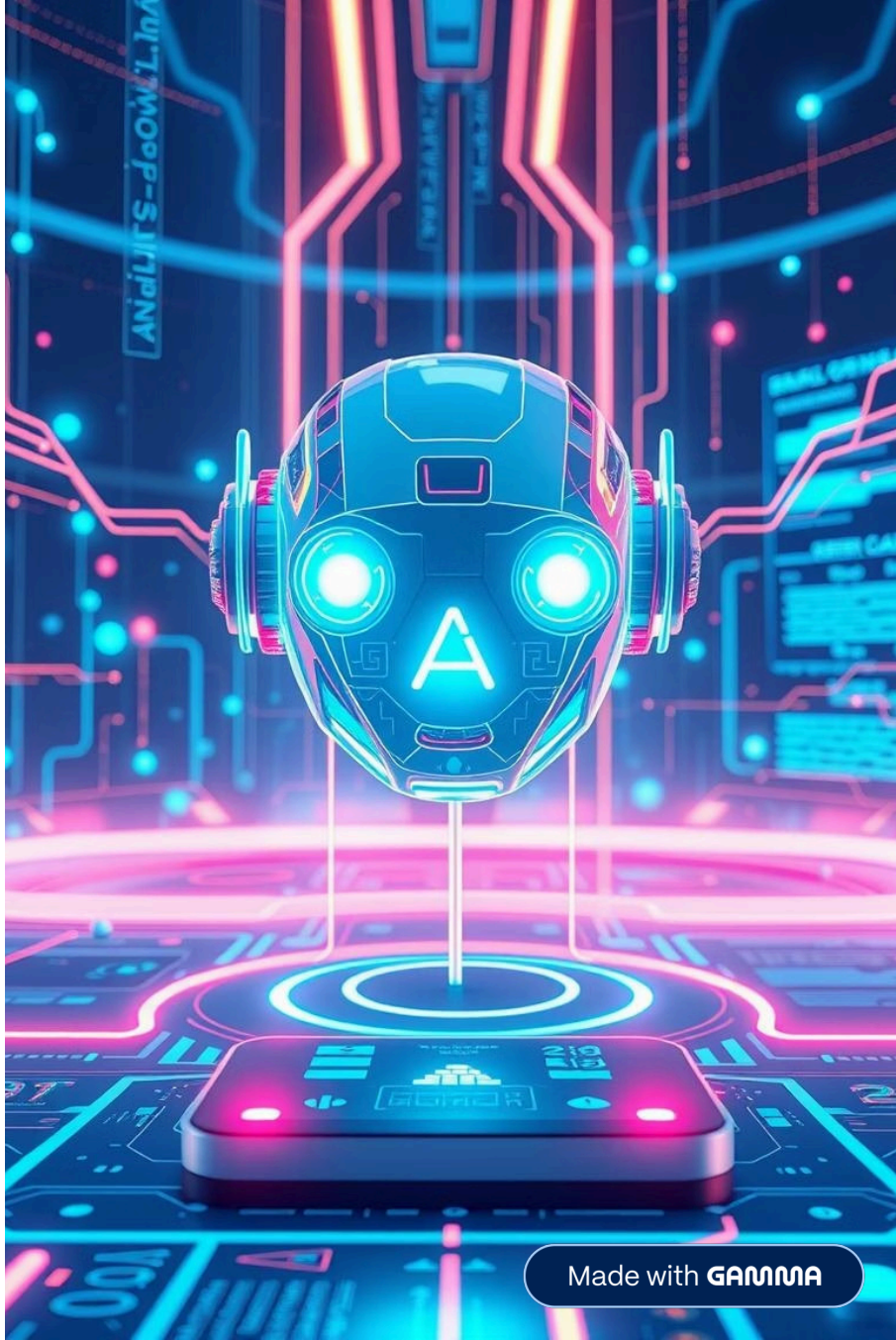
# AI Model Modalities, API calls and Closed-Source Providers

This presentation explores the fundamentals of AI model modalities, key closed-source large language model (LLM) providers, their core capabilities, and how to access them via APIs. We will also discuss important limitations and ethical considerations surrounding these technologies.

P **by Pat Pascual**



# What Are AI Modalities?

## Definition

Modalities are the types of inputs and outputs an AI model can handle, such as text, images, audio, and video.

## Multi-modal Models

Leading foundation models increasingly support multi-modal interactions, enabling reasoning across diverse data types.

