

Math, Magic, and Meaning

Unlocking AI's Hidden Superpower

(Presentation AND Notebook)

<https://github.com/nagoodman/embedding-tutorials>

Agenda

- Data Barn Delights
- What Are AI Embeddings?
- Real-World Applications
 - Semantic Search
 - Recommendation Systems
 - Response-Augmented Generation
- Challenges
- Future

Nicholas Goodman



LinkedIn: @nicholasandrewgoodman
X: @notachanceimstillthere

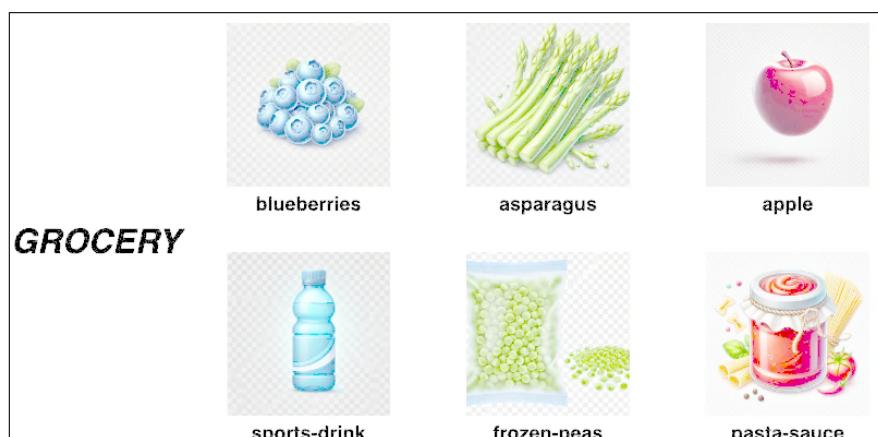
Barn Delights: Products as Embeddings



blue

green

red



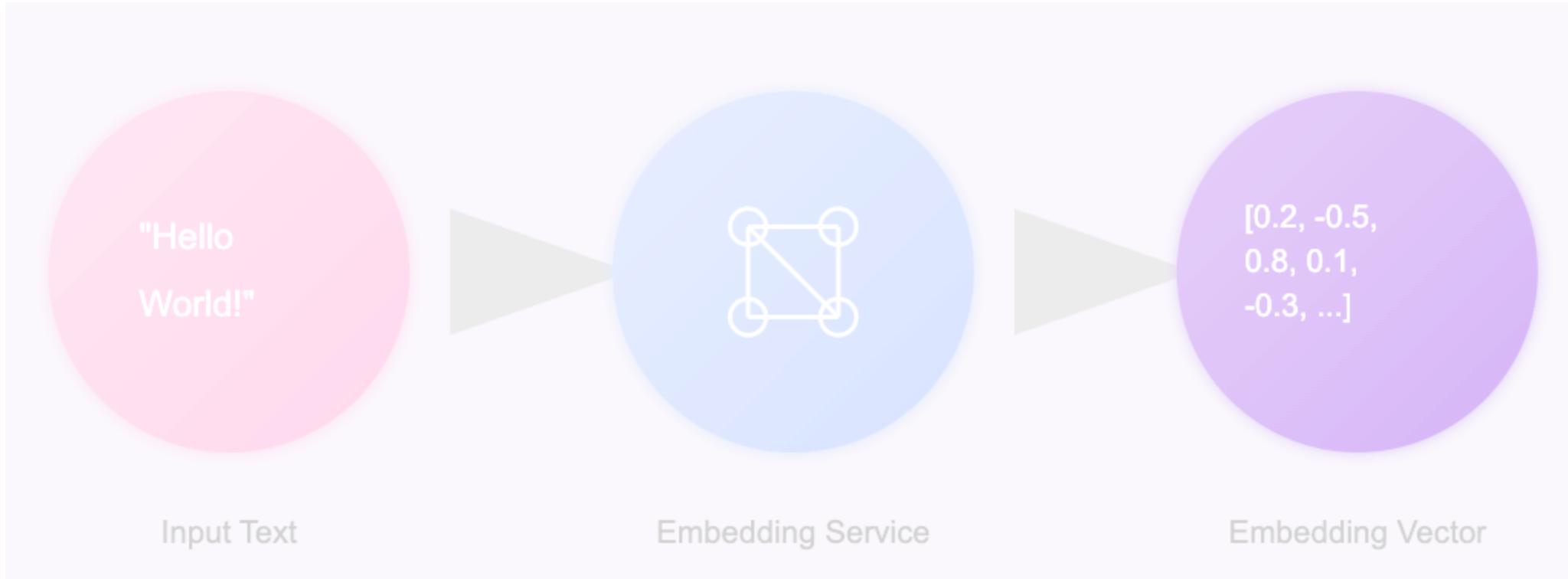
id	text	embedding
apple	esh red apple is perfect for snacking o...	[0.004357376601547003, -0.015314856544137001, ...]
asparagus	esh green asparagus is ideal for health...	[0.0060213059186935425, -0.024946179240942, 0....]
blueberries	fresh blue blueberries are perfect for s...	[0.00741775194182992, -0.026721514761447906, -...]
fertilizer	een fertilizer boosts plant growth and ...	[0.013712795451283455, 0.0327032133936882, 0....]
frozen-peas	een frozen peas are a versatile and he...	[-0.015473926439881325, -0.03986012190580368, ...]
mop	ed mop is designed for efficient floor c...	[0.05619010701775551, 0.03280946984887123, -0....]
pasta-sauce	d pasta sauce is perfect for adding fla...	[-0.05282636731863022, -0.037101104855537415, ...]
planting-bed	ed planting bed is ideal for creating or...	[0.0010808962397277355, -0.021274976432323456,...]
sponge	een sponge is perfect for all your clea...	[0.028249232098460197, 0.011501634493470192, 0....]
sports-drink	e sports drink is designed to keep you...	[0.03546168655157089, 0.03785005956888199, -0....]
watering-can	ue watering can is perfect for your gar...	[0.07881119847297668, 0.0558403879404068, 0.01...
window-cleaner	i window cleaner leaves surfaces strea...	[0.0016394169069826603, 0.05489180609583855, -...]

