

Integration of the Semantic Extractor with a Multimodal Chatbot: Use Case in LegalTech

Scenario:

In a law firm, lawyers need to interact with multiple document formats: scanned paper documents, images, text, and, in some cases, audio. A multimodal chatbot combined with a semantic extractor based on OCR + Computer Vision + LLM allows for the comprehensive processing of all these data types, facilitating access to and analysis of legal information in any format.

How Integration Works in a Multimodal Chatbot

1. Multimodal Interaction:

- Lawyers can interact with the chatbot in different ways:
 - Text: Entering queries directly in text.
 - Image/Documents: Uploading images or scanned documents (contracts, judgments, etc.).
 - Audio: Providing audio files (testimony recordings, statements, etc.).

2. Processing Information by Modality:

- Images and Documents: The chatbot uses the semantic extractor with OCR and Computer Vision to convert images into text, analyze the document structure, and extract the requested information. For example, the chatbot can be asked to search for a specific clause in a scanned contract or a resolution in a judgment.
- Text: The chatbot can directly handle text queries, connecting with the LLM to provide complex legal answers or summaries of key information.
- Audio: If the provided document is in audio format (e.g., testimony recordings), the multimodal chatbot uses a speech recognition layer to transcribe the audio and then applies the LLM to analyze and extract relevant information.

3. Multimodal Response and Follow-up:

- The chatbot can return responses in multiple formats. For example:
 - Text: Provides a summary or detailed information directly.
 - Image: Visually marks the relevant sections of a scanned document.
 - Audio: Responds verbally if the lawyer is interacting in audio format, improving accessibility.

4. Automation and Continuous Learning:

- The chatbot can store and learn from multimodal interactions, facilitating the processing of repetitive query types or similar documents.
- It can remember previous documents it has processed and continue extracting new information at the user's request.

Advantages of Integration in a Multimodal Chatbot

1. Natural and Flexible Interaction:

- Being multimodal, the chatbot allows lawyers to work with the modality they prefer, whether text, images, documents, or audio. This provides a more comprehensive and accessible user experience.

2. Comprehensive Processing of Different Formats:

- The combination of OCR, Computer Vision, and LLM in a multimodal environment allows for processing both scanned documents and images and audio, making it ideal for environments where legal documentation may be in various formats.

3. Personalized Responses Based on Medium:

- The chatbot can adapt to the input and output modality, offering personalized responses:

- For images, it can visually mark and highlight important parts.
- For audio, it can transcribe, analyze, and provide summaries in text or audio.
- For text, the chatbot uses its generation capability with the LLM to provide deeper legal analyses.

4. Optimization of Legal Review Time:

- With multimodal capability, lawyers can process documents and testimonies faster without needing to manually convert formats. A scanned document can be processed automatically to obtain key clauses, while an audio recording can be transcribed and analyzed in minutes.

5. Greater Efficiency in Complex Cases:

- In complex legal cases that require the analysis of multiple document types (contracts, judgments, audio recordings), the multimodal chatbot can handle the entire process in an integrated manner, accelerating the review of information and improving decision-making.

6. Accessibility and Flexibility in Work Methods:

- Lawyers can interact with the system from any device (PC, mobile) and through different means (typing, uploading documents, or dictating). This makes reviewing legal information much more accessible, even in situations where office access is not possible.

7. Scalability:

- The system is highly scalable and can handle massive volumes of multimodal data, making it suitable for law firms of any size, from small practices to large corporations with thousands of documents.

Example Workflow in a Multimodal Chatbot

- Image of a Contract: The lawyer uploads an image of a scanned contract.

- Chatbot: "What information would you like to extract from the contract?"

- Lawyer: "Search for the resolution clause."

- Chatbot Response: "The resolution clause is on page 5, paragraph 4, which states that...".

- Audio of a Testimony: The lawyer uploads an audio recording of a testimony.

- Chatbot: "Transcribing and analyzing the testimony..."
- Chatbot Response: "The testimony mentions that the defendant confirms that... at minute 3:25."

This integration allows the multimodal chatbot to be a key tool for law firms, facilitating the review of documents and audio more quickly and accurately, while providing a more efficient workflow in the analysis of complex legal cases.