## Examples of Modalities

- Text-to-text, text-to-image
- image-to-text, audio-to-text
- text-to-audio, an multi-modal dialogues


# Modalities Supported by Leading Models (2025)

| Model | Supported Modalities |
| --- | --- |
| GPT-4o (OpenAI) | Text, Image, Audio input/output (real-time); Native image generation; Code interpreter |
| Claude 3.5 (Anthropic) | Text, Image (vision-supported); GUI interaction (Computer Use); Audio planned |
| Gemini 2.5 (Google) | Text, Image (vision-native), Audio output (emotional & experimental); Multilingual |

# Overview of Key Closed-Source LLM Providers

| Organization | AI Personality | Mission Focus | Distinctive Edge |
|---|---|---|---|
| **OpenAI** | Helpful, cautious, globally responsible | Safe AGI for all humanity | Real-time multimodal, strong safety protocols |
| **Anthropic** | Transparent, ethical, reasoning-driven | Reliable AI via Constitutional AI | Long-context, structured ethical principles |
| **Google DeepMind** | Efficient, integrated, utility-focused | Multimodal AI for Google ecosystem | Vision-native, emotional voice, wide deployment |


# Core Capabilities Across Providers

| Capability | GPT-4 (GPT-4o) | Claude (Claude 3) | Gemini (1.5 / 2.5) |
|---|---|---|---|
| **Text Generation** | ✓✓✓ (Highly coherent, creative) | ✓✓✓ (Strong context handling) | ✓✓✓ (Powerful & detailed) |
| **Multi-modal Reasoning** | ✓✓✓ (Vision, audio, text) | ✓✓ (Text only, image/audio coming) | ✓✓✓ (Text + vision) |
| **Coding & Math** | ✓✓✓ (GPT-4o / Codex) | ✓✓ (Strong but less deterministic) | ✓✓✓ (Robust with long context) |
| **Tool Use & Functions** | ✓✓✓ (Assistants API, tools) | ✓ (Planning in Claude 3 roadmap) | ✓ (Experimental or limited in IDEs) |
| **Context Memory** | Up to ~128K tokens | Up to 200K tokens | Up to 1M tokens |
| **Safety & Alignment** | Strong (RLHF + moderation) | Very strong (Constitutional AI) | Strong (Safety layers + evals) |

# QnA and Recall

# Accessing Closed-Source LLMs via APIs

1

## What is an API?

An API acts like a waiter: you place an order (request), the chef (LLM server) prepares it, and the waiter delivers the result.

2

## How to Use APIs

1. Sign up on OpenAI, Anthropic, or Google AI Studio
2. Get your API key (digital ID badge)
3. Send requests using code (e.g., Python)
4. Monitor usage limits and costs

3

## Why Use APIs?

- Simple integration without hosting models
- Scalability handled by providers
- Built-in safety filters and alignment

