DATASTAX THE AI CONFERENCE

The Power of Vision Models in Building Image Recognition Apps

About Me

Melissa Herrera

- Developer Relations Engineer @DataStax
  - previously Lead Solutions Engineer
  - Focused on Python
- Tech Content and Gen AI Junkie
- Loves iced matcha lattes, concerts, and dancing

Let's Connect!

Agenda

What are Vision Models?

Fashion Buddy Demo

Overview of what a vision model is and does

Leveraging Vision Model Output as Data

Breaking down output of vision model

End to end diagram

Power of Vision Models in Gen AI

Applications and benefits

Complex tasks that use Vision Models

What are Vision Models?

Quick Demo!

Fashion Buddy 🛍️👗

Scan and try for yourself!

Fashion Buddy Demo

- How is this actually being done?

Fashion Buddy Demo

- How is this actually being done?
- What if I told you this was relatively “easy” to build?

Fashion Buddy Demo

- How is this actually being done?
- What if I told you this was relatively “easy” to build?
- What if I told you this is all being done with only 3 key things:
  - Vector Embeddings
  - Vector Database
  - Vision Model

What are Vision Models?

What are Vision Models?

What are Vision Models?

+

Concept of vision

What are Vision Models?

Computer Vision

- LLMs trained on text -> Vision Models trained on visual inputs (digital images and videos)

What are Vision Models?

Computer Vision

- LLMs trained on text -> Vision Models trained on visual inputs (digital images and videos)
- Allows machines to “see” and understand visual data

What are Vision Models?

Computer Vision

- LLMs trained on text -> Vision Models trained on visual inputs (digital images and videos)
- Allows machines to “see” and understand visual data
- Model Types:
  - Generative Adversarial Networks (GANs)
  - Diffusion Models
  - Large Vision Models (LVMs)

Generative Adversarial Networks (GANs)

- Consists of two competing neural networks: generator and discriminator
- Generator creates fake images, discriminator tries to distinguish real from fake
- Used to create highly realistic synthetic images

EXAMPLES:
NVIDIA Style GAN3 generates highly realistic faces, Pix2Pix for image-to-image translation tasks

Diffusion Models

- Works by gradually adding noise to images, then learning to reverse this process
- Creates high-quality images from random noise
- Can also handle image synthesis tasks
- Also helpful for image editing, inpanting, super resolution

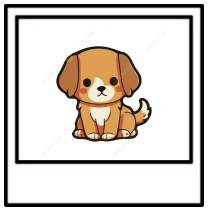
EXAMPLES:
OpenAI's DALL-E2 for text to image generation, Google's Imagen for high-fidelity generation, Stable Diffusion

Large Vision Models (LVMs)

- Just like LLMs, trained on vast image datasets
- Performs various tasks like object detection, image classification
- Adaptable to different domains with additional training

EXAMPLES:
Google Gemini Pro Vision, OpenAI CLIP, Anthropic Claude

Understanding Vision Models

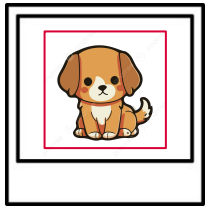


Input
Image/Video

```
Describe the
object in the
photo as best
as possible.
```

Prompt

Vision Model



Identifies object
in image

```
The photo shows a
cartoon dog. It is
sitting down with
its legs crossed.
The dog is brown
and white with big
brown eyes. It has
a short, curled
tail. The dog is
drawn in a simple,
cute style. The
background is
white.
```

Generated
Output

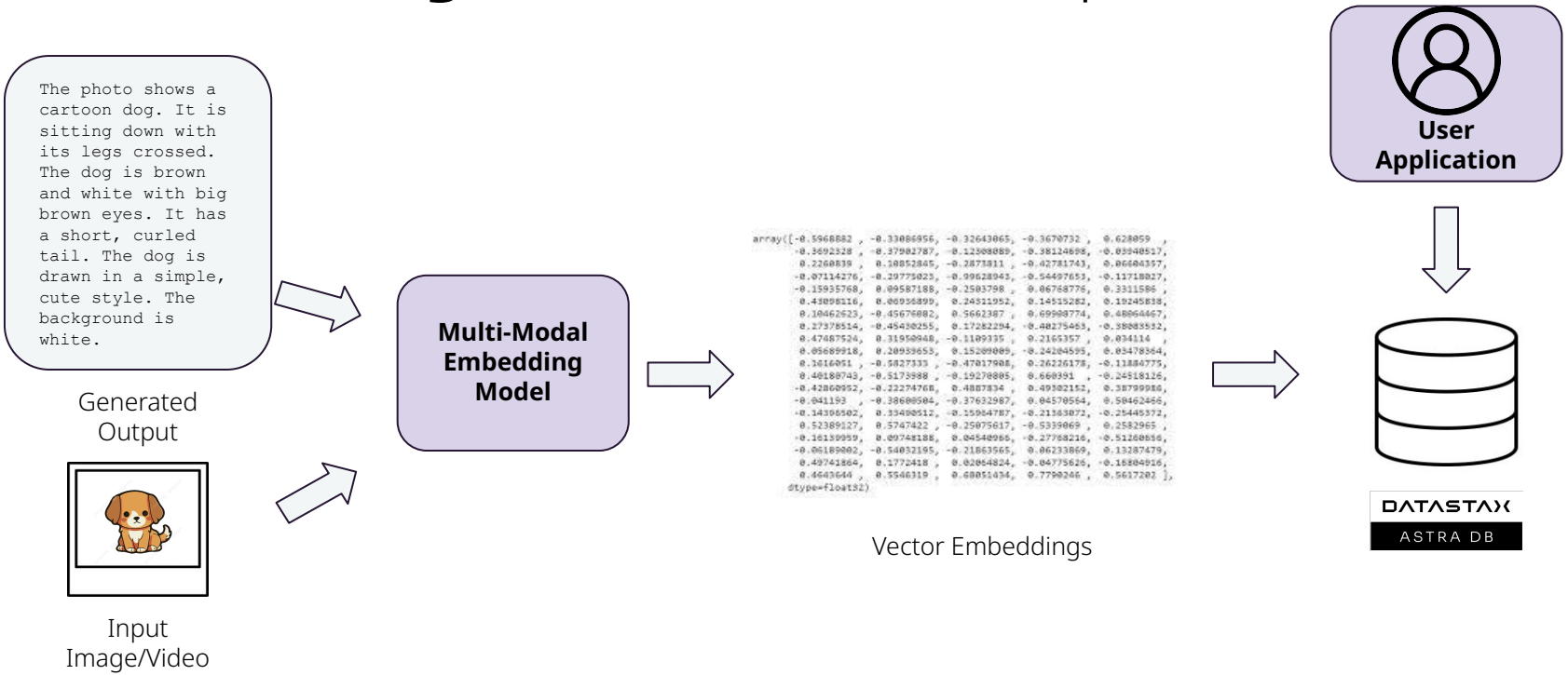
Let’s see this in action!

