

Building a Customer Persona Digital Twin

Simulating a Trend-Conscious, Sustainable Shopper Using RAG



Fashion Brands Struggle to Connect with Niche Customer Segments

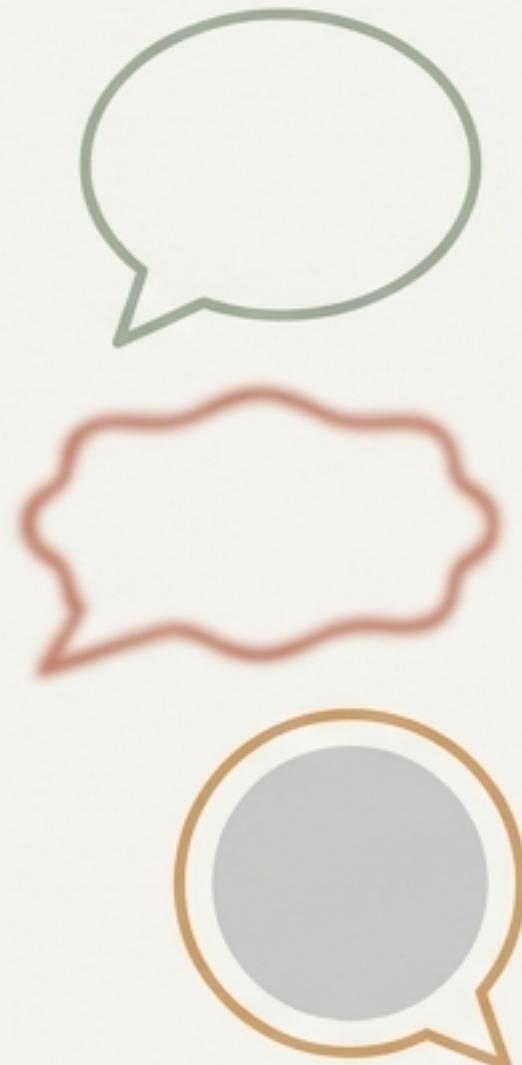
The Challenge

Brands need to test how sustainability-aware customers will react to new messaging and styling advice, but traditional methods are slow and expensive.

The Problem with Generic LLMs

- Polished but Impersonal: They provide generic, 'stylist textbook' advice that lacks a genuine customer voice.
- Prone to Hallucination: They often invent trends or misrepresent brand values, making them unreliable for market research.
- Lack Nuance: They fail to capture the complex, real-world tradeoffs customers make between trendiness, sustainability, and practicality.

Our Goal: Build a persona-level digital twin that mimics how a real customer segment thinks, speaks, and decides.



Meet Our Persona: The Trend-Conscious Sustainable Shopper



Who She Is

A fashion-forward but practical individual, influenced by social media trends but actively avoids the excess of fast fashion.



Core Traits

- Values curated, confident recommendations.
- Prioritizes 'wearable' translations of current trends.
- Seeks versatility and longevity in her wardrobe.



Core Values

- **Materials:** Prefers natural fibers and ethically produced goods.
- **Habits:** Believes in capsule wardrobes, cost-per-wear, and thrifting.
- **Mindset:** A strong anti-fast-fashion bias guides her purchasing decisions.

Grounding the Persona in Authentic Customer Voices

To create a realistic persona, we can't rely on the LLM's generic knowledge. We need to ground it in real-world conversations and opinions. This is achieved with Retrieval-Augmented Generation (RAG).

Source 1: Reddit



Subreddit: r/femalefashionadvice

Why it was chosen: Captures the authentic, unfiltered language of real users. It reveals practical tradeoffs, detailed sustainability discussions, and genuine brand perceptions.