What Are AI Embeddings?

- **Definition:**
 - Vector representations of data in high-dimensional space.
- **Concept:**
 - Data points like words, images, or products are transformed into coordinates in a mathematical space.



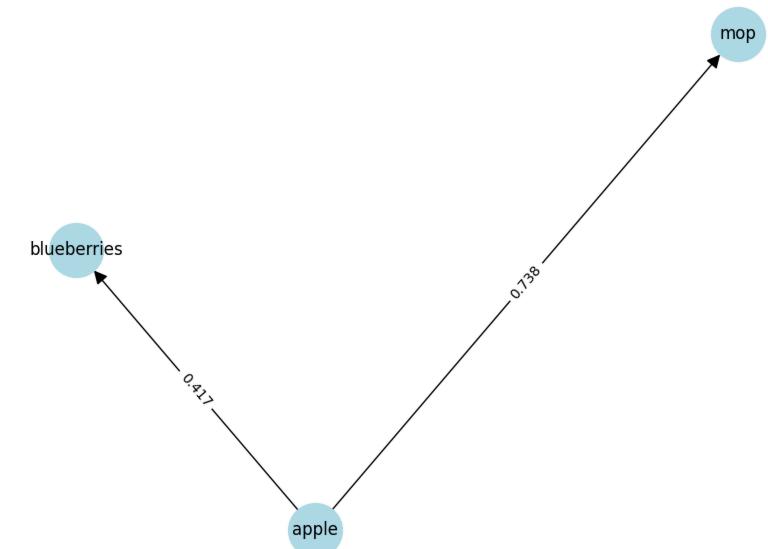
What Are AI Embeddings? (cont.)



What Are AI Embeddings? (cont.)

Why Use Them?

- Enable machines to understand emotions, meaning, and relationships.
- "Apple" and "Blueberries" are closer in this space than "Apple" and "Mop."



What Are Embeddings Used For?