Leveraging Vision Model Output as Data

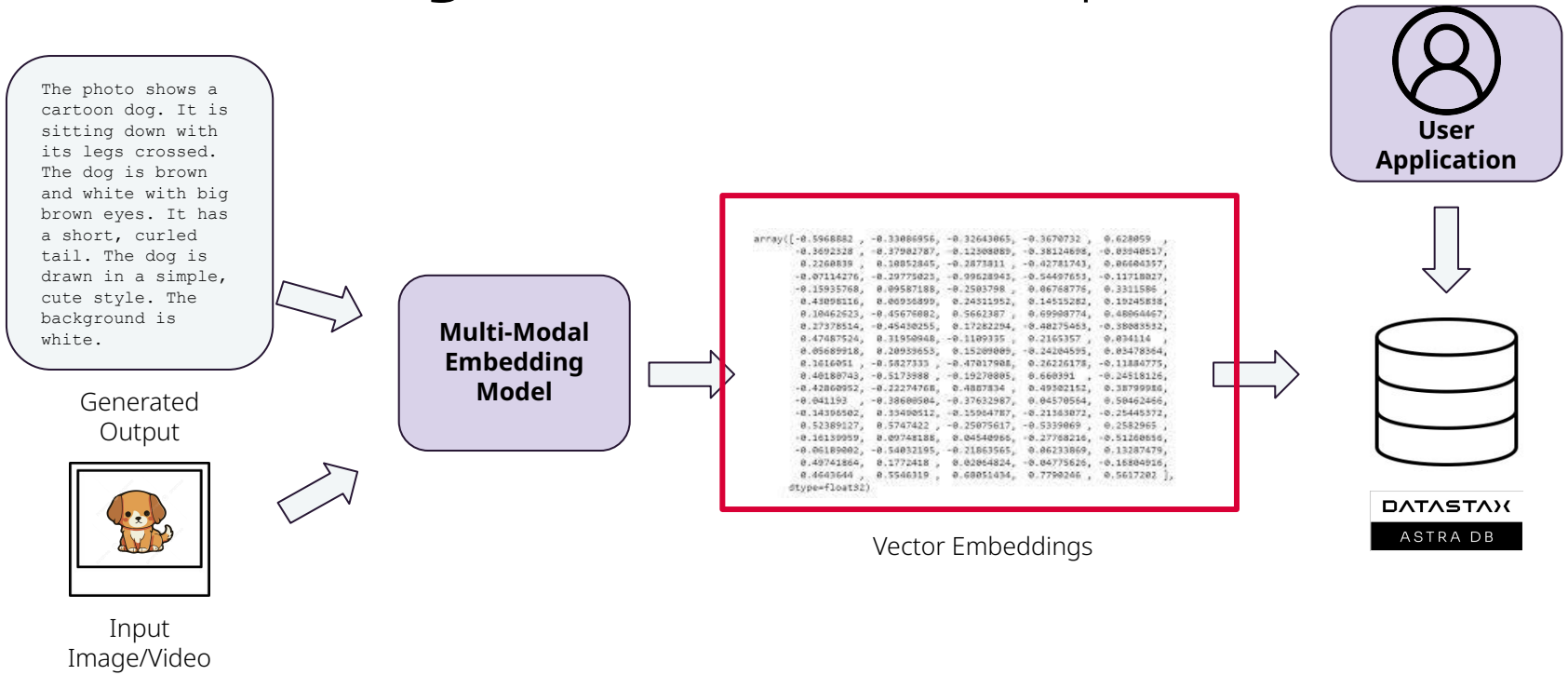
I have an output…now what?

Let’s turn this into data that we can use!