# Limitations of Large Language Models

### Hallucinations

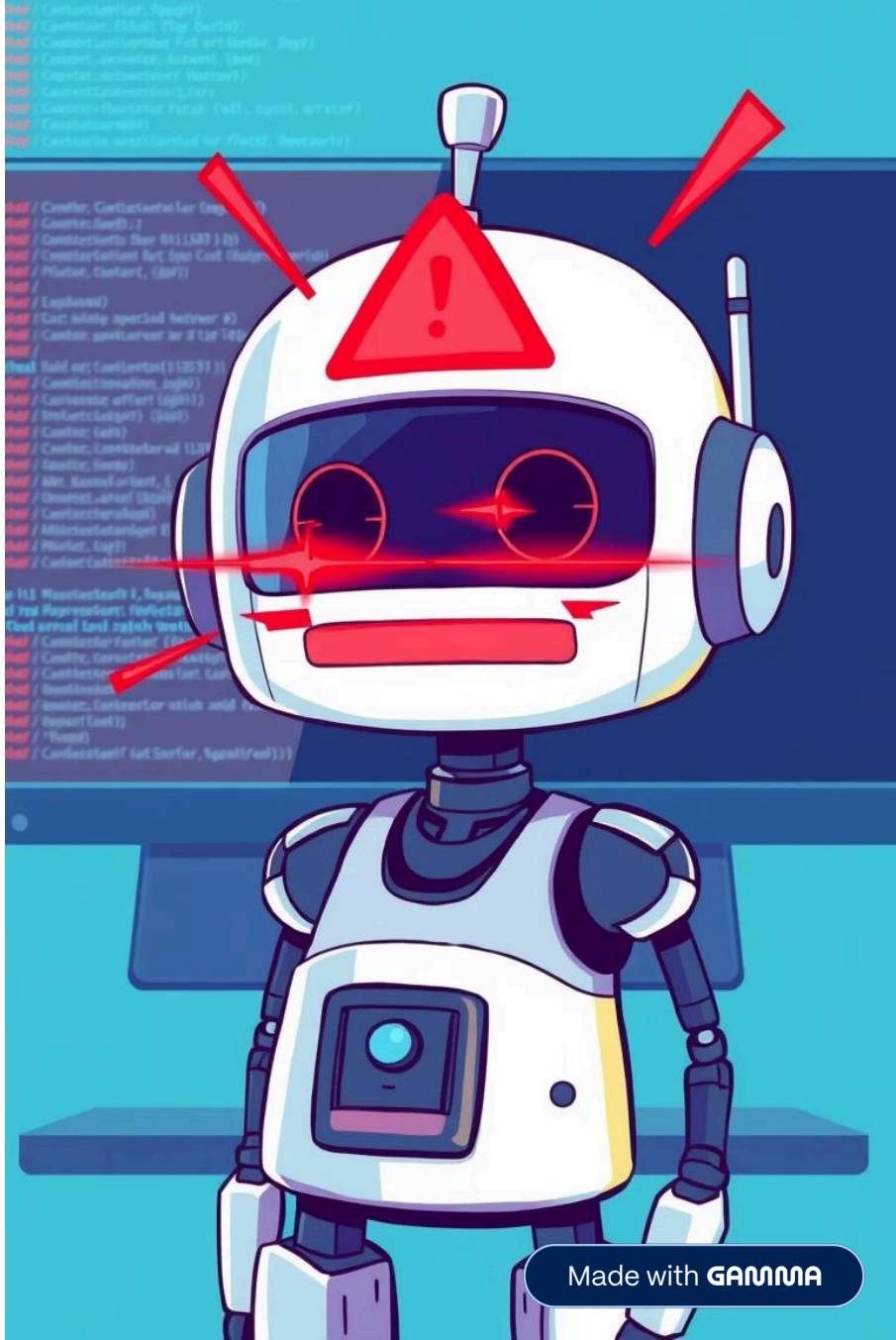
Models can generate inaccurate or fabricated information.

### Prompt Sensitivity

Output quality varies significantly based on input phrasing.

### Cost

Usage is billed per token, which can add up with heavy use.



# Cost Comparison

| Organization | Model Tier | Approx. Cost (API usage) | Notes |
|---|---|---|---|
| OpenAI | GPT-4o | ~$5.00 per 1M input tokens | Includes text, image, and audio processing |
| | GPT-4 | ~$30.00 per 1M input tokens | High accuracy, no audio or real-time multimodality |
| | GPT-3.5 | ~$0.50 per 1M input tokens | Most affordable, widely used in chatbots |
| Anthropic | Claude 3 Opus | ~$15.00 per 1M input tokens | Premium reasoning, long context (200K+ tokens) |
| | Claude 3 Sonnet | ~$3.00 per 1M input tokens | Balanced performance and cost |
| | Claude 3 Haiku | ~$0.25 per 1M input tokens | Optimized for speed, lowest cost |
| Google DeepMind | Gemini 1.5 Pro | ~$10.00 per 1M input tokens (estimated) | High reasoning and multimodal support |
| | Gemini Flash | ~$0.35 per 1M input tokens (estimated) | Optimized for speed and cost-efficiency |
| | Gemini 2.5 (Exp) | Not publicly disclosed | Emotion, audio support, premium experimental tier |

# Ethical Considerations in AI Deployment

### Bias & Fairness

Models may produce prejudiced or harmful outputs if not carefully managed.

### Data Privacy

Responsible handling of sensitive input data is critical.

