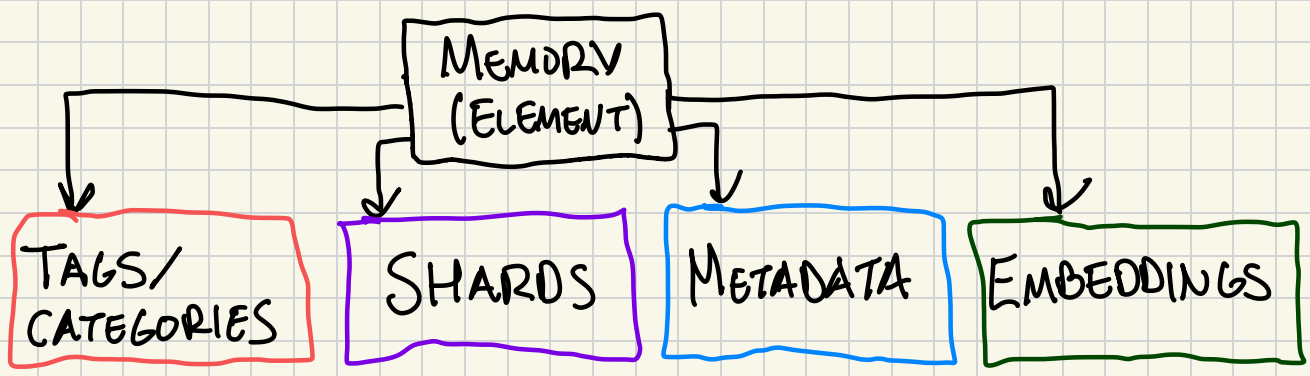
Updated Memory Schema
May 4 2025

A SINGLE UNIT OF CONTEXT
CAN BE RAM OR LONG-TERM



TAGS: Human Readable filters used for labeling

→ TYPES: Pre-defined (system categories like "career" or "relationship")
User-defined (freeform tags manually entered/ created by user)
System-suggested (Use LLM to suggest tags like "imposter syndrome")

→ USED FOR: Symbolic filtering, often first pass of retrieval filter

```
"tags": [
  { "label": "burnout", "type": "user" },
  { "label": "career", "type": "system" }
]
```

SHARDS: Groupings across time, meaning, etc. for retrieval and context weaving

→ TYPES: Semantic, emotional, Personas
temporal, conversational, Event-based

→ EACH MEMORY CAN BELONG TO MANY SHARDS

MEMORY ↔ SHARD1, SHARD2, SHARD3

METADATA: Used for internal state tracking. INCLUDING RAM

```
"metadata": {
  "created_at": "2025-05-03T18:01:00Z",
  "source": "manual_entry",
  "author": "user",
  "pinned": false
}
```

Created/updated at
Source: manual, import_gpt4, chat_context
Author: user, ai_agent, llm
pinned: yes OR No
→ RAM PROMOTION

FOR RAM
→ Add pin_reason
"pin_reason": "user-pin"
→ Add expires_at?

EMBEDDINGS: Core of semantic searching and Clustering

```
"embedding": {
  "vector": [0.123, 0.456, ...],
  "model": "text-embedding-3-small-v1"
}
```

→ Stored as a VECTOR
→ Includes MODEL VERSION of model that generated the embeddings

MongoDB Stores "embedding_id"
For instance "vector_entry_1"

You generate an embedding for the memory → Add and store that embedding in a vector DB

embedding = embed("I like yogurt")

```
{
  "_id": "memory_001",
  "text": "I feel stuck in my career.",
  "tags": ["career", "stuck"],
  "metadata": {...},
  "embedding_id": "vec_001"
}
```

MONGODB

```
{
  "id": "vec_001",
  "vector": [0.123, -0.045, ..., 0.021],
  "metadata": {
    "memory_id": "memory_001",
    "tags": ["career"]
  }
}
```

VECTORDB