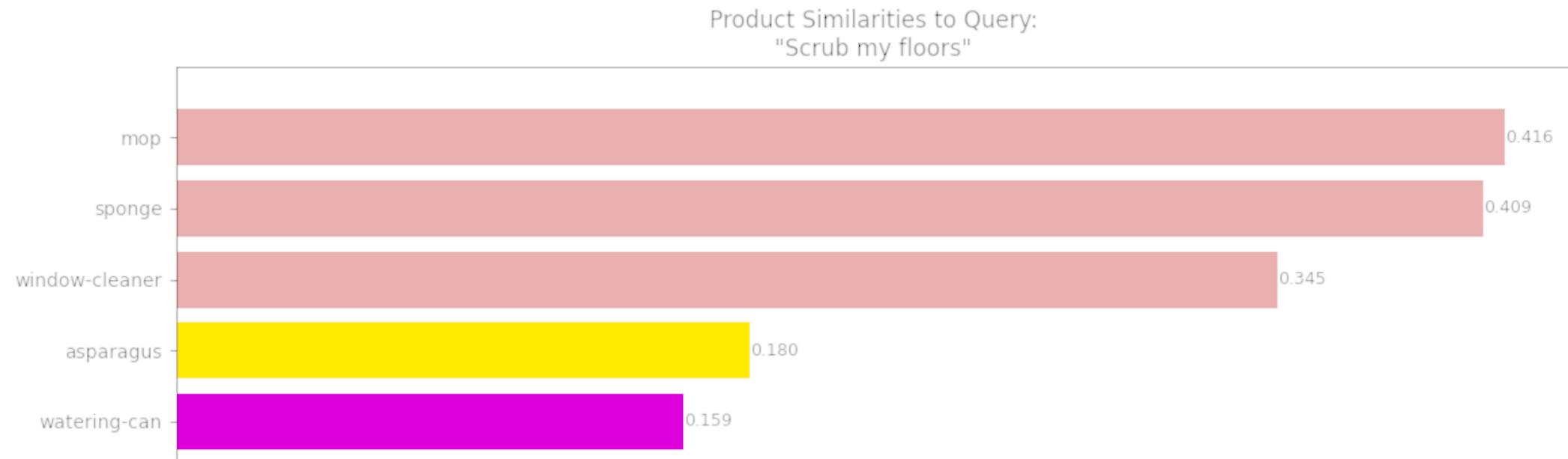
Semantic
Search

Recommendation
Systems

Retrieval-
Augmented
Generation

1. Semantic Search

- **Problem:** Queries don't always match document keywords.
- **Embedding Solution:** Find relevance by meaning, not exact matches.
- **Barn Delights Example:**



2. Recommendation Systems

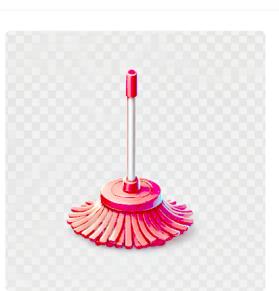
- **Embedding Role:** Group similar items based on similarity.



sponge

Similarity: 0.49908

This green sponge is perfect for all your cleaning needs. Use it to scrub dishes, countertops, and more. With its durable material, this sponge can handle tough grime while being gentle on surfaces. It is designed to absorb water efficiently, making your cleaning tasks quick and easy. Simply rinse after use and it's ready for the next job.



mop

Similarity: 0.46217

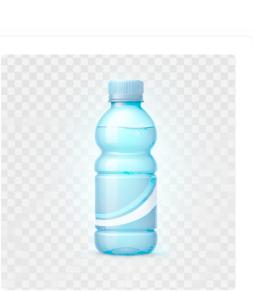
This red mop is designed for efficient floor cleaning. Ideal for all types of surfaces, it easily picks up dirt and spills. The mop head is highly absorbent and can be used wet or dry. Simply dip in water or cleaning solution, wring out excess, and mop your floors. The sturdy handle provides a comfortable grip, making cleaning a breeze.



watering-can

Similarity: 0.38806

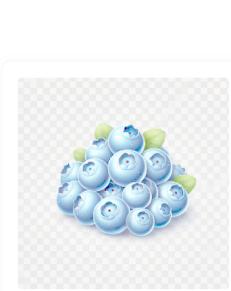
This blue watering can is perfect for your gardening needs. Its ergonomic design makes watering plants easy and efficient. With a large capacity, you can water multiple plants without frequent refills. The spout delivers a gentle, even flow to ensure plants receive the right amount of water. Ideal for indoor and outdoor use.



sports-drink

Similarity: 0.37068

This blue sports drink is designed to keep you hydrated and energized. Perfect for athletes and active individuals, it replenishes electrolytes lost during exercise. Enjoy it before, during, or after workouts to maintain peak performance. The refreshing flavor makes it a tasty way to stay hydrated.



