

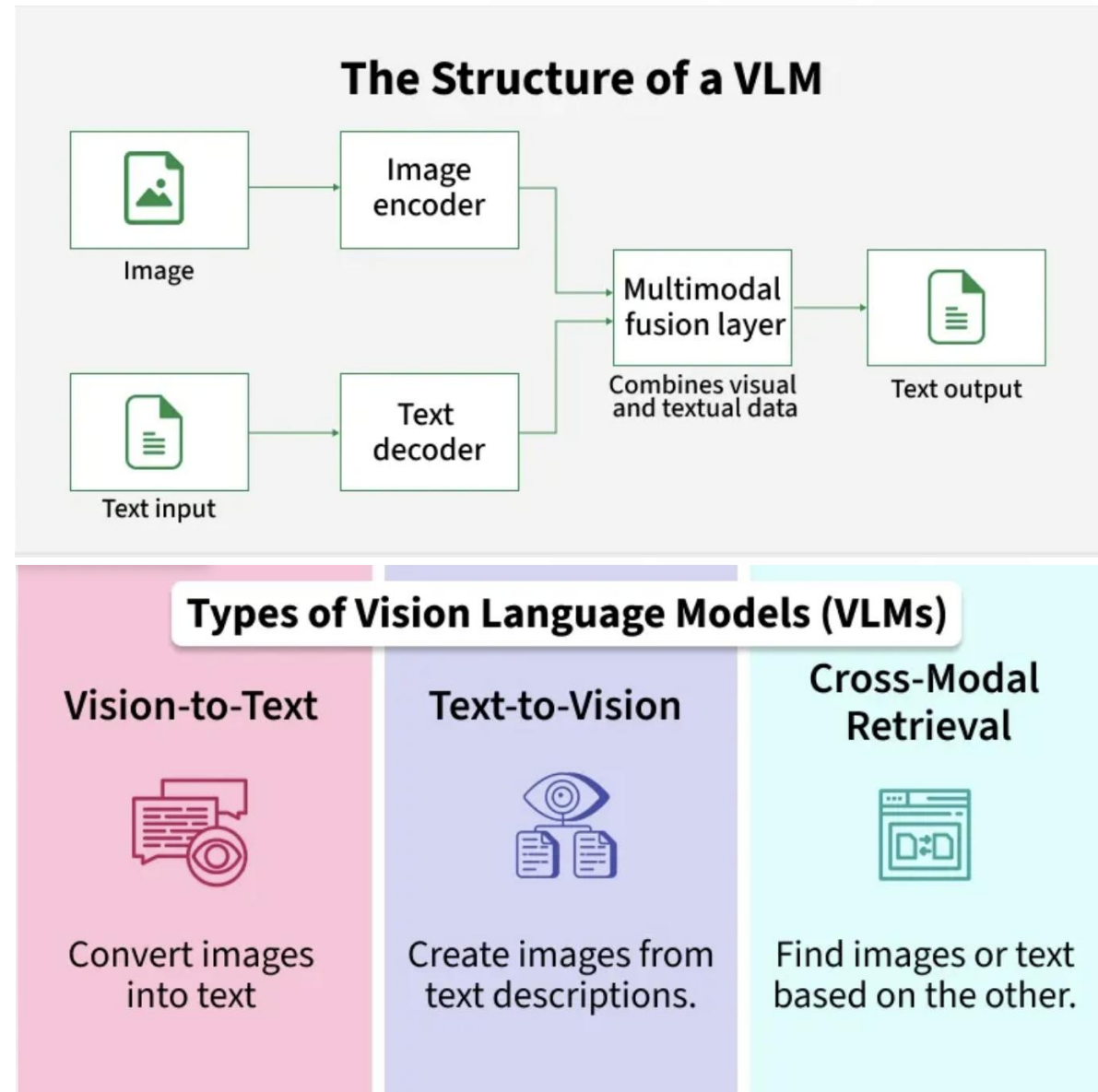
State of the Art & Use of Vision Language Models (VLMs) to Support Enterprise AR

Category: Software

**This topic is at the
intersection of
AR and AI**

You've heard of LLMs, but what are Vision Language Models?

- Vision-language models (VLMs) are AI systems that combine a large language model with a vision encoder to process and understand both image and text inputs.
- There are many approaches
- To learn more, visit
 - [this July 2025 blog post](#)
 - [this November 2024 blog post](#)



Problems this Research Would Address

- The field of Vision Language Models (VLMs) is emerging from research and could be maturing sufficiently to improve automatic AR experience delivery
- **Shortage of detailed knowledge** of these developments prevents customers and technology providers from making informed decisions.
 - Without published industry requirements regarding architectures, evaluation strategies, and mainstream datasets used in developing VLMs, as well as the key challenges and future trends in AR projects, products or services may be overlooked or not developed.
 - Investments in VLM hardware or software may not be fully leveraged if they are rapidly out of date, or inappropriate to target use cases or environments.

Whose problem would be addressed?

- **Enterprise decision makers** Choosing AR hardware and software without reliable VLM benchmarks means budget commitments are based on vendor claims, not evidence. This research gives buyers a neutral framework to evaluate options before they commit.
- **Providers of AR software or services** VLM technology is moving fast. Without current, verified information, providers risk building on platforms that will not scale or that misalign with enterprise conditions. This research gives them the data to prioritise features and select the right technology partners.
- **Hardware manufacturers** Device makers need to know which VLM architectures enterprise customers will require. Without that signal, hardware roadmaps are built on assumptions. This research surfaces those requirements directly from the enterprise community.

How would this research be conducted?

1. Perform **desk research** to document latest advances in VLMs for AR and related fields. Identify those hardware or software VLM providers that specifically train for VLM use in enterprise conditions.
2. **Perform analyses** on research findings to identify trends, top categories of opportunity and risk, and document these in a report
3. **Compile samples** (videos and running code in public GitHub repositories) and instructions for their use (including any hardware) that members can follow to test and evaluate VLMs (including open source)
4. **Develop a decision support tool** for members to assess if/where there are suitable options for VLM use in enterprise AR deployments or products and services

Project Deliverables

- **Report of research findings** describing current state of the art and future trends in VLM technology
- Approaches for VLM evaluation (with visual and text instructions) that specifically generate performance measurements (benchmark)
- **Decision support tool for members** to use when identifying the most suitable options for introducing VLMs in existing or new projects
- **Executive summary** of findings for public release and **member exclusive webinar**

Benefits to AREA Members

- **Increased knowledge**, awareness and insights into VLMs and trends to watch and how to use these will permit better informed decisions and investments
- Through identification of **use case and requirements**, use of a tool to evaluate VLMs that are suitable for enterprise AR conditions
- **Identification of potential technology partners** that will differentiate future products and services by way of high performance/low latency VLMs
- Provides an opportunity for AREA members to lead the market in defining the creation of benchmarks and documentation of performance (accuracy, reliability and resilience).