Converting Vision Model Outputs

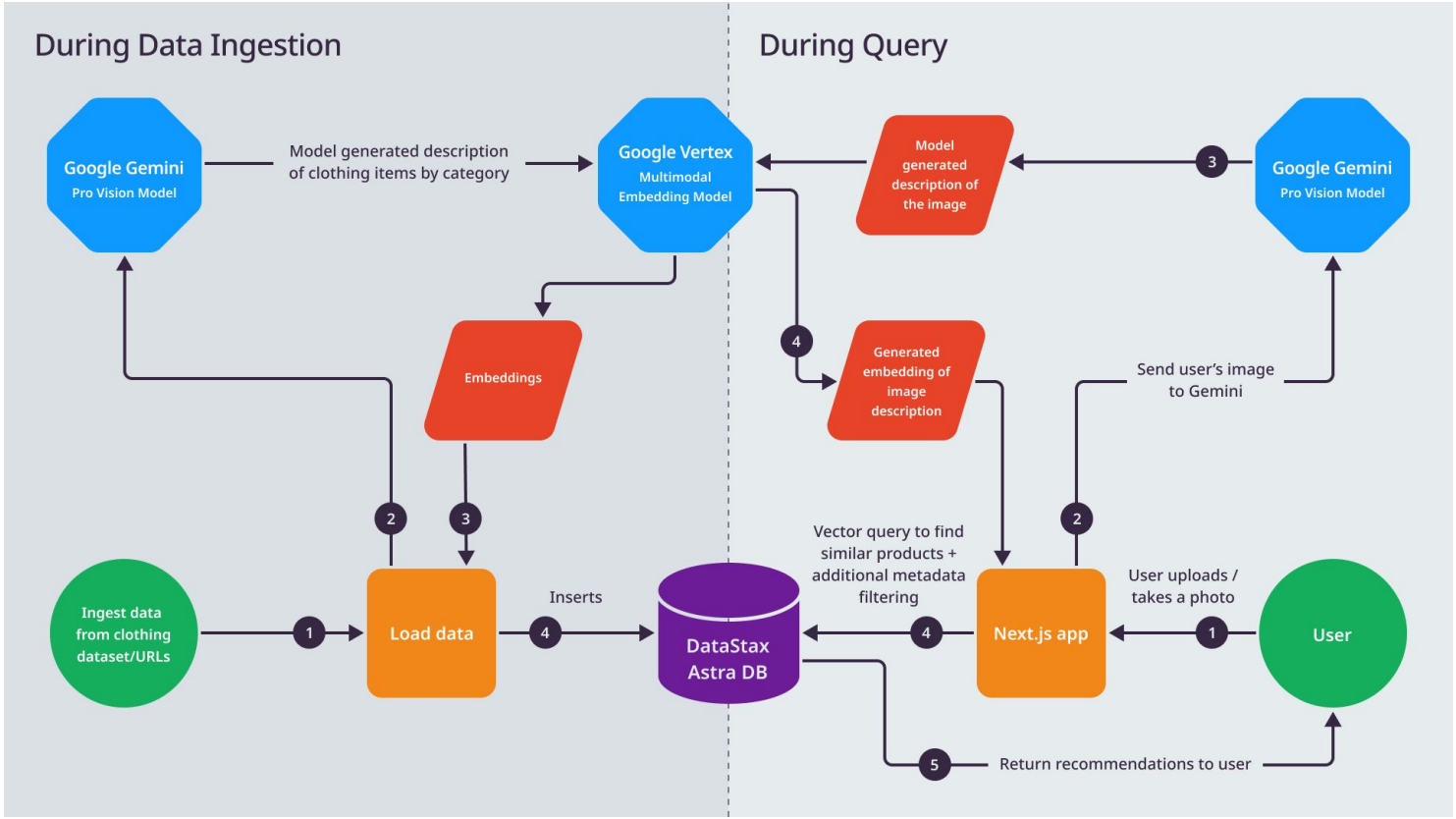


Converting Vision Model Outputs

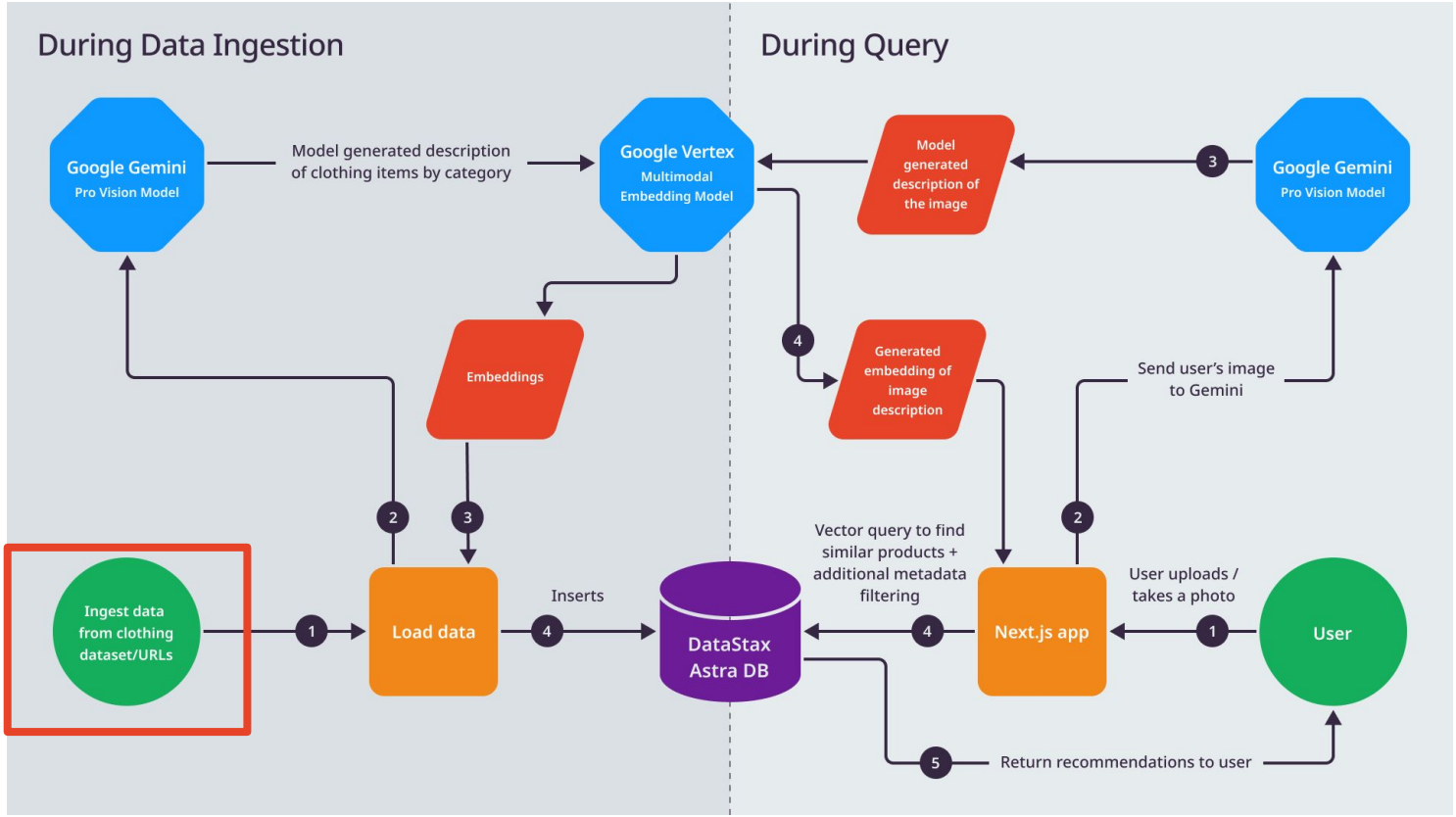


Let’s see this data in Astra DB

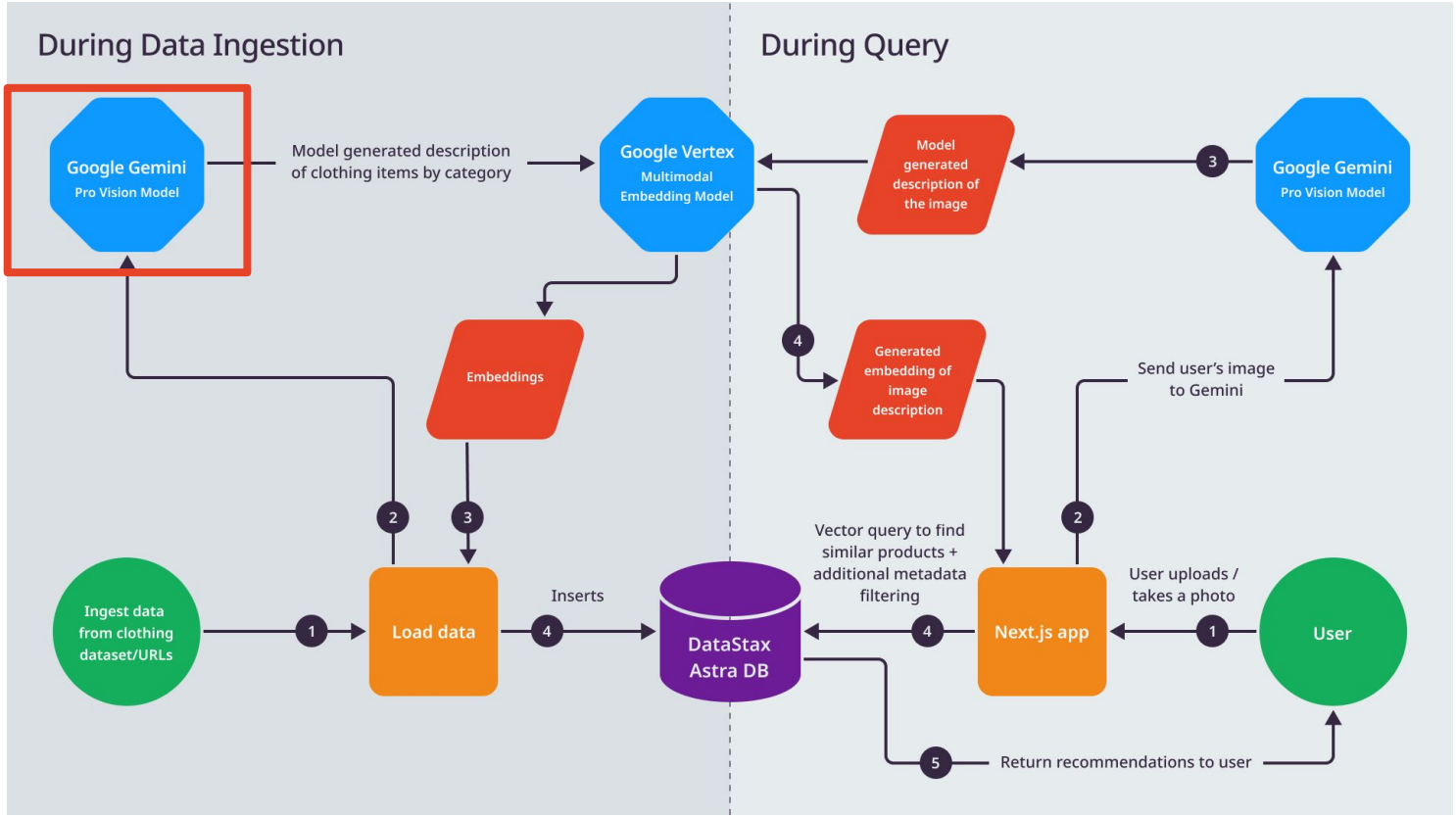
Fashion Buddy Diagram



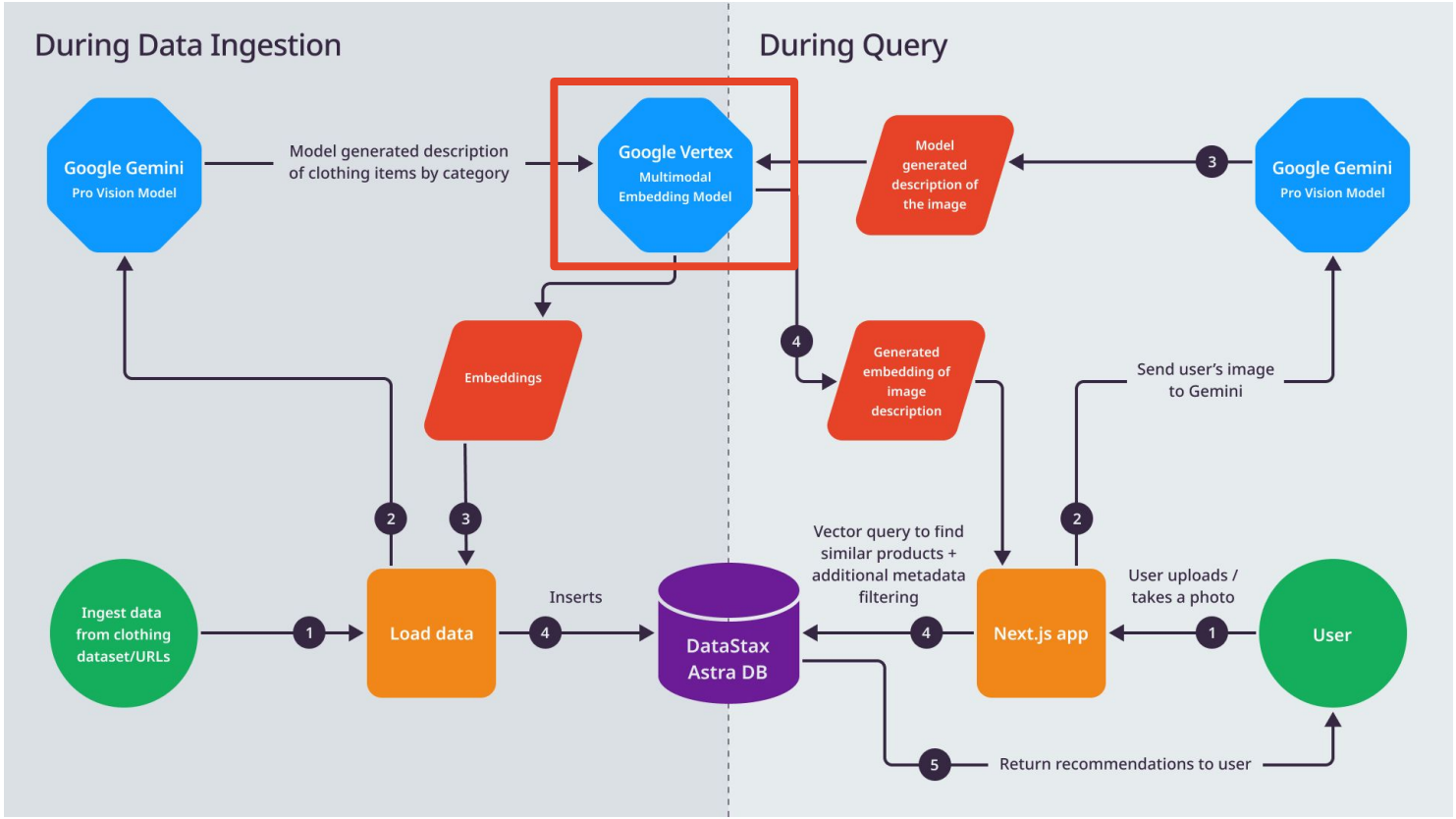
Fashion Buddy Diagram



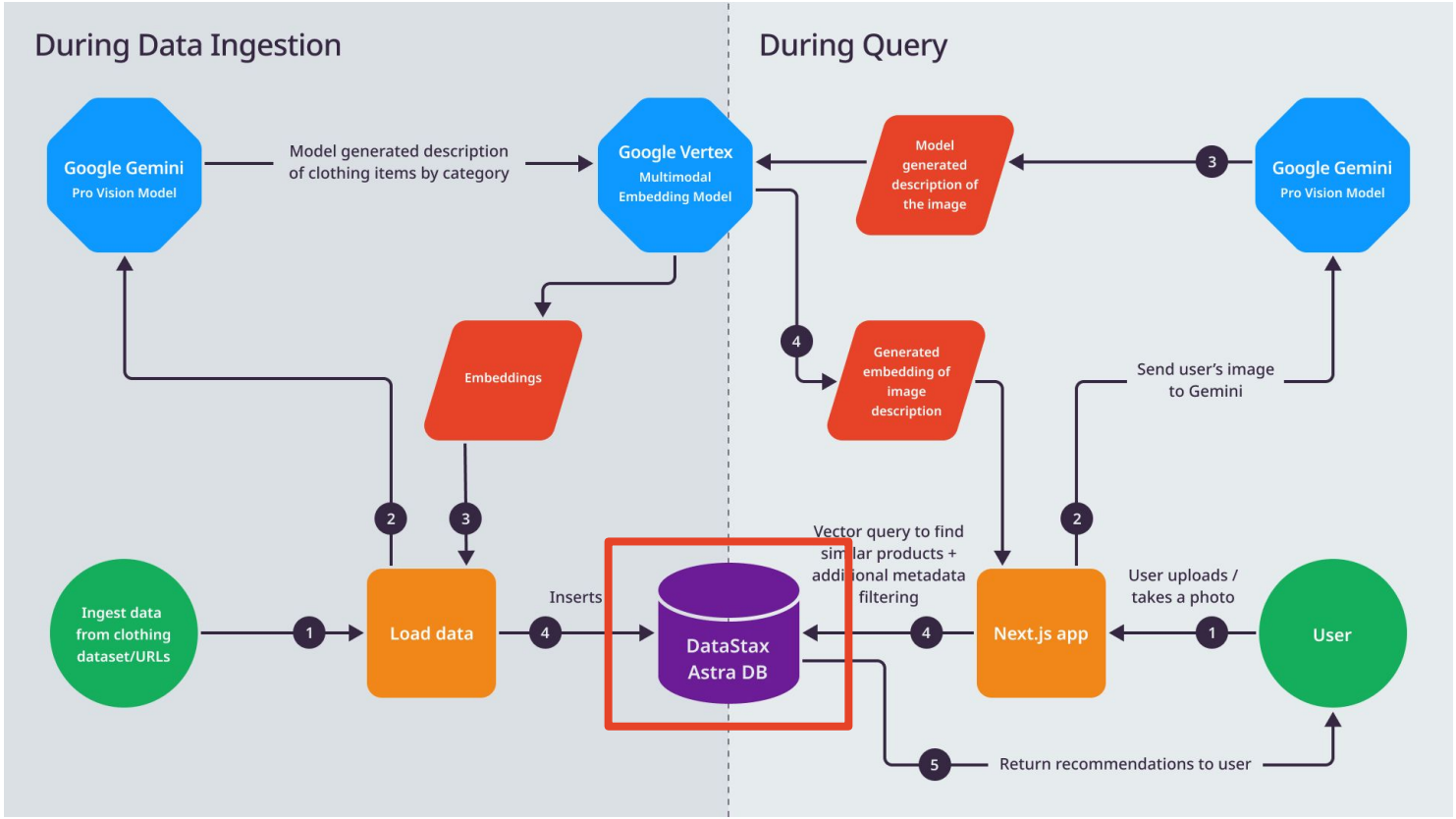
Fashion Buddy Diagram



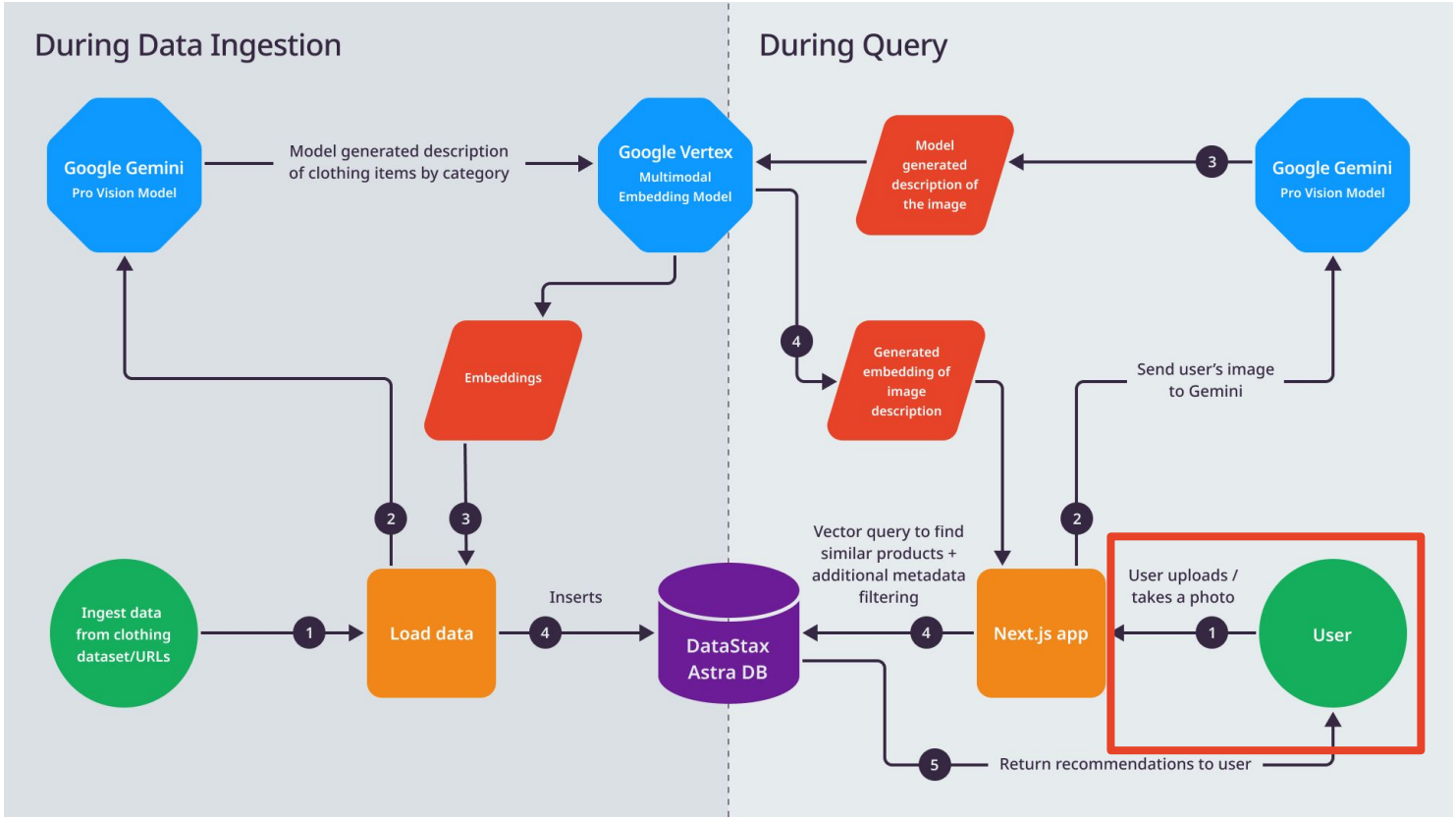
Fashion Buddy Diagram



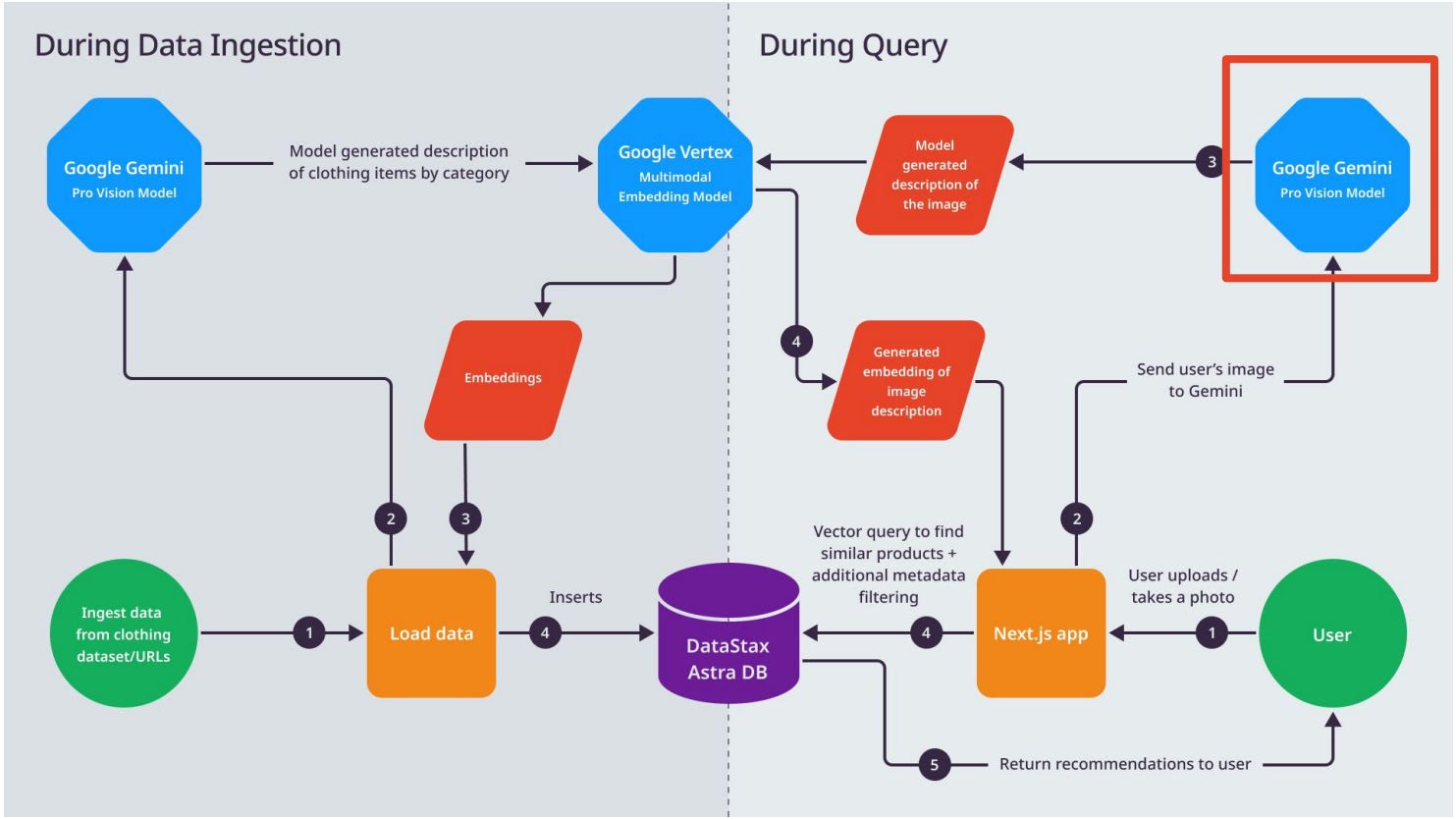
Fashion Buddy Diagram



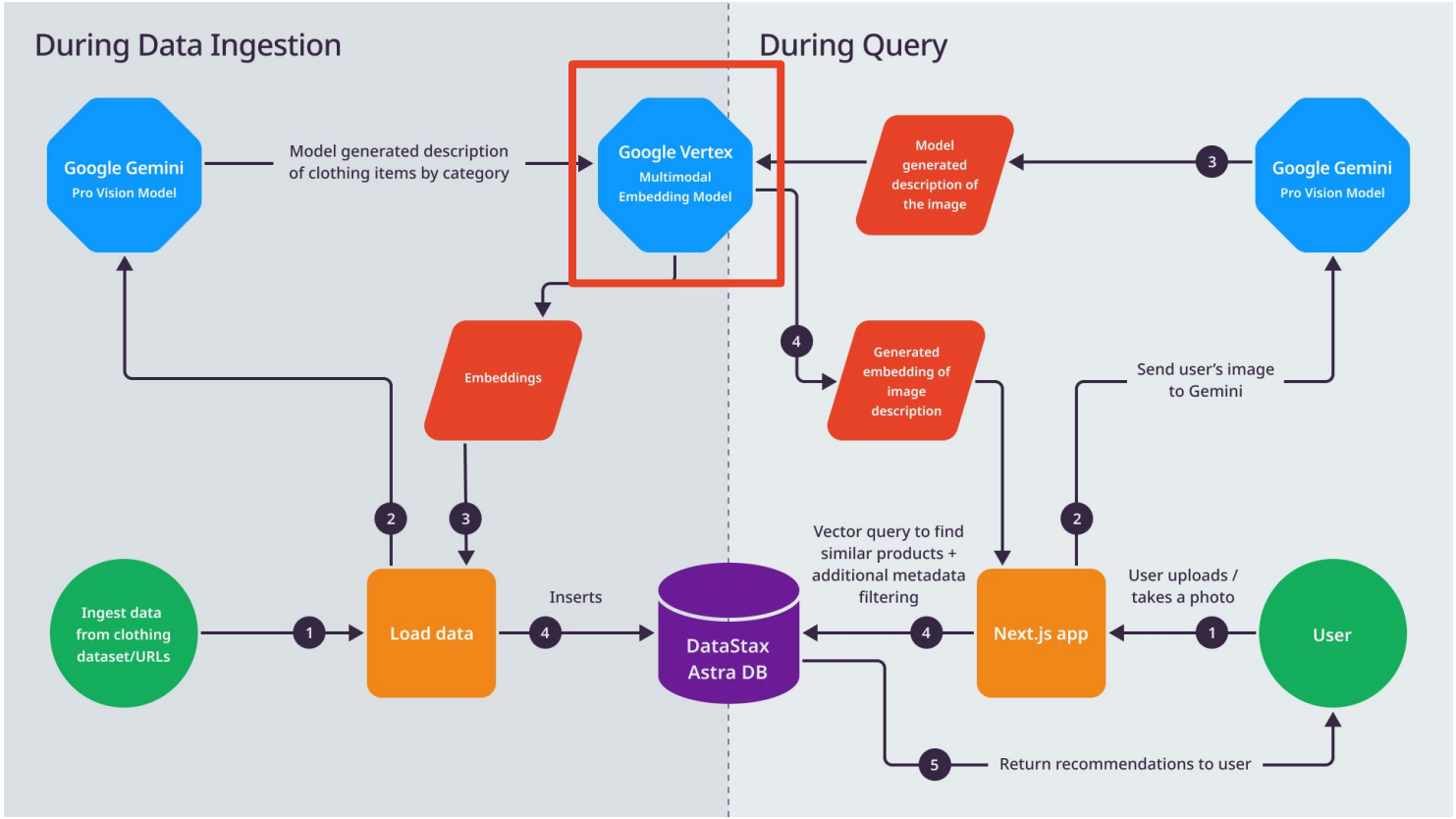