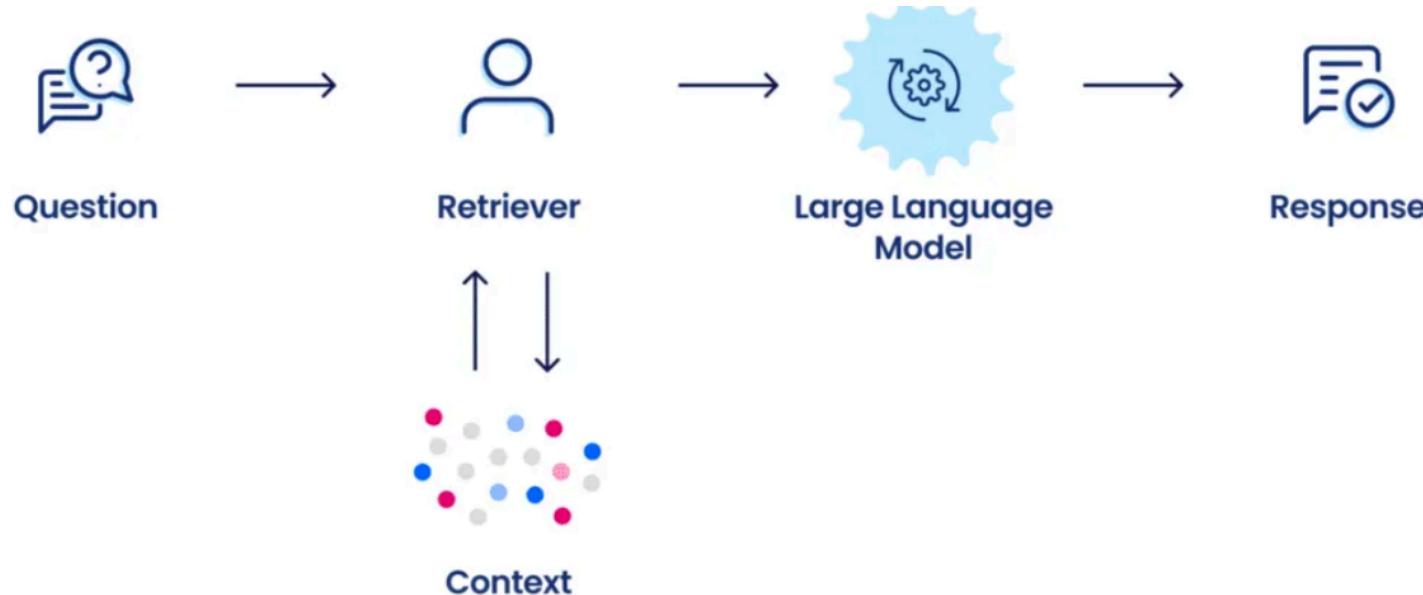
blueberries

Similarity: 0.28139

These fresh blue blueberries are perfect for snacking or baking. Rich in antioxidants and vitamins, they are a healthy choice. Enjoy them on their own, in smoothies, or as a topping for yogurt and cereals. Blueberries are versatile and add a burst of flavor to any dish.

3. Retrieval-Augmented Generation (RAG)

- **Concept:** Use embeddings to find relevant information for generative AI.
- **Barn Delights Example:**
 - Query: "How do I start a garden?"
 - RAG Response: "You'll need seeds, fertilizer, and tools like a trowel."



How Do You Build Embeddings?

1. Prepare the Data:

- Clean, structure, and categorize your data.

2. Select a Model:

- Example: `text-embedding-3-small` for text data.
- Example: `embed-english-v3.0` for text+images

3. Generate and Use:

- Generate embeddings for search, recommendations, and analytics.

4. Vector Database:

- Either in your own database or using a vector database like [Pinecone](#)

Challenges with Embeddings

- 1. Data Quality:** Poor data = poor embeddings.
- 2. Chunking:** Granularity of embeddings is important.
- 3. Generalization:** Terms specific to a domain don't differentiate enough
- 4. Scalability:** Embeddings for massive datasets are big (3-5x)
- 5. Bias:** Models can inherit biases from training data.

Future of AI Embeddings

- 1. Cross-Modal Embeddings:** Unify text, images, and audio.
- 2. Real-Time Applications:** Faster, more dynamic embeddings.
- 3. Enhanced Personalization:** Tailor recommendations with more precision.

Questions?

Thank You for Exploring Embeddings!

tSNE (map hyper dimensions to reduced)