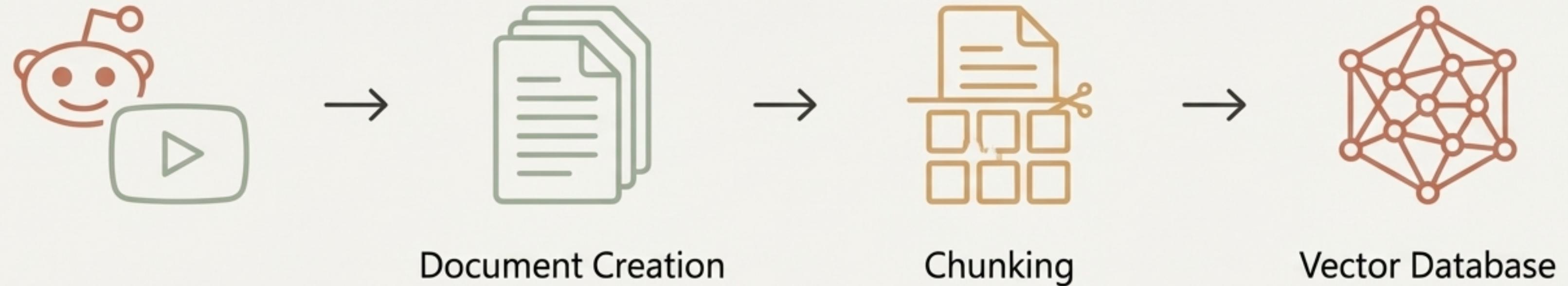
Source 2: YouTube



Source Type: Fashion & seasonal trend analysis videos.

Why it was chosen: Adds high-level trend interpretation and seasonal fashion commentary. Audio was transcribed to text to capture expert but accessible style guidance.