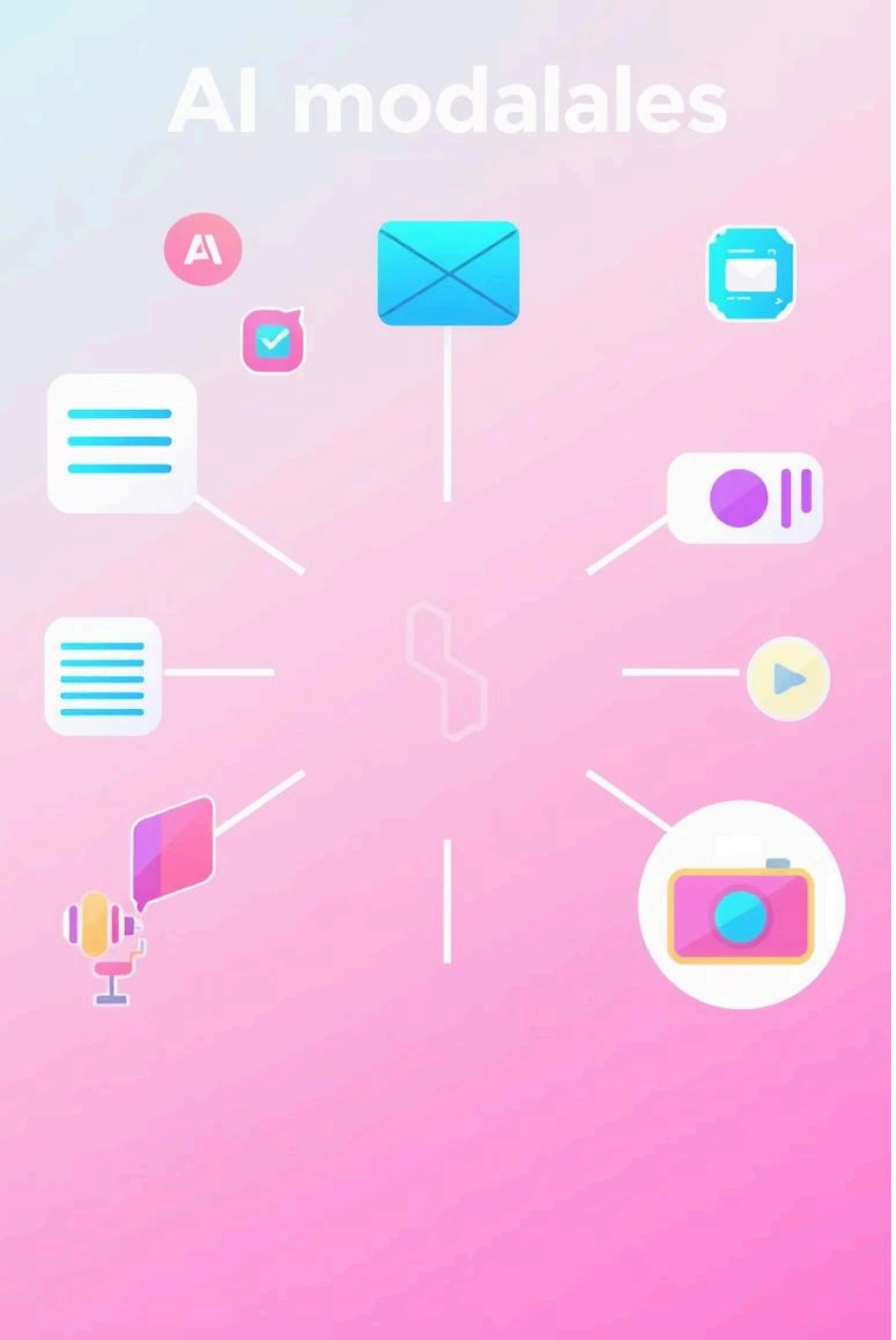
### Deployment Ethics

Avoid misuse in sensitive domains like law and healthcare.

# Regulations to Consider for AI Use

- **GDPR (EU)** – Data protection and privacy regulations in Europe.
- **HIPAA (US)** – Health data privacy rules in the United States.
- **AI Act (EU)** – Proposed regulations for AI safety and transparency.
- **NIST AI RMF (US)** – Framework for responsible AI risk management.

# QnA and Recall




# Activity #2

Let's create your very first closed-source chatbot.

# Activity # 3 – Business Use Case

1

**Identify the top 3 challenges or inefficiencies in your target domain or industry.**

Example: “Employees waste time writing repetitive reports” or “Customers drop off due to unclear product descriptions.”

2

**Define who experiences these problems.**

Example: “Sales teams,” “College students,” or “E-commerce product managers.”

3

**Clearly state how LLMs (via API) will uniquely solve the problem or create value.**

- Simple integration without hosting models
- Scalability handled by providers
- Built-in safety filters and alignment

# Activity # 3 – Business Use Case

4

### Describe the function or purpose the tool

Example: “A summarizer for call center logs,” or “A tone-converter for customer emails.”

5

### List and define each tasks.

Example: “Auto-generate social media captions,” “Extract customer pain points from reviews,” or “Rephrase technical content for non-experts.”

