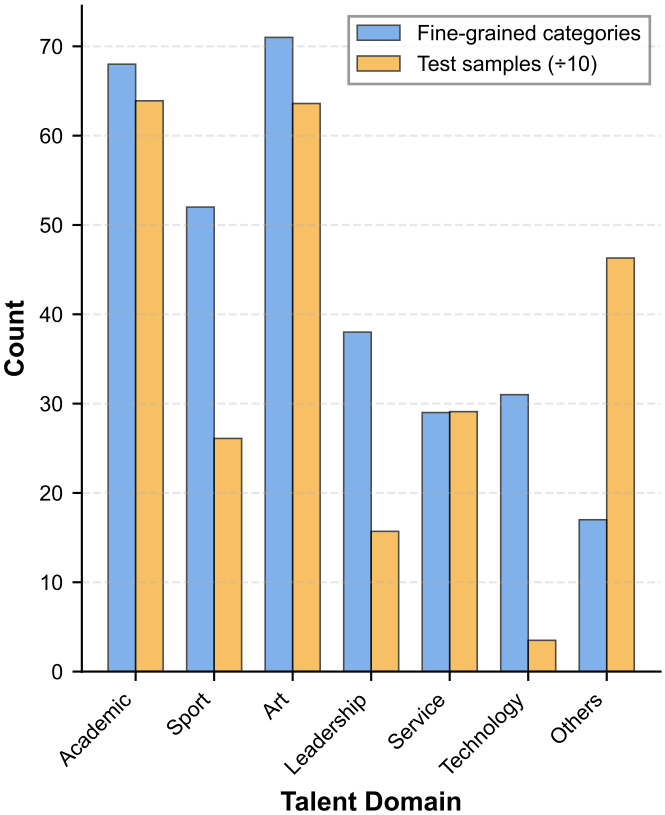
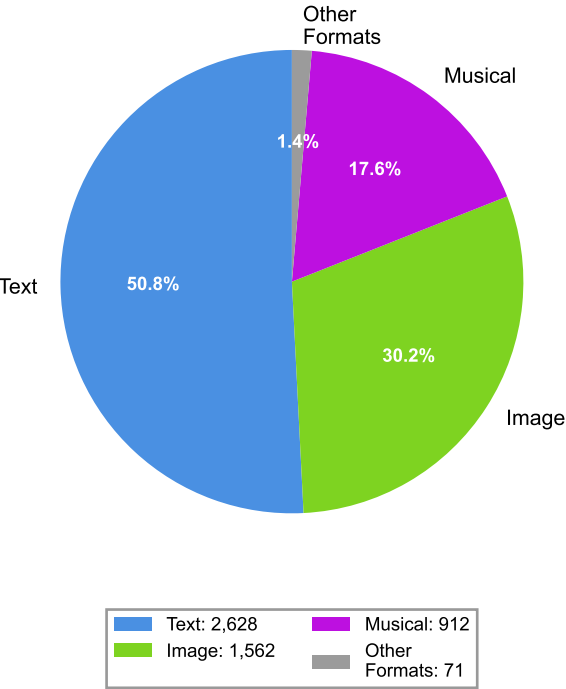


Figure S3: Dataset Composition and Characteristics

B. Domain Coverage (Categories & Samples)



A. Artifact Type Distribution (n=5,173)



C. Dataset Statistics Summary

Metric	Value
Total analyses	5,173
Training set	4,491 (86.8%)
Test set	682 (13.2%)
Train/test split	87:13
Total children	479
Fine-grained categories	306
Top-level domains	7
Artifact modalities	8