Final Project Presentation

GROUP 4

YU ZHANG

YONGQI YU
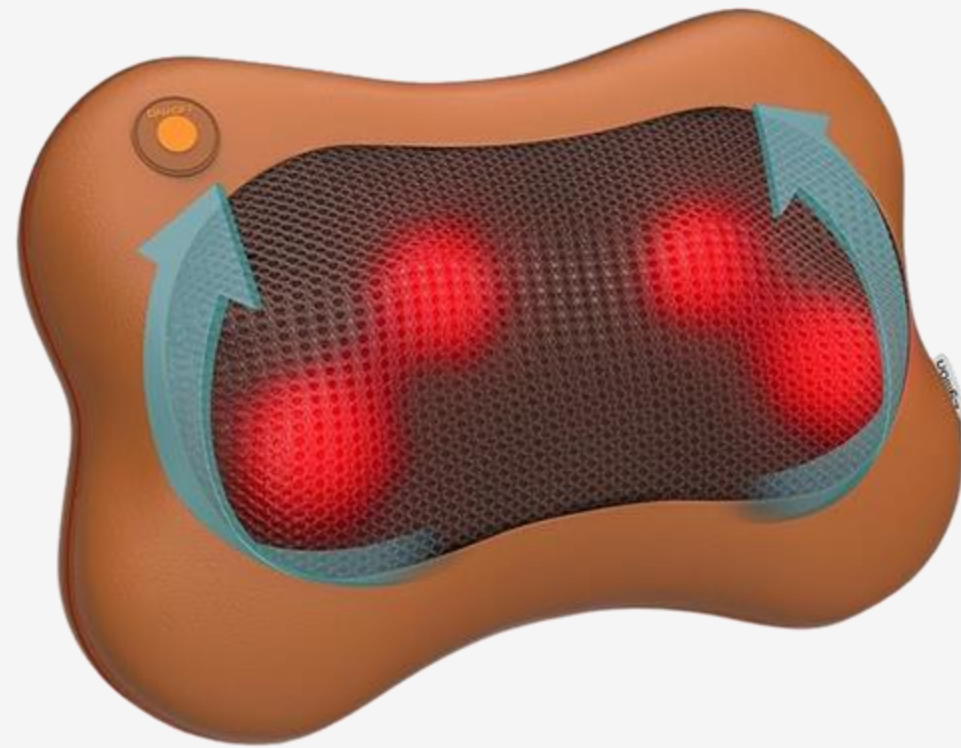
LIBIN CHEN

JOANNA CHANG

Outline

- **Product Overview**
- **Pipeline Selection**
  - Model
  - Chunking strategy
- **Analytics Engine Components**
  - Summarization, Visual feature, Product feature, Sentiment analysis
  - Topic extraction, Image generation summary
- **Image Generation**
- **AI Agentic Workflow**
- **Challenges and Limitations**
- **Future Enhancements**

Overview of Product Categories

ZYLLION SHIATSU BACK AND NECK MASSAGER WITH HEAT

Health & Household

4.4 ⭐ (50,884 reviews)

8BITDO RETRO MECHANICAL KEYBOARD

PC Accessories

4.8 ⭐ (2,183 reviews)

HARIO V60 CERAMIC POUR OVER COFFEE SET

Home & Kitchen

4.6 ⭐ (1,921 reviews)

Rationale for Selection

DISTINCT CATEGORIES

Three distinct product categories enhances the **diversity** of the evaluation and ensures **varied textual cues** for analysis

VISUALLY & SEMANTICALLY RICH

The selected products are **visually and semantically rich**, providing ample opportunities for exploring text-derived visual signals and enhancing model performance

DIVERSE REVIEW VOLUMES

Review counts range from 2,000 to 50,000, and all products have consistently high average ratings between 4.4 and 4.8

Pipeline Overview

Massager differs in
- Review numbers
- Chunking strategy

	Massager	**Keyboard**	**Coffee Set**
Model	gpt-5o	gpt-5o	gpt-5o
How reviews are obtained (top reviews)	Web scraping (Selenium package)	Hard-coded	Hard-coded
# of reviews	50 reviews	5 reviews	5 reviews
Chunking Strategy	chunk_size = 3000 overlap = 200	None	None

Analytics Engine Components

Keyboard
- No summarization
- Fewer feature extraction

Coffee set
- No product feature extraction

	Massager	**Keyboard**	**Coffee Set**
Summarization	per-chunk summaries	None	Overall summary, pros/cons, keywords
Visual feature extraction	1. Colors mentioned 2. Materials described 3. Size/Dimensions 4. Shape/Design elements 5. Textures 6. Visual features (buttons, straps, etc.) 7. Overall appearance	1.colors 2.materials 3.shape design 4.visual features 5.overall aesthetic	1. materials 2. colors 3. shapes 4. textures 5. patterns 6. functional visual element 7. usage scenes
Product feature extraction (functional/ design)	1. Functional Features (what it does) 2. Design Features (design elements) 3. Material Features 4. Size/Portability 5. Usage Context (where/how used) 6. Key Selling Points	1.functional_feature 2.design_features 3.connectivity 4.unique_selling_points	None (combined with visual feature extraction)