Fashion Buddy Diagram



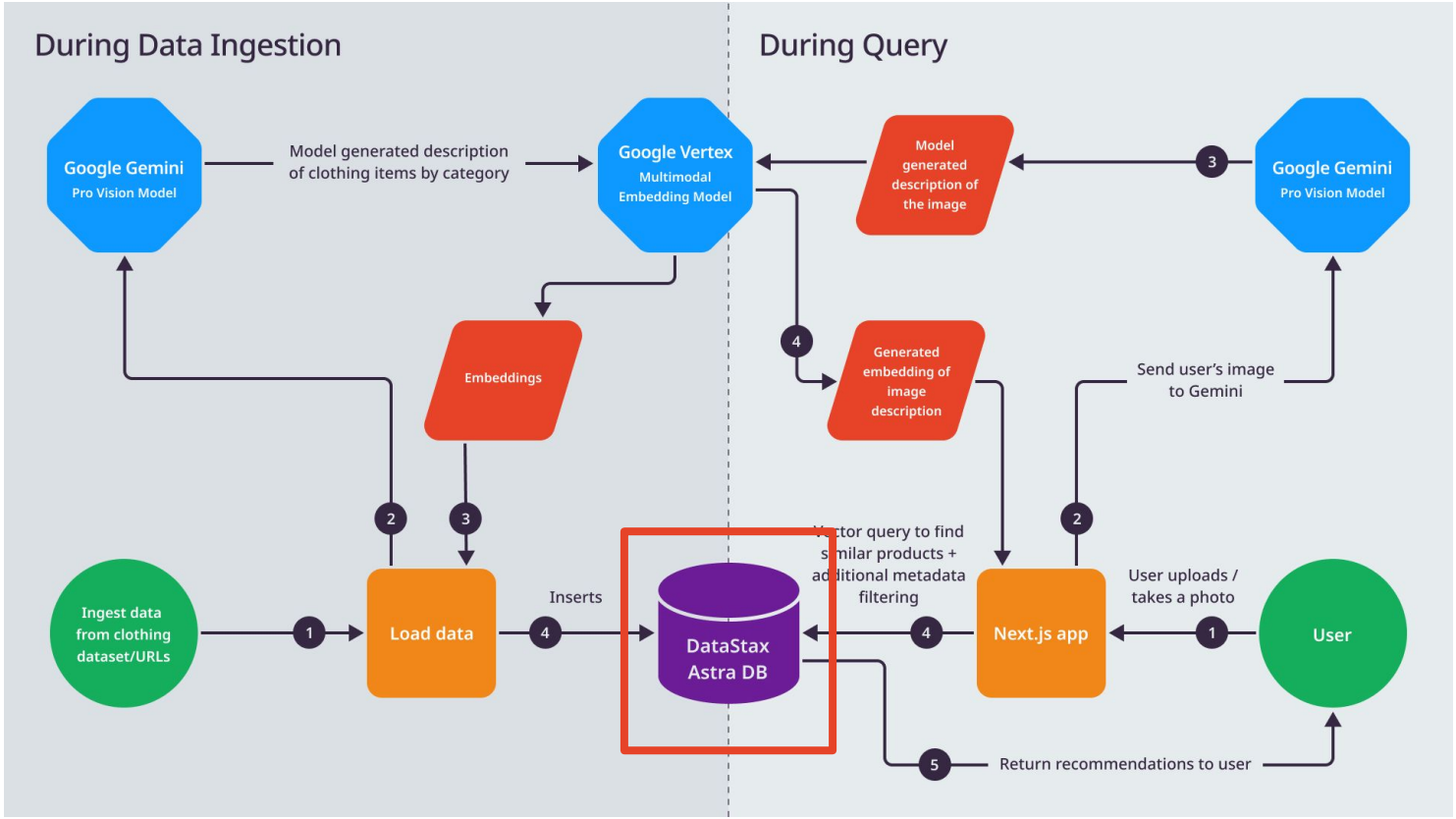
Fashion Buddy Diagram



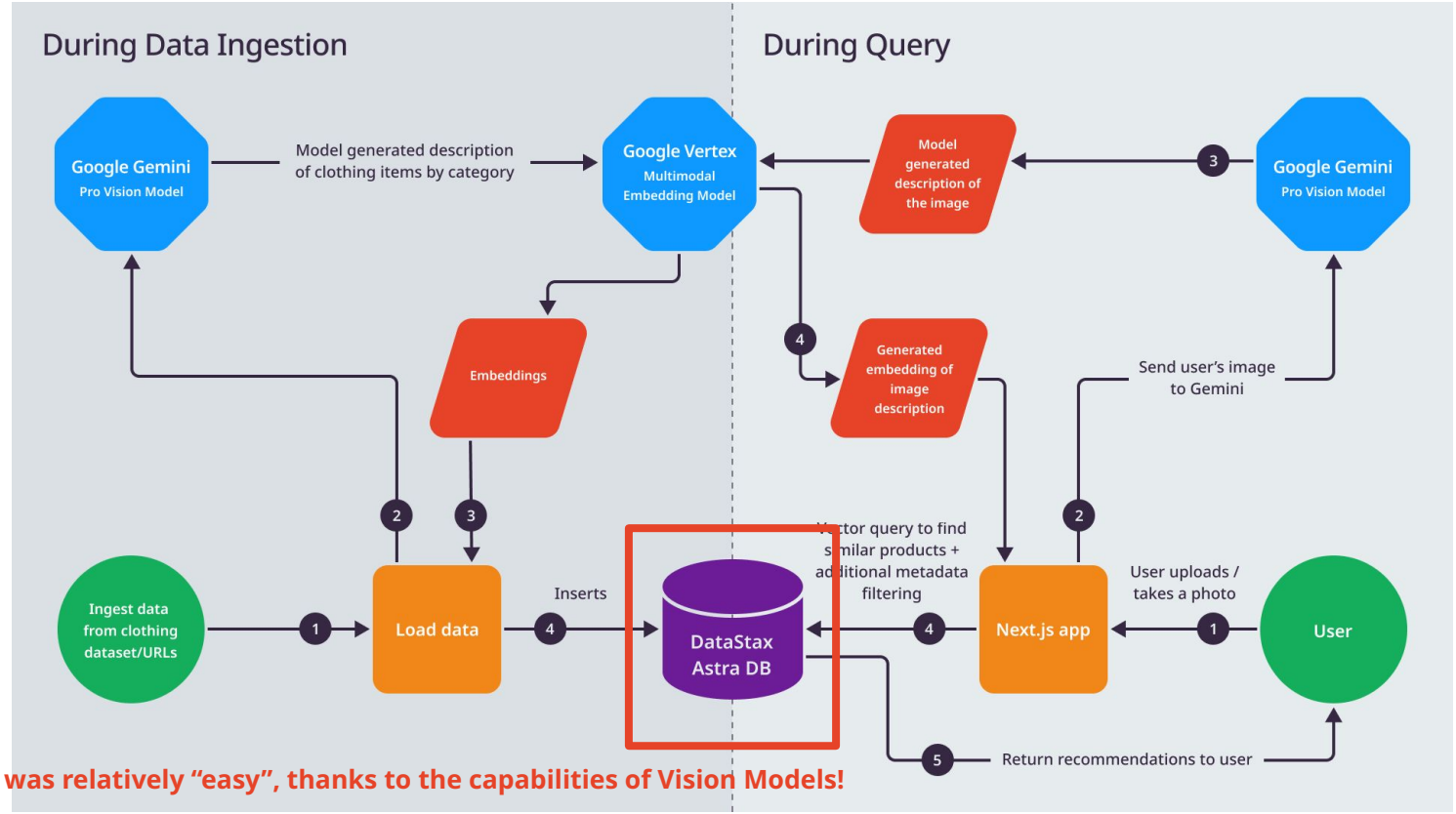
Fashion Buddy Diagram



Fashion Buddy Diagram



Fashion Buddy Diagram

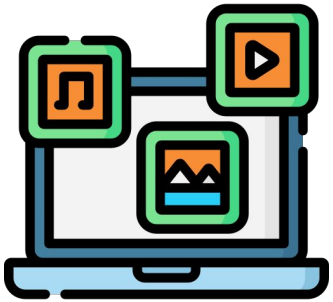


This was relatively "easy", thanks to the capabilities of Vision Models!

Let’s look at some code

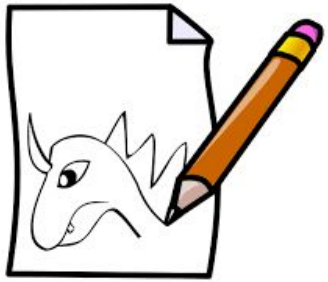
Power of Vision Models in Gen AI

Applications and Benefits



Apply intelligence to unstructured data

Visual Inspection

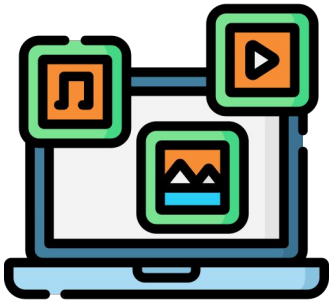


Generative Capabilities