6

### Describe the overall process. Focus on how the data will flow through the API.

Example: “User uploads a file → API extracts key points → Summary is sent back to user.”

# Activity # 3 - Business Use Case

7

### Define the limitations of the tool

Example: “It can only respond to love-related questions”, “Only works with English input”

8

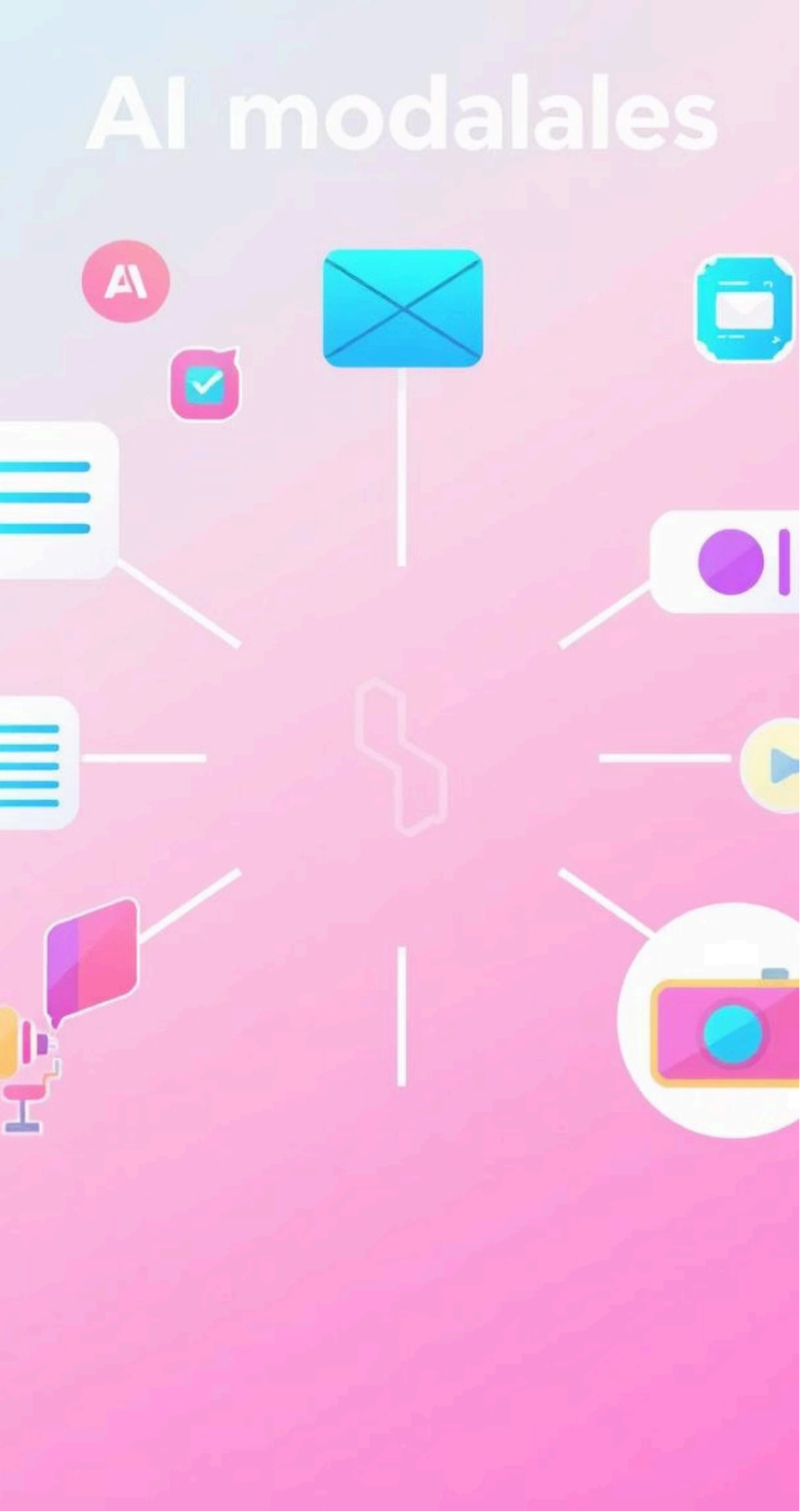
### Define how the solution generates revenue or value

Example: “Internal time saved = operational cost savings”

9

### Define how you measure success

Example: “Average time saved per user”, “Uptake rate of auto-generated drafts”




# Activity #4

Let's build your business usecase! Generate a system_prompt using this:

You are an intelligent assistant designed to help solve real-world problems using a language model via API. Below is the user’s specification. Read it carefully and perform only what is defined in the tasks and process.

