

Seven AI Patterns

Vienna is calling. Answer the damn phone

● Why short text breaks embeddings



Not enough signal

Two or three words gives the model almost nothing to embed. A 40-word paragraph produces a rich, specific vector. “Bob Smith” produces a vague one — it ends up near every other two-token name.



Surface-form variance

Smith / Smyth / Smithe / Смит. Robert / Bob / Bobby / Rob. Embeddings care about meaning, not about orthography, and there isn't much “meaning” in a proper noun.



Cultural & transliteration drift

Accents. Non-Latin scripts. “Müller” vs “Mueller.” “李娜” vs “Li Na.” These are the same person to a human and nearly unrelated to a vector.

● Give GenAI 5 seconds. Get a richer record.

STAGE 3



SYSTEM PROMPT · person

"Generate realistic name variants for use in a semantic search index. Always include: the original name, LastName-FirstName form, accent-free spelling, common nicknames or diminutives, transliteration variants if relevant. Return a plain text list — one variant per line."

Everything goes into the embedding. The record side is now fat with signal.

● Now do the same on the query. Many-to-many.

STAGE 4

ENRICHED RECORD

Bob Smith
Robert Smith
Smith, Robert
Bobby / Rob / Robb
Smyth / Smithe
ABC Technology
bob@abctech.com

MATCH

ENRICHED QUERY

Robert Smyth
Rob Smyth
R. Smyth
Smyth, Robert
robertt@abc-tech.com
237 More Park Dr, CA
Alphabet Tek?

Orthogonal variants meet in the middle. Typos, nicknames and transliterations fall on the same point in vector space.

● Same pattern, more context per record



ENTITY → COMPANIES

WHAT WE EMBED

- Original org name
- Abbreviations & acronyms
- With / without legal suffix
- Address — full and normalized
- List of staff names
- List of staff emails

REAL-WORLD MATCH

User types:

ABC Technology
237 More Park Drive, California

Finds:

Alphabet TekSystem
2337 Moorpark, Santa Monica CA

✓ *Prevents duplicate data entry.*

● Project names? Same playbook.



ENTITY → PROJECTS

PROJECT NAME → VARIANTS

Financial Audit Preparation Law Office

Monetary Audit Preparation Lawyer

Prepare for review of financial documentation

Accounting document analysis — Law Office



Your user typed “prep for law office audit.” They'll find it.

The script only enriches the project NAME field. That's deliberate — and leads straight into the next slide about what we do NOT embed.

● Semantically clever. Numerically blind.

THE HONEST LIMIT



Meaning, similarity, paraphrase

“law office audit” matches “Financial Audit Preparation Law Office” because the WORDS mean similar things.



Dates, amounts, ranges, exact strings

“Projects in Q2” vs a due-date of 2026-05-14 — a vector will not know those are the same thing. Numbers are not meanings.



Right tool rule:

semantic search for meaning + a plain WHERE clause for dates and amounts.

Stop trying to make one tool do both. That's the #1 mistake I see with new FileMaker AI work.

● Re-enrich and re-embed on a schedule

EXTRA CREDIT

Projects, products and companies change. So do their embeddings.



New staff? New price? New project description? The server re-enriches the text and re-embeds. Search stays accurate without a human touching anything.

(The Update Embedded Record script already does this — aim a Schedule at it.)

● Teach the index what “cheap” means



ENTITY → PRODUCTS

Users search “cheap red shoes.” The database stores \$39.99 / crimson / Nike.

PRICE TIERS

Shoes:

< \$50 → cheap, budget
\$50–150 → mid-range
> \$200 → luxury, premium

Cars:

< \$20k → basic, economy
> \$50k → luxury, flagship

COLOR FAMILIES

Red →

crimson, cherry,
cardinal, burgundy,
wine, garnet, brick,
tomato, rose, maroon

Silver → platinum,
ash, pewter, titanium

CATEGORY SYNONYMS

Shoes →

sneakers, trainers,
kicks, running shoes

SUV →

crossover, 4x4,
sport utility

“Cheap” is not a price. It's a category-relative judgment. Bake that in at write-time.

● The cost story: basically free

~100

tokens in

the search string + a compact system
prompt

~200

tokens out

5–10 name or query variants — that's
it

small

model OK

gpt-4o-mini, haiku, local models. No
frontier needed.



In practice: false positives are rare.

Real-world noise (typos, nicknames, abbreviations) is what we were after. The enrichment surfaces genuine overlap — not random neighbors.

● Where a CS purist will push back



“Use hybrid search”

Production systems often combine BM25 (lexical) + semantic + structured filters, merged with rank fusion. Our Stage 3+4 is a strong subset — adding a keyword pass for SKUs is cheap and worth doing when you have exact-token fields.



“Your index goes stale”

When data changes and you don't re-enrich, search accuracy drifts. That's why the scheduled re-embed matters. Build it in from day one.



“Price-tier text is baked in”

If pricing strategy shifts (“now \$150 is mid-range for shoes”), the existing embeddings encode the old worldview. Re-enrich the catalog. Not the end of the world — just a scheduled job.

All of these are solvable. None of them invalidate the approach.

QUESTIONS?

Build the search your users think they have.

REACH FOR IT WHEN

- You store short-text entities
- Users type sloppy searches
- Duplicate prevention matters

DON'T REACH FOR IT WHEN

- You need exact-match by ID
- Data is numeric or date-based
- A WHERE clause does the job