## Core Principle: Decoupling Representation from Classification

By separating the process into two stages, we gain the semantic power of deep learning while avoiding the high operational costs of end-to-end LLM classification [1] and the privacy risks of enrollment-based methods [2].

## Stage 1: Semantic Representation

The computationally intensive work of understanding language is done **once**, **offline**.

- A deep embedding model converts raw course text into a structured, reusable semantic vector.
- This captures the nuanced meaning and context of the course description.

## Stage 2: Pairwise Classification

The classification of course pairs becomes computationally cheap and fast.

- A traditional machine learning model simply compares the pre-computed vectors.
- This allows for rapid, on-demand comparison of any two courses in the database.

## The Benefit: The Best of Both Worlds

We leverage the power of transformers for deep semantic understanding without incurring their high inference costs for every comparison, creating a highly scalable system.



Mark S. Kim (SESU)

Automating Course Articulation July 8, 2025

Automating Course Articulation

Introduction

Core Principle: Decoupling Representation from Classification

By supervising the process into the stages, wa gain the semanticip power of they having which sanding the high questionist and offer-steer LMC deficiently all they have present such as develows [27].

Stage 1: Semantic Representation
The computationally intensive sork of members of the computational presents of the computational of computational presents of the computational present of the computational presents of

Core Principle: Decoupling Representation from Classification

We leverage the power of transformers for deep semantic understanding without incurring their high inference costs for every comparison, creating a highly scalable system.