🧩 **Industry Challenges**

👤 **Who Experiences the Problems**

✨ **LLM Solution Value**

🎯 **Tool Function or Purpose**

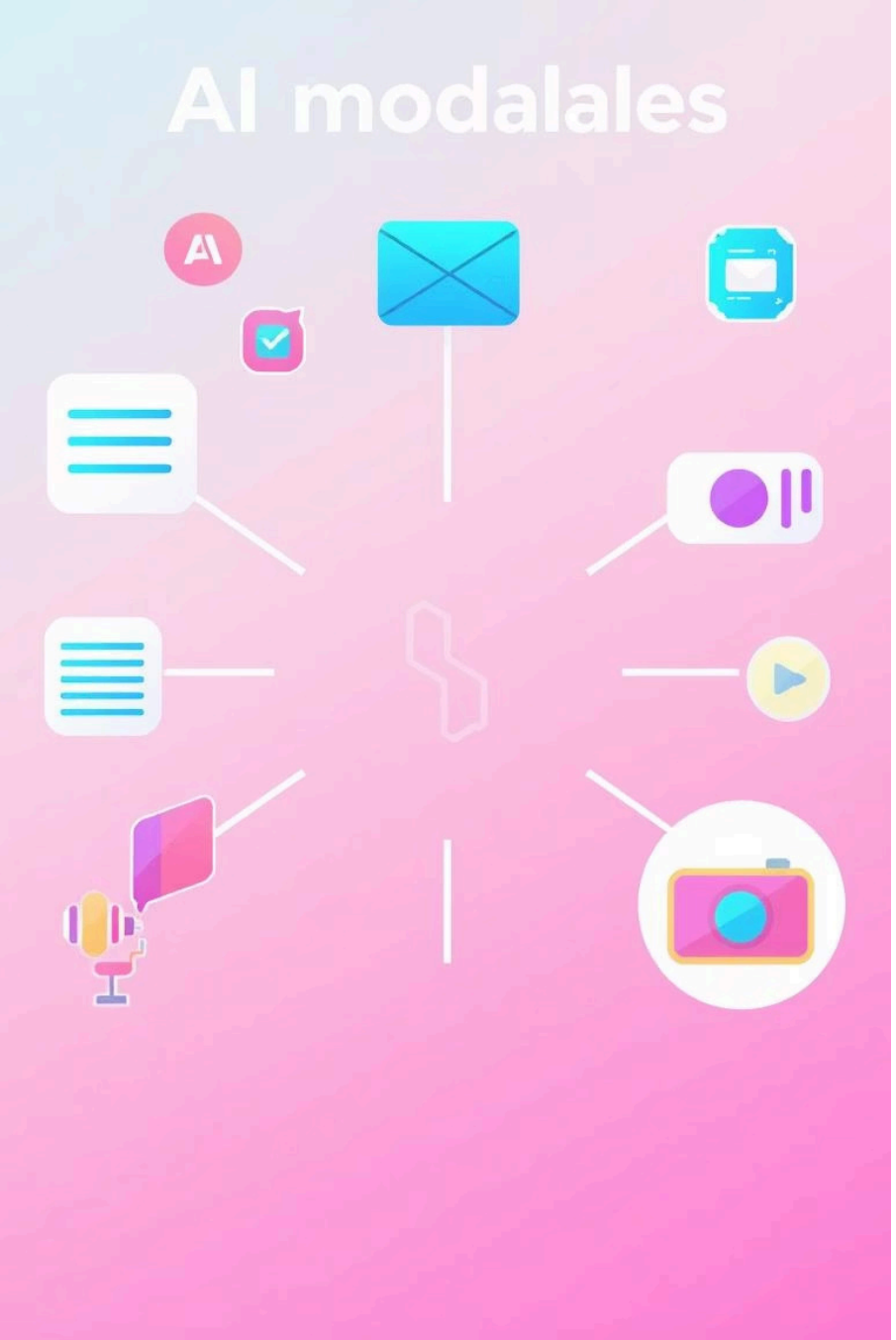
🛠️ **Tasks Performed**

🔄 **Data Flow / Process Overview**

🚫 **Tool Limitations**

💰 **Revenue Streams**

📈 **Key Metrics**




# Activity #5

Deploy you Activity #4 via Streamlit.

# QnA and Recall