Visual feature & Product feature (conti.)
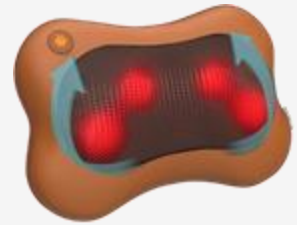

```
visual_prompt = f"""
Analyze the following customer reviews and extract ALL visual information about the product.

Product: Zyllion Shiatsu Back and Neck Massager with Heat (Model: ZMA-13)

{condensed_review_text}

Extract:
1. Colors mentioned
2. Materials described
3. Size/Dimensions
4. Shape/Design elements
5. Textures
6. Visual features (buttons, straps, etc.)
7. Overall appearance

Format as JSON:
{{
    "colors": ["color1", "color2"],
    "materials": ["material1", "material2"],
    "size_dimensions": ["mention1", "mention2"],
    "shape_design": ["feature1", "feature2"],
    "textures": ["texture1", "texture2"],
    "visual_features": ["feature1", "feature2"],
    "overall_appearance": "summary description"
}}
"""
```

```
visual_prompt = f"""
Analyze the following reviews for the 8BitDo Retro Mechanical Keyboard.
Extract ONLY visual/physical attributes.

Reviews:
{all_review_text[:15000]}

Return JSON with keys:
colors, materials, shape_design, visual_features (specific parts like knobs, LEDs), overall_aesthetic.
"""
```

```
def build_visual_feature_prompt(text):
    return f"""
You are an expert in extracting visual and physical product characteristics from text.

From the following product reviews, extract ONLY concrete visual or physical features of the product.
Include attributes such as: shape, color, material, texture, size, special patterns, and usage environment.

Return the output in JSON with this structure:
{{
  "materials": [],
  "colors": [],
  "shapes": [],
  "textures": [],
  "distinctive_patterns": [],
  "functional_visual_elements": [],
  "common_usage_scenes": []
}}
```

```
features_prompt = f"""
Analyze the product description and customer reviews to extract key product features.

PRODUCT DESCRIPTION:
Title: {product_description.get('title', 'N/A')}

Key Features:
{features_text}

CUSTOMER REVIEWS:
{condensed_review_text}

Extract:
1. Functional Features (what it does)
2. Design Features (design elements)
3. Material Features
4. Size/Portability
5. Usage Context (where/how used)
6. Key Selling Points

Format as JSON:
{{
    "functional_features": ["feature1", "feature2"],
    "design_features": ["feature1", "feature2"],
    "material_features": ["feature1", "feature2"],
    "size_portability": "description",
    "usage_context": ["context1", "context2"],
    "key_selling_points": ["point1", "point2"]
}}
```
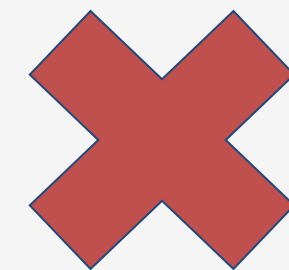
```
feature_prompt = f"""
Extract key functional and design features from these reviews.

Reviews:
{all_review_text[:15000]}

Return JSON with keys:
functional_features, design_features, connectivity, unique_selling_points.
"""
```

Analytics Engine Components

- No topic extraction for keyboard
- Recommended prompts are used directly to generate images

	Massager	**Keyboard**	**Coffee Set**
Sentiment analysis	1.Overall sentiment 2.Sentiment themes 3.Visual sentiment 4.Satisfaction score	1.Overall sentiment 2.Sentiment themes 3.Visual sentiment	1.Overall sentiment 2.Sentiment distribution 3.Reasoning 4.Summary
Topic extraction	1.Topic name 2. Description 3. Keywords 4.Visual relevance	None	1.Topic 2.Description 3.Representative keywords 4.Representative comments
Image generation summary	1.Visual description 2.Key visual elements 3.Recommended prompts for image generation	1.Visual description 2.Key visual elements 3.Recommended prompts for image generation	Recommended prompts for image generation

Image Generation

Transform textual product attributes extracted from customer reviews into realistic product images using image generation models.

