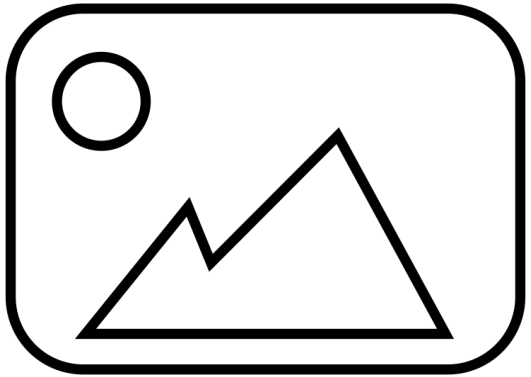
Shared Misclassifications: A Data-Centric Problem

To diagnose the source of errors, we analyzed their overlap across all embedding models—from the large, general-purpose models to our small, fine-tuned specialist. The results provide strong evidence that the errors are systematic.

The Finding: Models Agree on What's Hard

A significant portion of failures are not random but are systematic products of the course catalog data itself.

- A large number of misclassified course pairs were common to all model combinations.
- This indicates that these "hard" examples consistently challenge a wide range of semantic models.
- Such errors often arise from annotation artifacts, inherent ambiguity in the source text, or insufficient information to support a clear classification.



Interpretation

These shared failures strongly suggest that the errors