Transforming Raw Data into Retrievable Knowledge



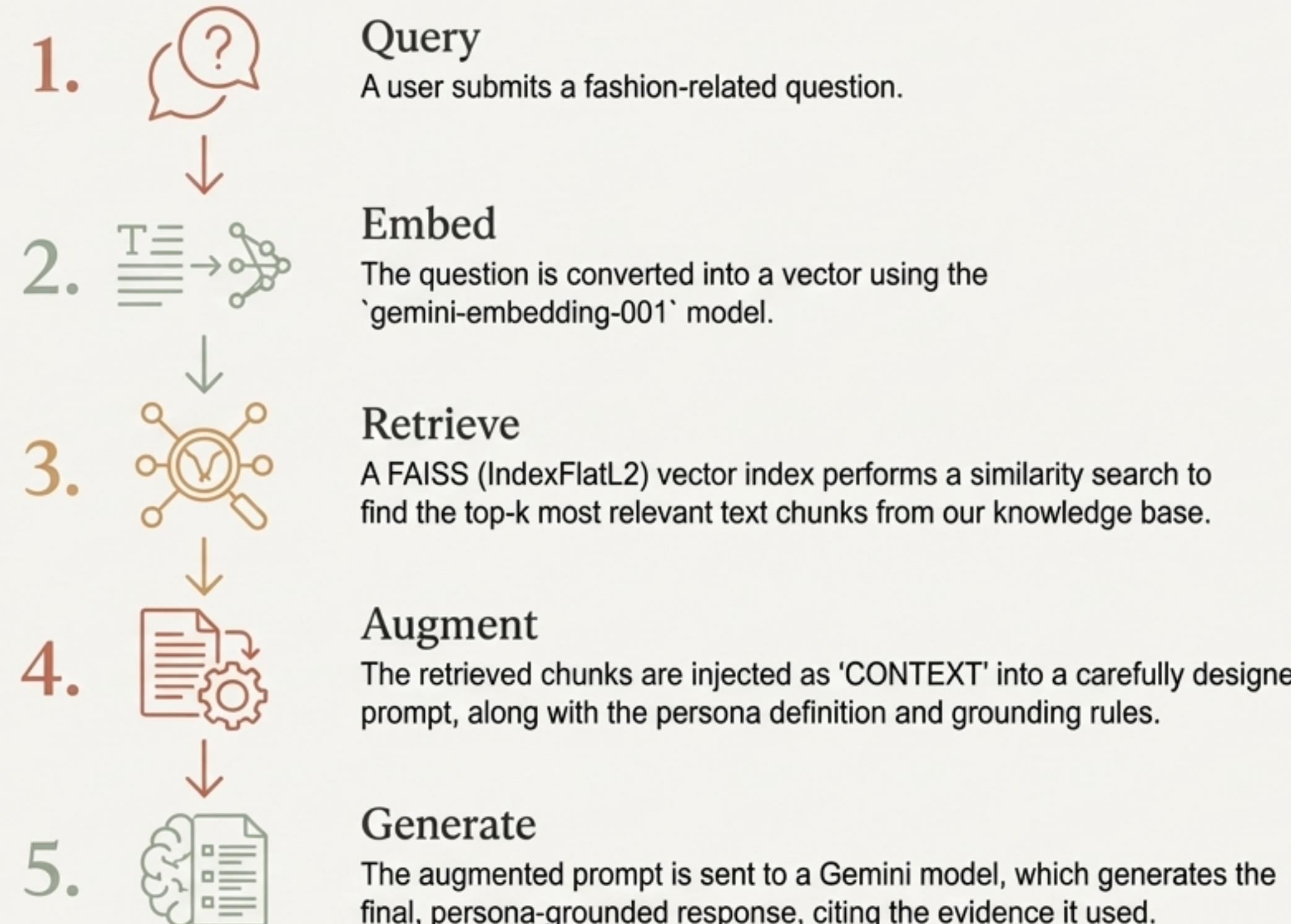
Process

1. **Document Creation**: Reddit posts and their top comments were merged into single documents. YouTube audio was transcribed into text documents.
2. **Chunking Strategy**: All documents were split into smaller, uniform chunks of approximately 250 words.
3. **Metadata Tagging**: Each chunk was tagged with its source (Reddit/YouTube), a unique Chunk ID, and a reference URL or post ID for traceability.

The Rationale for Chunking

- **Improves Retrieval Precision**: Smaller chunks allow the system to find more specific and relevant information for a given query.
- **Optimizes LLM Context**: Keeps the information fed to the LLM concise and efficient, preventing context overload.

The RAG System Architecture



Engineering the Persona's Voice and Behavior

The prompt is more than just an instruction; it's the persona's **constitution**. We iterated to move from a simple instruction to a sophisticated mimicry and testing tool.



Prompt v1: Initial Persona & Structure

- Defined the persona: "fashion-forward + sustainability-aware."
- Included a core grounding rule: "Use ONLY the CONTEXT."
- Enforced a structured output (Style Direction, Outfit Formulas, etc.).
- Cited chunk IDs as "signals."



Prompt v2: Enhanced "Mimicry" & Business Value

- **Stronger Voice:** Added constraints like "concise, confident, trend-aware."
- **Explicit Bias:** Instructed the model to avoid fast fashion and synthetic-heavy hype.
- **Added Business Use Case:** Included a "Messaging Test" section to generate marketing lines that would appeal to the persona.
- **Required Evidence:** Mandated an "Evidence Used" list for transparency.

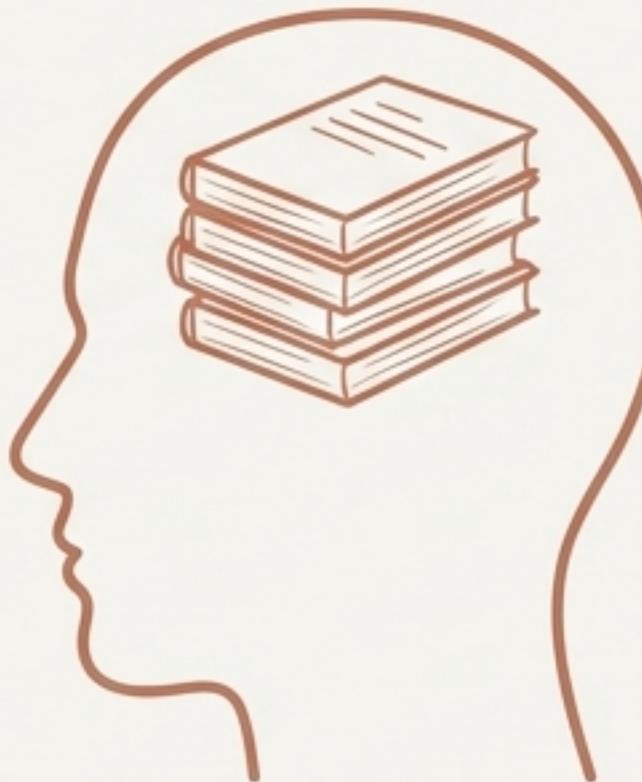
The Showdown: Proving the Value of Grounded Generation

To measure success, we conducted a head-to-head comparison to demonstrate the tangible improvements provided by our RAG pipeline.



Baseline LLM

VS.



RAG Persona Twin

The Contenders

- **Baseline LLM**: Gemini answering a user's query directly, using only its pre-trained knowledge.
- **RAG Persona Twin**: Our system, answering the same query but with its response grounded in the retrieved Reddit and YouTube context.

The Goal

To show that the RAG twin provides superior persona realism, grounding, and trustworthiness.