We evaluate:

1. How well models convert text to visual appearance

2. How different prompts affect fidelity & realism

3. Model differences between OpenAI and SDXL

Models Used

OpenAI GPT-Image-1

- Better at following instructions
- More accurate geometry
- Lower hallucination
- Slightly simpler textures

Stable Diffusion XL (SDXL)

- Better at texture realism & lighting
- Very aesthetic images
- Higher variability
- Occasional shape/component errors

Our Prompt Versioning Strategy

For each of 3 products:

3 prompt versions

- **Prompt 1: Natural Language Prompt**

  - Reads like a normal description, minimal structure

- **Prompt 2: Condensed Attribute-Focused Prompt**

  - Removes storytelling language

  - Focuses on essential product features

  - More consistent than prompt 1

- **Prompt 3: Structured Photorealistic Prompt**

  - Most structured and detailed

  - Includes materials, geometry, textures

  - Includes lighting, background, photography style

```
"product1_massager": {
    "v1": """Create an image of a compact, pillow-like back and neck massager in rich brown fabric.
    The massager should feature soft silicone nodes on each side, an ergonomic design that fits body
    contours, and Velcro straps for securing it to a chair. Include user-friendly buttons and a
    zipper for a replacement cover. The overall look should convey a sturdy and modern aesthetic,
    suitable for therapeutic use in home or office settings.""",

    "v2": """Ultra-realistic photo of an ergonomic shiatsu massage pillow.
    Brown faux leather + fabric material, curved compact shape, four rounded massage nodes,
    simple control buttons, velcro straps. Bright studio lighting, sharp detail.""",

    "v3": """Realistic studio product photo of a compact, pillow-shaped shiatsu back and neck massager.
    Key visual features:
    - rich brown fabric cover made of nylon/polyester
    - four raised soft silicone massage nodes forming smooth rounded bumps
    - ergonomic curved design that fits body contours
    - Velcro straps on the back for attaching to a chair
    - user-friendly side control buttons and a visible zipper for a replaceable cover
    Style: clean white seamless background, soft diffused lighting, crisp detail and accurate textures."""
},
```

Product 1 (Massager)

OpenAI

SDXL
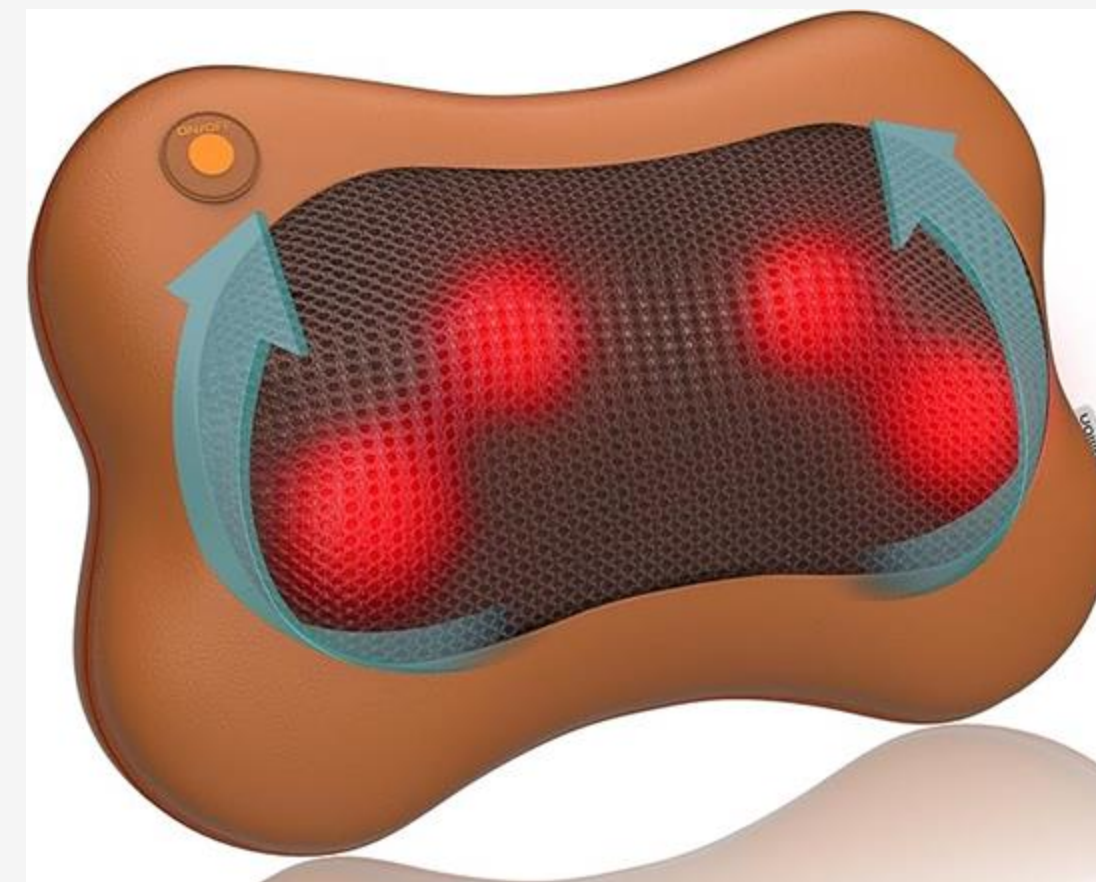
Key Findings:

- OpenAI generated images closely match the real massager's shape and color.
- The pillow-like outline and four glowing massage nodes are accurately captured.
- Button placement is generally correct, though design differs slightly.
- Models did not reproduce the blue rotation arrows seen in the real product.
- SDXL outputs show higher realism but lower accuracy, often drifting into neck pillows.

