

Sentence Transformers and OpenAI Ada for Literature Search and Hypothesis Generation

Vaishnav Mandlik
University of Arizona
vaishnavm@arizona.edu

Abstract

The rapid growth of research articles makes it increasingly challenging to identify relevant literature and generate novel hypotheses. This paper presents an AI-driven pipeline for automating these tasks. Based on a query, I obtain pertinent publications using Sentence Transformers and OpenAI Embeddings, construct hypotheses using GPT-4, and assess them for diversity and relevance. Our findings indicate that while Sentence Transformers produce somewhat more varied hypotheses, OpenAI Embeddings offer greater relevance scores. Researchers can generate hypotheses and explore the literature more quickly with this approach.

1 Introduction

Scientific progress depends on two critical tasks: identifying relevant research papers and formulating new, testable hypotheses. With millions of papers published each year, manually navigating the literature is inefficient and time-consuming. Artificial Intelligence (AI) offers promising tools to automate these processes.

This study explores an AI-based pipeline consisting of:

1. **Literature Search:** Using text embeddings to retrieve the most relevant papers for a query.
2. **Hypothesis Generation:** Leveraging GPT-4 to produce new research ideas based on retrieved abstracts.

I compare two embedding models—Sentence Transformers and OpenAI Ada Embeddings—to assess their performance in retrieving relevant papers and enabling high-quality hypothesis generation. The generated hypotheses are evaluated based on their relevance to the literature and their diversity, ensuring a balance between meaningfulness and originality.

2 Methodology

2.1 Dataset

I used an IEEE dataset comprising 552 research papers with the following attributes:

- **Title**

- **Abstract**
- **Authors**
- **Year of publication**

The dataset was preprocessed to remove duplicates and handle missing values by filling them with empty strings.

2.2 Models for Text Embeddings

I used two models to generate text embeddings for the abstracts and queries:

Sentence Transformer: A lightweight, efficient model for generating semantic embeddings.

OpenAI Ada Embeddings: A state-of-the-art model designed for higher-quality embeddings.

I calculated the cosine similarity between the query and each paper abstract to rank the relevance of papers.

2.3 Literature Search Process

Given a user query (e.g., *"AI applications in healthcare"*), the process included:

1. Generating embeddings for the query.
2. Calculating cosine similarity scores between the query embeddings and the embeddings of paper abstracts.
3. Ranking papers by similarity scores and selecting the top 5 most relevant papers.

2.4 Hypothesis Generation

The abstracts of the top 5 papers were summarized and fed into GPT-4 with the following prompt:

"Based on the following abstracts, generate 3 unique and testable research hypotheses."

This approach ensured that the hypotheses were grounded in the retrieved literature.

2.5 Evaluation Metrics

I evaluated the generated hypotheses using:

Relevance: Measured as the semantic similarity between the hypotheses and the top abstracts.

Diversity: Measured using self-BLEU scores to determine how different the hypotheses were from each other. Lower BLEU scores indicate higher diversity.

