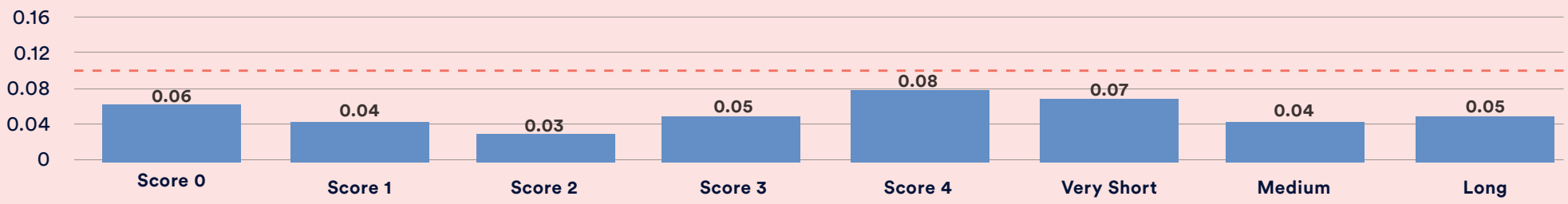
Red line indicates industry-standard minimum ML scoring threshold (QWK ≥ 0.70). All traits exceed this threshold.

Chart 2: Accuracy Metrics by Trait



Exact accuracy ranges from 75–84% while adjacent agreement (±1) ranges from 97–99% across traits, consistent with human-rater monitoring expectations.

Chart 3: Preliminary Fairness Check (Structural SMD)



Preliminary distributional SMD checks (by score/length bands) show no systematic drift.
Full subgroup SMD analysis (EL status, race/ethnicity) will be computed using partner-specific human-scored validation sets.