Product 2 (Keyboard)

Prompts

Initial prompt (v1):

*Create a image of a retro-style **audio device** with a boxy shape made of thick plastic. The device has a creamy grey body adorned with bold red accents. It features large 'Super Buttons' on the front, a prominent central volume knob, and a soft glowing power LED indicator. The overall design exudes a nostalgic aesthetic reminiscent of classic electronics, highlighting concave keys and a user-friendly layout.*

Updated prompt (v2):

*Ultra-realistic retro-style **mechanical keyboard** modeled after classic 8-bit consoles. Matte creamy grey ABS plastic housing, bold red accent buttons, oversized Super Buttons, concave retro keycaps, a metallic central volume knob, and an illuminated LED indicator. Bright studio lighting, sharp detail.*

Product 2 (Keyboard)

OpenAI

SDXL

Key Findings:

- OpenAI captures the color scheme (cream gray + red) but often misinterprets the product as a retro console or button box rather than a full keyboard.
- Layout accuracy is low: OpenAI struggles with accurate key arrangement, legends, and overall keyboard proportions.
- SDXL outputs show stronger realism but frequently hallucinate incorrect forms (e.g., audio equipment, multi-piece boards, random knob placement).
- Both models capture the retro aesthetic, but neither achieves true product-level accuracy in geometry or layout.

Product 3 (Hario Pour Over Coffee Set)

OpenAI

SDXL

Key Findings:

- OpenAI accurately generated all core components (ceramic dripper, glass server, paper filters, scoop) with correct overall arrangement.
- The dripper shape (V60 cone + spiral ribs) is captured reasonably well
- The glass carafe shape matches the real product, but measurement markings are inconsistent, and may missing.
- Paper filters are recognized correctly as stacked fan shapes, but thickness and orientation vary.
- SDXL outputs show high realism but significant semantic drift, like introducing wood bases, metal stands, or unrelated brewing accessories.

Image Generation Summary

- **OpenAI → higher accuracy**, better at capturing correct shape, layout, and components.

- **SDXL → more photorealistic**, but often hallucinates unrelated product forms.

- **Structured prompts** produce the most consistent and faithful results.

- Both models struggle with **fine details**, UI icons, and precise component geometry.

- Customer reviews provide enough signal for **approximate product reconstruction**, but not exact replication.

AI Agentic Workflow

Researcher Agent

- Function: Data ingestion
- Live Mode: Selenium-based scraping of Amazon product pages (descriptions, reviews)

Analyst Agent

- Function: Use LLMs (GPT-4o) to parse unstructured text
- Output: Extracts objective visual features and computes sentiment analysis

Creative Agent

- Function: Convert analyst output into a high-fidelity image-generation prompt optimized for diffusion models

Visualizer Agent

- Function: Use DALL·E 3 to generate a visual prototype from the creative prompt

AI Agentic Workflow

Demo

Challenges and Limitations

NOISY REVIEWS

User reviews often contain **inconsistent language** and irrelevant information, making it difficult to extract reliable insights for image generation.

DATASET GENERALIZABILITY

Relying on a **small subset** of customer reviews (5-50), the extracted visual features may not fully represent the product, limiting the generalizability and accuracy of the generated images.

METRIC FOR GENERATED IMAGE

There is no fixed or standardized metric to **evaluate** the quality of the generated images or to objectively compare them with real product images, making assessment largely subjective.

Future Enhancements

CHUNKING

Implementing more effective **chunking strategies** will preserve full contextual meaning across reviews, allowing the LLM to extract richer and more accurate visual features. This enhancement would reduce information loss and ultimately lead to more faithful, higher-quality product image generation.

GENERATION IMPROVEMENT

Different **structure** of extracted information could be tested to achieve higher quality.
Fine-tuning the generation model using a selected similarity metric would help the model produce images that more closely match the real product's appearance.

PIPELINE ENHANCEMENT

Iteratively **feeding** both the generated images and the original image back into the agent can progressively improve image quality.
Try a **multiple agent** workflow with each agent assigned specific task with high performance.

Thank You for Your Attention