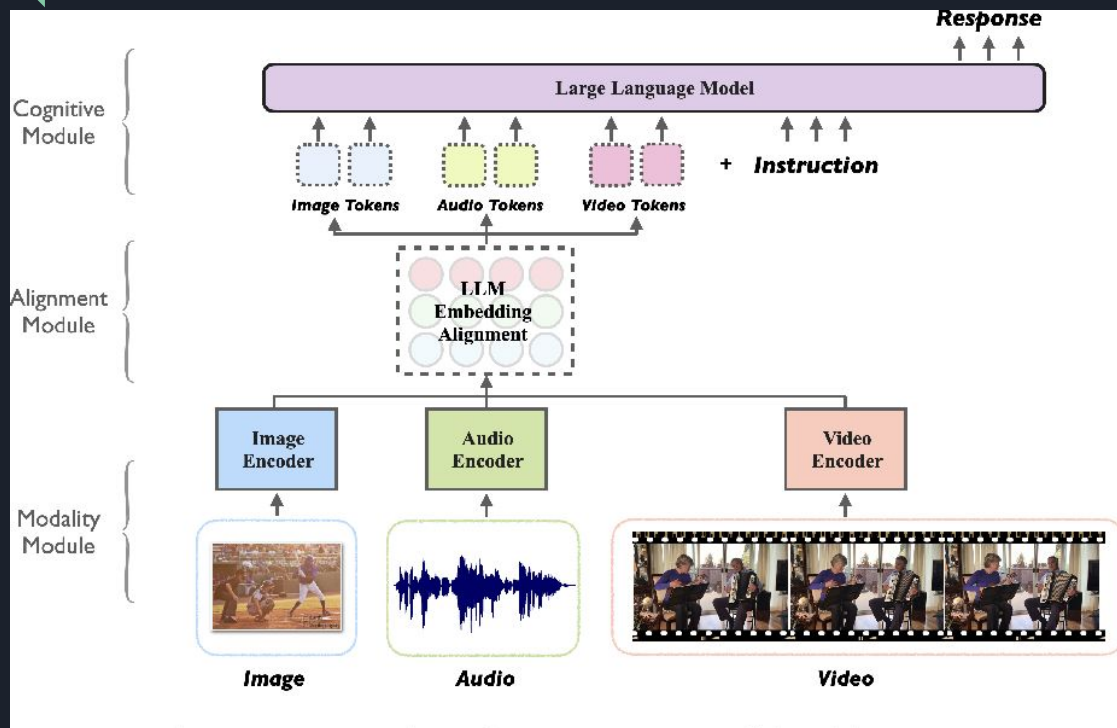


A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

# What are Multimodal LLMs

How to use Generative AI to create content

# Transformer models that process more than just Text



Modules process images, audio, and video and feed into the same cognitive space

This transforms the other forms of media through a process called embedding

Then feeds into the LLM



# MultiModal LLMs are not just theory

## Models like:

- Kosmos-1 by Microsoft
- PaLM-E by Google
- GPT-Vision by OpenAI

## Products like:

- GPT-Vision as part of ChatGPT
- BingChat as part of Microsoft



# Advantages of Multimodal LLMs

- **Improved accuracy:** Multimodal LLMs can achieve higher accuracy on tasks that require understanding multiple types of data. For example, they can better understand the context of a conversation if they can see the facial expressions of the people involved.
- **New applications:** Multimodal LLMs can be used for a wider range of applications than traditional LLMs. For example, they can be used to generate captions for images, translate languages, and even create art.
- **More natural interactions:** Multimodal LLMs can interact with humans in a more natural way, as they can understand and respond to multiple types of input. For example, they can understand a question asked in spoken language and respond with a written answer.
- **Better understanding of the world:** Multimodal LLMs can develop a better understanding of the world by learning from multiple types of data. This can help them to make better decisions and predictions.



# Use cases

- **AR and VR:** Multimodal LLMs can process images from the natural world and reason about them, allow them to be used in machinery and robots.
- **Education:** Multimodal LLMs are being used to create more personalized and engaging learning experiences. For example, they can be used to generate interactive lessons that adapt to the student's individual needs.
- **Healthcare:** Multimodal LLMs are being used to improve the diagnosis and treatment of diseases. For example, they can be used to analyze medical images and identify potential health problems.
- **Customer service:** Multimodal LLMs are being used to provide more efficient and personalized customer service. For example, they can be used to answer customer questions in real time and resolve issues quickly.



# Multimodal LLMs still have limitations

- Processing images and audio is more ambiguous than processing text
- Data quality and interpretation problems might multimodal LLMs less accurate than a human
- Multimodal LLMs are still subject to biases based on the training data.