Applications of AI to Assessment: Leveraging Vector Databases

Jake Cho, PhD

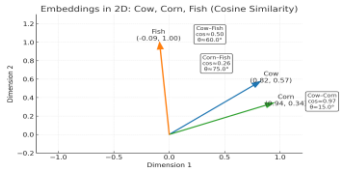
Motivation and Context

High-stakes assessment programs increasingly depend on large, dynamic item banks, where maintaining fairness, consistency, and psychometric quality is essential. Traditional workflows require extensive manual review to detect enemy items, predict item difficulty, and classify new content—processes that are time-consuming. Recent advances in vector databases and local large language models (LLMs) provide new opportunities to automate and enhance these psychometric workflows. By representing each test item as a high-dimensional vector, semantic similarity between items can be quantified, enabling instant enemy item detection through real-time similarity search, pretest item difficulty prediction via vector-based comparison to calibrated items, and data-driven insights to help SMEs and psychometricians identify the most relevant domain for new/revised items through similarity mapping.
The goal of this project is to develop a proof-of-concept system that demonstrates how embeddings, vector search, and local LLMs can enhance psychometric review and item bank management.

Major Terms

Embedding

Embeddings are numerical representations of text that capture its meaning in a continuous, high-dimensional vector space.
In simpler terms, they convert each test item (stem, options, or prompt) into a list of numbers that describe its semantic content.



Vector DB

Traditional **Relational Databases (RDBs)** store structured data like item IDs, domains, and psychometric parameters for exact-match queries.
In contrast, a **Vector Database** stores numerical *embeddings* that capture the semantic meaning of text, enabling similarity searches based on conceptual closeness rather than keywords.

Using **PostgreSQL with the pgvector extension** seamlessly integrates both—allowing relational data management and real-time vector similarity search in one system. This hybrid design is simple, scalable, and ideal for AI-augmented test development.

Local SLM

A **local language model** is a language model deployed on-premises or within a private server, rather than through external cloud APIs. This ensures **data confidentiality**, as sensitive assessment content—such as item stems, keys, or psychometric metadata—never leaves the organization's secure environment.

For small- to mid-size certification and licensure programs, a **Small Language Model (SLM)** offers a cost-effective alternative, supporting secure use cases such as:

- Automated item review and feedback generation
- Item modeling and item development
- Integration with vector databases for semantic search

Applications

Instant Enemy Item Detection

AI-powered vector similarity search identifies pairs of items that are too similar in content or structure—so-called *enemy items*—which can compromise fairness or threaten item security.

- **Real-time detection:** Each newly written item is converted into a high-dimensional vector and compared against the existing operational pool within seconds.
- **Informative feedback to item writers:** Highlights subtle conceptual overlap across items, guiding refinement and reducing overlooked redundancy.
- **Enhanced test security:** Prevents unintended item exposure and content redundancy in LOFT or adaptive test forms.

Pretest Item Difficulty Prediction

The system estimates an item's expected Rasch difficulty (b-parameter) prior to calibration by locating the most semantically similar calibrated items within the vector space. By leveraging these similarity relationships, the model infers the probable difficulty level of new items, providing early psychometric insight before empirical data collection.

- **Similarity-based inference:** Retrieves the top *k* most similar calibrated items to estimate pretest difficulty.
- **Empirical validation:** Compares predicted and actual difficulties to confirm accuracy.
- **Improved form balance:** Supports assembling equivalent forms early in development, even before pilot data are available.

Assisted Classification

New / revised items are AI-assisted and preliminarily categorized by domain or cognitive level through similarity mapping, which compares each new item to an existing labeled item bank to inform accurate and consistent classification across content areas.

- **Vector-based domain alignment:** The model identifies the most similar items and assigns the majority domain label among them.
- **Streamlined SME workflow:** Provides an initial domain suggestion that SMEs can verify and adjust, speeding up classification.

Implementation

Sample Item Pool

Sample Item Pool:

- Domains: 8
- Topics: 64
- Active Items: 1,510 (≈89%)

Embedding in Vector DB and GUI

- Set up a database, load the item bank, and generate embeddings by combining the stem and choices columns using the nomic-ai/nomic-embed-text-v1.5 model (768 dimensions).
- Develop a Streamlit GUI that provides the required functionality.

Next Step

After validating the prototype, the system will be evaluated using the full operational item pool. Pending successful performance, the project will advance to a full-scale implementation phase.