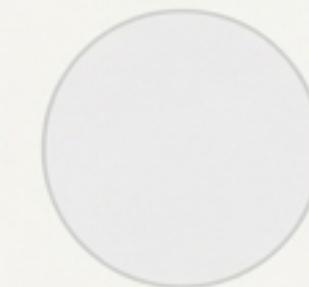
Defining ‘Good’: A Rubric for Persona-Mimicry

We developed a multi-dimensional rubric to assess the quality of the generated responses against our project goals.

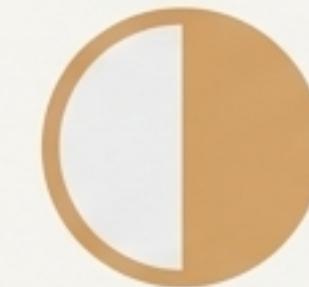
Evaluation Dimensions

- **Persona Match:** *Does it sound like our persona?* (Tone, priorities, anti-fast-fashion stance).
- **Groundedness:** *Is it based on real data?* (Claims supported by retrieved chunks; refuses to answer when info is missing).
- **Actionability:** *Can a user act on this advice?* (Provides specific outfit formulas, materials, next steps).
- **Sustainability Alignment:** *Does it align with sustainability values?* (Mentions natural fibers, thrifting, cost-per-wear).

Scoring Scale



0: Poor / Missing



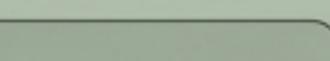
1: Partial / Generic



2: Strong / Clearly meets goal

Evaluation Results: RAG Delivers a More Authentic and Reliable Persona

Query: “How would you style for fall if you want to be fashion-forward but sustainable?”

Model	Persona Match	Groundedness	Actionability	Sustainability	Total Score
Baseline LLM	1	0	2	2	5 / 8 
RAG Persona Twin	2	2	2	2	8 / 8 

Key Finding*: The RAG Persona Twin achieves a perfect score by excelling in Persona Match and Groundedness, the two areas where the Baseline model fails completely.

The Story Behind the Scores

The baseline model provides generally helpful and actionable fashion advice, but it lacks grounding in my specific customer persona and does not cite or constrain itself to any real user data. In contrast, the RAG-powered persona twin closely matches the defined ‘Trend-Conscious Sustainable Shopper,’ consistently emphasizing **capsule wardrobes, natural fibers, and cost-per-wear**. The RAG model is also more grounded, explicitly citing Reddit and YouTube sources and refusing to answer when information is not present in the context, which improves trust and reliability.

A Feature, Not a Bug: Principled Refusal Builds Trust

We tested the system's reliability with a query it could not answer from its knowledge base.

Query: "What are the BEST sustainable shoe brands for fall 2025?"

Baseline LLM



Hallucinates plausible-sounding but unverified brands. **(Groundedness = 0)**

RAG Persona Twin

Information on 2025 trends is not available in my knowledge base.



Correctly refuses to answer, stating that information on 2025 trends and brands is not available in its context. **(Groundedness = 2)**

The Insight

The RAG model's refusal is a sign of a well-engineered, trustworthy system. It avoids making up information, which is critical for a business use case where accuracy is paramount. This behavior is safer and more reliable.

Limitations and a Vision for the Future

⊖ Current Limitations

- **Dataset Size'**: The knowledge base is built from a small set of Reddit threads and YouTube videos.
- **Simple Chunking'**: A word-based strategy may occasionally split concepts awkwardly.
- **Strict Grounding'**: The 'Use ONLY context' rule, while preventing hallucination, can sometimes lead to overly conservative answers.

⊕ Future Improvements

- **Expanded Data Sources'**: Incorporate ethical fashion blogs and official brand sustainability reports.
- **Advanced Retrieval'**: Implement a reranker to improve the quality of retrieved chunks.
- **Scaled Evaluation'**: Develop an automated scoring harness to test the system across dozens of queries.
- **Dynamic Personas'**: Introduce 'sliders' to adjust the persona's priorities (e.g., budget, climate, workplace formality).

Key Takeaways



1. RAG is essential for creating authentic, persona-accurate digital twins. By grounding generation in real-world data, we move beyond generic responses to genuine mimicry.



2. Grounding directly improves trust and reduces hallucination. Citing evidence and refusing to answer unanswerable questions are hallmarks of a reliable AI system.



3. This persona-twin approach has immediate business value. It can be used as a powerful tool for:

- Testing marketing messaging and ad copy.
- Analyzing emerging consumer trends.
- Developing sustainable fashion strategy and product lines.