

# Visual Language Model

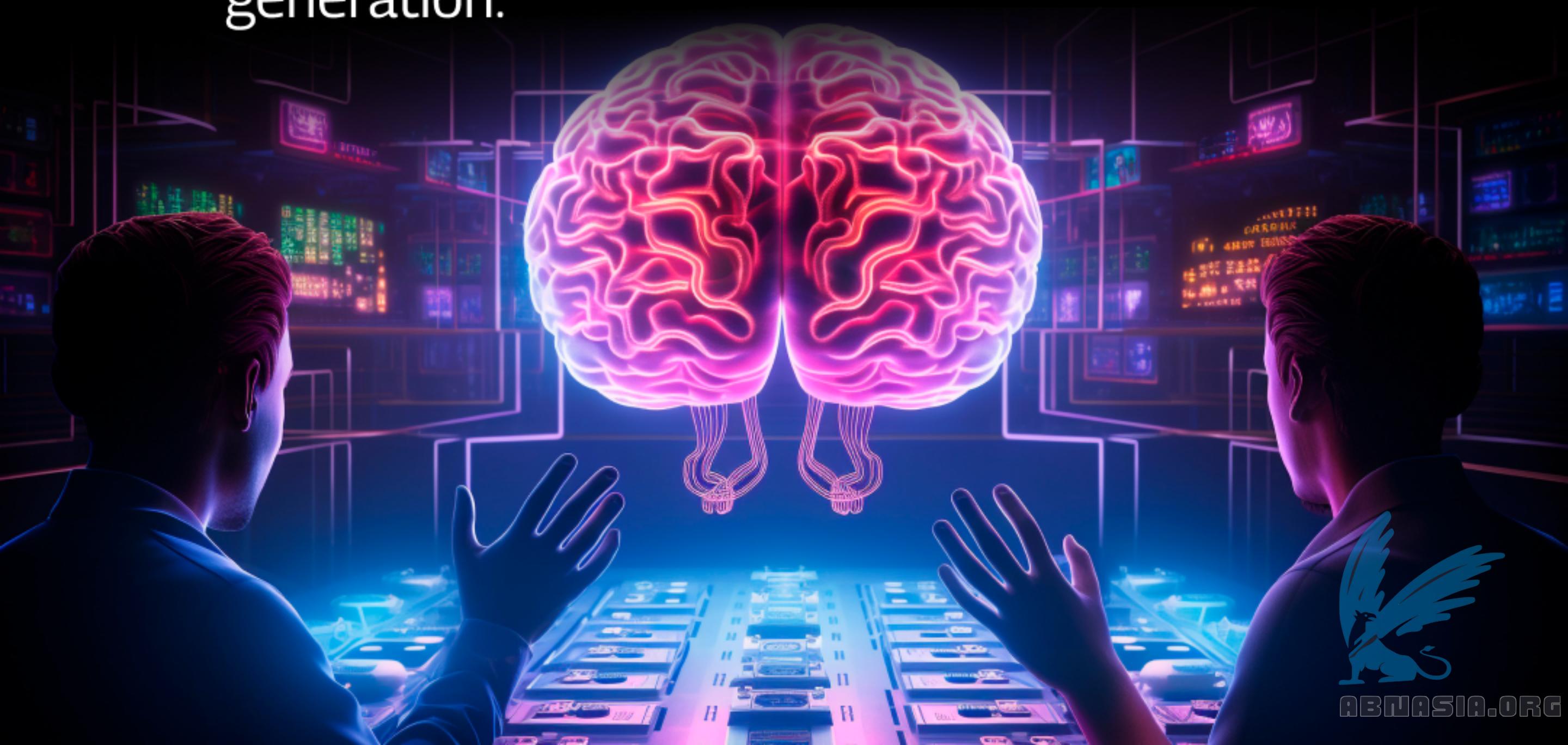
Seeing the World, Understanding the Words



swipe to know more

# What are VLMS?

- It combines visual and textual data to understand and generate content across both modalities.
- It processes images and text together for deeper visual content interpretation using natural language understanding.
- VLMS are used in tasks like image captioning, visual question answering, and content generation.



# The Working Principle of VLMS

## Step-1

### Image Processing

Use ViT (Vision Transformer) to process images.

## Step-2

### Text Processing

Apply language models like BERT to process text

## Step-2

### Input Fusion

Combine processed images and text for context-aware outputs. accuracy by 40%, enhancing understanding.



Get 1:1 On Demand Doubt Clearing Sessions



# Benefits of Visual Language Models



## Richer Understanding:

AI comprehends both visual and textual context.



## Better Accuracy:

Outputs are more reliable due to multimodal learning.



## Creative Automation:

Automates tasks like captioning, product search, and art creation.

# Tools based on VLMs

- **CLIP:** Aligns images with text for better understanding.
- **BLIP:** Enhances language-image learning.
- **DALL·E:** Creates images from text descriptions.

“

**McKinsey reports a 20% boost** in operational efficiency for businesses using AI models based **VLMs** like **CLIP** and **DALL·E**.