...And more!

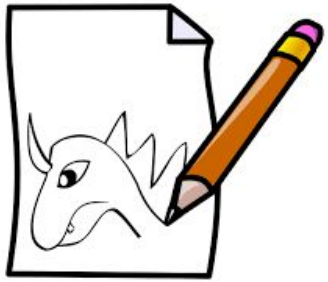
Image and Video Editing

Applications and Benefits



Apply intelligence to unstructured data

Visual Inspection



Generative Capabilities

...And more!

Image and Video Editing

Vision Models tackle complex tasks in Generative AI so that we don't need to...it's like MAGIC!

The Possibilities Are Endless!

Existing/Common Use Cases

- Image Search (Currently being done with Google Lens, Amazon, etc)
- Facial Recognition
- AI-Assisted Creativity (Adobe Photoshop)

Non-Traditional Use Cases

- Makeup Color Matching (prototype built)
- Architectural Work
- Room/Interior Design, Artistic Transformation
- Environmental Protection/Detecting Ecosystem Changes
- Historical Artifact reconstruction

The Possibilities Are Endless!

Existing/Common Use Cases

- Image Search (Currently being done with Google Lens, Amazon, etc)
- Facial Recognition
- AI-Assisted Creativity (Adobe Photoshop)

This was already being done but a with a full TEAMS to back it….now with Vision Models even smaller devs can tackle these tasks.

Non-Traditional Use Cases

- Makeup Color Matching (prototype built)
- Architectural Work
- Room/Interior Design, Artistic Transformation
- Environmental Protection/Detecting Ecosystem Changes
- Historical Artifact reconstruction

The Possibilities Are Endless!

Existing/Common Use Cases

- Image Search (Currently being done with Google Lens, Amazon, etc)
- Facial Recognition
- AI-Assisted Creativity (Adobe Photoshop)

Non-Traditional Use Cases

- Makeup Color Matching (prototype built)
- Architectural Work
- Room/Interior Design, Artistic Transformation
- Environmental Protection/Detecting Ecosystem Changes
- Historical Artifact reconstruction

Expand into more unique use cases with the power of Vision Models

Let’s wrap it up! Today We...

- Introduced Fashion Buddy's use of Vision Models
- Learned about Vision Models
- Explored outputs retrievable from Vision Models
- Experimented with various outputs
- Understood how to leverage this data in an end-to-end solution (Fashion Buddy)
- Discovered applications and benefits of Vision Models for empowering Image Recognition Apps and Gen AI

Resources

Fashion Buddy

Sign Up for Astra DB

Connect with Me

Thank you!

THE AI CONFERENCE