Core VLM Components

OCR

- Optical Character Recognition
- A technology that enables machines to read and extract text from images or scanned documents.
- Use in VLMs: Allows models to understand and process text inside images, such as signs, forms, or screenshots.
- Used in Kosmos-2 and ChatGPT Vision to read charts, menus, or handwritten notes.

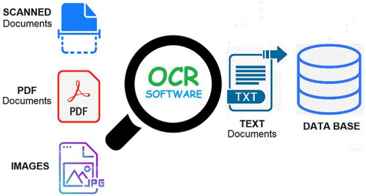

Figure 13: Optical Character Recognition