# Algorithms Driving VLMs

- **Attention Mechanism:** Focuses on important aspects of both images and text.
- **Self-Supervised Learning:** Trains models without needing labelled data.
- **Transformer Networks:** Power large-scale data processing for text and images.

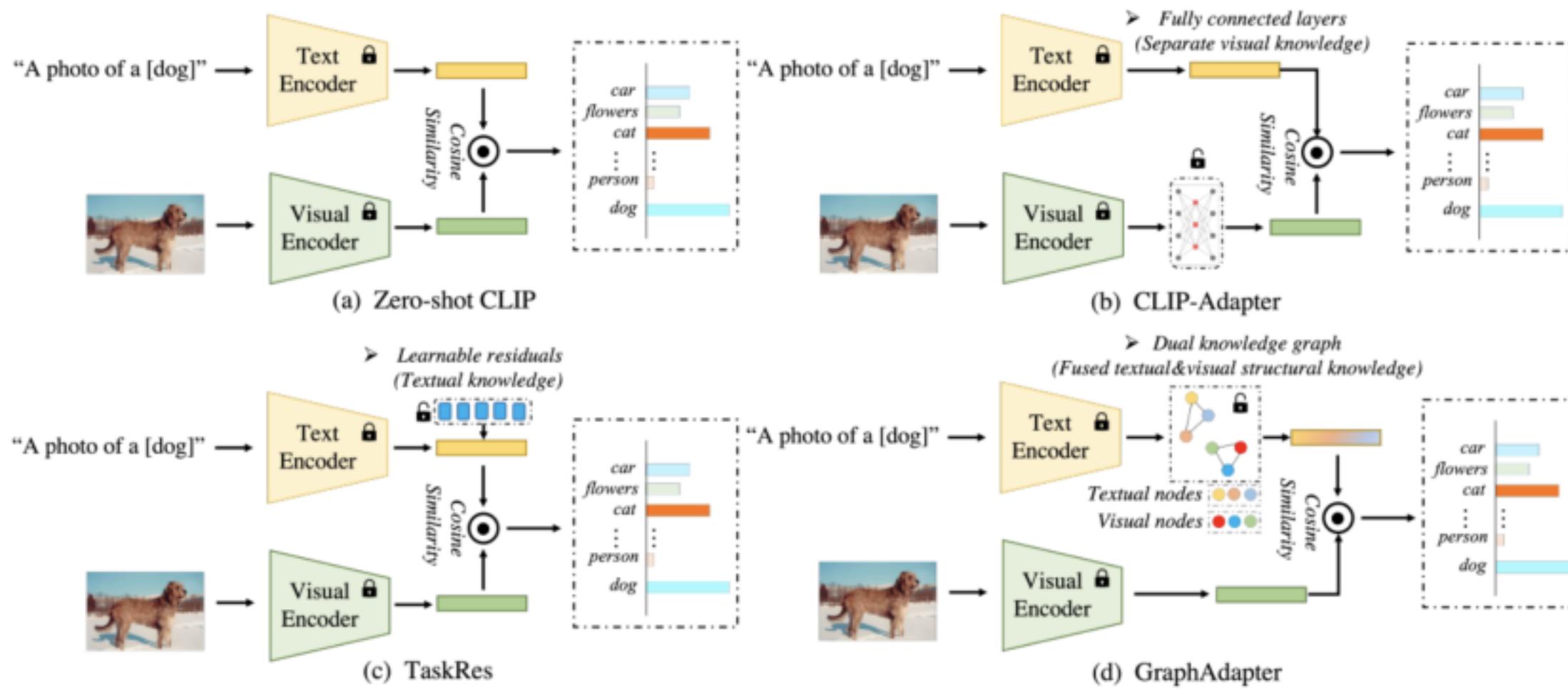


# Tuning VLMs with Dual Knowledge Graphs

- Enhanced Understanding: Dual knowledge graphs improve text-image relationship mapping.
- Boosted Accuracy: Fine-tuning with these graphs increases model accuracy.

“

**Stanford AI** research shows dual knowledge graphs boost **VLM accuracy by 30%**.



# Challenges in Training VLMs

- **Massive Data Needs:** Requires vast datasets of images and text.
- **Bias:** Risk of picking up biases from training data.
- **Computing Power:** High computational resources required for training



Get 1:1 On Demand Doubt Clearing Sessions

# The Future of VLMs

- VLMs will revolutionize design, film, and gaming with complex content.
- AI assistants will become more intuitive with VLMs.
- VLMs will enable richer AR experiences through contextual interactions.
- VLMs will improve human-AI collaboration by streamlining processes.

“

**PwC** estimates VLMs could boost the global economy by **₹1,300 lakh crore** by **2030** through virtual environments and AR.

”

**OpenAI** research shows models like **CLIP** improve text-image alignment



Get Certification from:



IIT  
Guwahati

Woolf University/

Microsoft



ABN ASIA.ORG