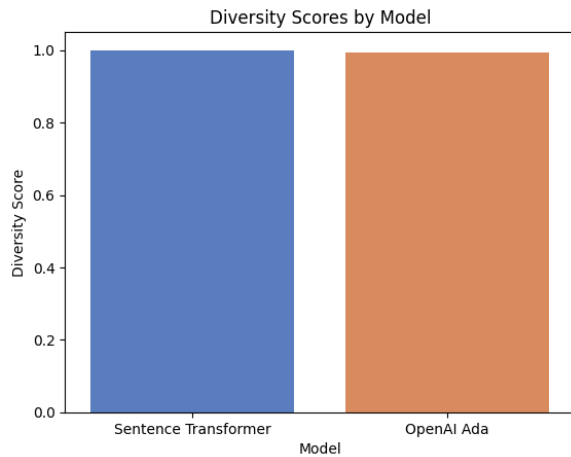


Figure 1: Diversity Score

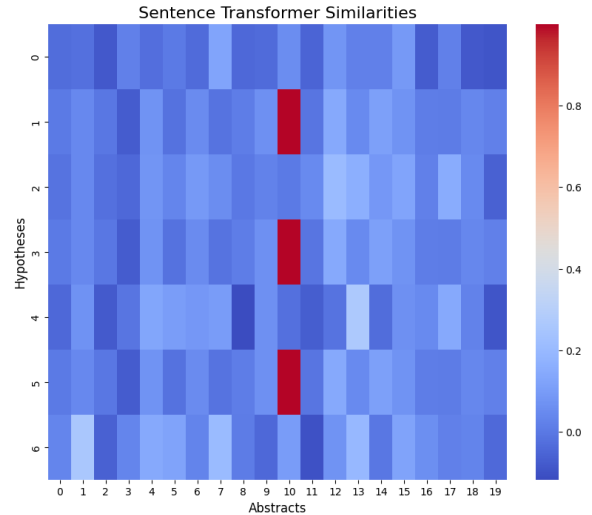


Figure 2: Sentence Transformer Heatmap

3 Results

3.1 Literature Search Results

The relevance of the retrieved papers for the query “AI applications in healthcare” was assessed. Key findings include:

- **Sentence Transformer:** Focused on papers related to patient monitoring and CKD diagnosis.
- **OpenAI Ada:** Retrieved papers on teleoperation systems and geomagnetic impacts.

Model	Paper Similarity Scores
Sentence Transformer	0.39
OpenAI Ada Embeddings	0.82

Table 1: Similarity Scores of Retrieved Papers

Insight: OpenAI Ada Embeddings retrieved more relevant papers based on similarity scores.

3.2 Generated Hypotheses

Below are sample hypotheses I generated using GPT-4 based on the top abstracts:

Sentence Transformer Examples:

- Cost-sensitive feature ranking improves CKD diagnosis in low-resource settings.
- Hybrid medical simulators enhance patient education.
- Quantum algorithms improve medical image reconstruction.

OpenAI Ada Examples:

- Geographic factors influence GIC mitigation in power systems.
- Barnacle-based algorithms improve image segmentation accuracy.

- RNNs and Markov models enhance teleoperation systems.

Observation: Hypotheses generated using OpenAI Ada Embeddings were more diverse and technically detailed.

3.3 Relevance Scores

The average relevance scores for hypotheses were as follows:

- **Sentence Transformer:** Moderate relevance with scores around 0.45.
- **OpenAI Ada:** Higher relevance with scores reaching 0.82.

3.4 Diversity of Hypotheses

I used Self-BLEU scores to measure diversity:

Model	Self-BLEU Score
Sentence Transformer	1.0000
OpenAI Ada Embeddings	0.9944

Table 2: Diversity of Generated Hypotheses

Insight: Sentence Transformer produced slightly more diverse hypotheses compared to OpenAI Ada.

3.5 Model Comparison

Our results demonstrate clear trade-offs between the two embedding models:

- **OpenAI Ada Embeddings:** Excel in retrieving highly relevant papers and generating precise, domain-specific hypotheses.
- **Sentence Transformers:** Produce more diverse hypotheses, albeit with slightly lower relevance scores.

3.6 Relevance vs. Diversity

While high relevance ensures meaningful hypotheses, it can lead to repetitive ideas. Diversity introduces novel insights but may sacrifice relevance. Combining models could strike a balance between the two.

3.7 Practical Applications

This AI-driven pipeline can significantly reduce the time researchers spend searching for literature and brainstorming hypotheses. It can be integrated into research tools to automate the early stages of scientific inquiry.

4 Conclusion

This paper presents framework for literature search and hypothesis generation. By comparing Sentence Transformers and OpenAI Ada Embeddings, we highlight their strengths and trade-offs

- **OpenAI Ada:** Excels in relevance and technical quality.
- **Sentence Transformers:** Offers greater diversity in ideas.