# Multimodal Models

-Multimodal models combine ML techniques like LLM + Image recognition in one system. Ex GPT4 with that you can add a image to the context along with input text for the LLM together they pass information and give a better answer this can also be in the form of speech or redios .etc and LLM dont even need to be involved but are very popular and can make:

Multimodal Agents: which are Multimodel models + agents Like GPT 4o which is a agent + can take img, speech for ex for can ask it "Best gpu under 500$ that looks like this" and attach a picture of a gpu you like.

# MCP ( model context protocall)

MCP = a design pattern or framework that structures how LLM's interact with context, memory, tools and environments in a clean and extensible way (not official yet) ⇒ uses Agents

-LLM Alone take text in → generate text But lack structure for tools, long term memory, manging goals etc MCP does it

MCP includes

| component | |
|---|---|
| Model | the LLM |
| context | all info like past messages |
| protocol | Defines how to structure input output + tool calls |
| tool layer | Allows LLM to make API calls |
| Memory layer | stores long term knowledge |
| Planner/agent | optional (for multi step workflows) |

How it works

1) Prompt → passed into protocall
2) protocall → packages it with memory, context, goal, tool list
3) LLM → received this input and decides
   - ans directly ? - call a tool ? etc
4) protocal handels action and update
5) final response returned to user