



ARTICLE

# Transformation of Verbal Descriptions of Process Flows into Business Process Modelling and Notation Models Using Multimodal Artificial Intelligence: Application in Justice

Silvia Alayón<sup>1,\*</sup>, Carlos Martín<sup>1</sup>, Jesús Torres<sup>1</sup>, Manuel Bacallado<sup>1</sup>, Rosa Aguilar<sup>1</sup> and Guzmán Savirón<sup>2</sup>

<sup>1</sup>Department of Computer and Systems Engineering, University of La Laguna, San Cristóbal de La Laguna, 38200, Spain

<sup>2</sup>Provincial Court of Las Palmas de Gran Canaria, Section 4 (Commercial), Las Palmas de Gran Canaria, 35016, Spain

\*Corresponding Author: Silvia Alayón. Email: salayon@ull.edu.es

Received: 19 September 2025; Accepted: 04 December 2025; Published: 26 February 2026

**ABSTRACT:** Business Process Modelling (BPM) is essential for analyzing, improving, and automating the flow of information within organizations, but traditional approaches based on manual interpretation are slow, error-prone, and require a high level of expertise. This article proposes an innovative alternative solution that overcomes these limitations by automatically generating comprehensive Business Process Modelling and Notation (BPMN) diagrams solely from verbal descriptions of the processes to be modeled, utilizing Large Language Models (LLMs) and multimodal Artificial Intelligence (AI). Experimental results, based on video recordings of process explanations provided by an expert from an organization (in this case, the Commercial Courts of a public justice administration), demonstrate that the proposed methodology successfully enables the automatic generation of complete and accurate BPMN diagrams, leading to significant improvements in the speed, accuracy, and accessibility of process modeling. This research makes a substantial contribution to the field of business process modeling, as its methodology is groundbreaking in its use of LLMs and multimodal AI capabilities to handle different types of source material (text and video), combining several tools to minimize the number of queries and reduce the complexity of the prompts required for the automatic generation of successful BPMN diagrams.

**KEYWORDS:** Process modelling; verbal description; BPMN; LLM; multimodal AI

## 1 Introduction

Business process modelling (BPM) is essential for analysing, improving and automating the flow of information within organisations. It provides answers to the questions “Who does what, when, and how?” for each activity, helping organisations to understand and structure their workflows more effectively. A key tool in this field is Business Process Model and Notation (BPMN), an internationally standardised modelling language [1,2].

Today, most organisations model their processes to ensure efficiency and improve quality [3]. However, complex processes are often unclear and require considerable time and resources for analysis [4]. Traditional modelling approaches have significant limitations. They rely on manual interpretation by specialised engineers, which makes them slow and error-prone processes requiring a high level of expertise. The inherent complexity of organisational processes, coupled with the need for interdisciplinary collaboration, can result



in communication issues, domain-specific terminology and unrealistic expectations, all of which make it challenging to create accurate models [5].

To overcome these limitations, this paper presents an innovative solution that builds on recent advances in artificial intelligence (AI). We present a methodology that can automatically generate complete and accurate BPMN diagrams from verbal descriptions of processes obtained directly from video recordings of experts. This methodology uses Large Language Models (LLMs) and multimodal AI.

This research is novel in several key respects:

- Direct modelling from video to BPMN: unlike previous works, which required verbal descriptions to be transcribed into text first, our methodology uses the multimodal nature of the latest LLMs to work directly with video recordings of the expert, eliminating the need for pre-processing steps.
- Optimised combination of LLMs: We propose combining two LLMs, NotebookLM (based on Gemini) and ChatGPT (based on GPT-4), to minimise the number of queries and the complexity of the prompts required. This achieves greater efficiency and accuracy in model generation.

An LLM is a deep neural network-based system that has been trained on a large corpus of text data to interpret natural language and generate human-like responses. LLMs can be used for a variety of natural language processing tasks, including language translation, text summarisation, answering questions and generating content [6]. Another important advantage of LLMs is their multimodal nature. They can process various inputs, not only text but also images or audio, in the form of instructions, and convert them into different types of output. Prominent examples of LLMs include OpenAI's GPT series [7], Google's Gemini [8] and Anthropic's Claude [9]. We consider LLMs to be suitable for process modelling because they can generate structured results from textual descriptions. As stated in [10], using LLMs for process modelling represents a transformative change as it significantly reduces the need for manual effort and specialised knowledge.

This work aims to develop and validate a replicable methodology that can efficiently and accurately transform verbal descriptions of workflows into BPMN models by leveraging the capabilities of multimodal AI systems. This methodology was evaluated by generating BPMN models, which were then reviewed and validated by a subject matter expert. While the specific application is in the field of Commercial Courts in Spain, the methodology can be generalised and applied to other organisations. This research makes a significant contribution to the field of business process modelling by offering an innovative solution that improves the speed, accuracy and accessibility of the modelling process.

## 2 Related Work

Although business process modelling has evolved considerably, challenges remain in accuracy, efficiency and accessibility. Traditional methods depend on manual interpretation by specialists, making them slow and error-prone. Over the past two decades, researchers have explored various strategies to automate the transformation of textual or verbal process descriptions into formal models.

### 2.1 Rule-Based and Business-Rule Approaches

Early automation efforts relied on predefined business rules to map textual descriptions onto process structures. The work [11] introduced an event-condition-action (ECA) framework that formalised business rules as the basis for workflow construction, establishing explicit mappings between rules and control-flow fragments. Ref. [12] performed a representational analysis of process- and rule-modelling languages, identifying expressiveness overlaps and limitations in capturing constraints through rules alone. Later, Refs. [13,14] proposed algorithms that generate BPMN models from Attribute-Relationship Diagrams

(ARDs) and rule sets, demonstrating the feasibility of automatic model synthesis but also revealing low flexibility when handling ambiguous or incomplete descriptions.

These rule-based methods were foundational, yet their dependency on manually engineered rule sets made them brittle and difficult to scale to natural-language input.

## **2.2 Grammar-Based and Classical NLP Approaches**

A subsequent research line applied Natural Language Processing (NLP) to extract structured information from text. The work [15] demonstrated how part-of-speech tagging and syntactic parsing can identify activities, roles and decisions in textual process descriptions. Ref. [16] reversed the direction by generating natural-language explanations from process models to support model validation, contributing to the integration of textual and diagrammatic representations. The authors of [17,18] refined extraction pipelines by combining grammar rules with semantic analysis to detect tasks and events. More recently, Ref. [19] extended these pipelines to user stories, showing that data-driven methods reduce manual rule design but remain sensitive to linguistic ambiguity.

Overall, traditional NLP approaches improved recall and portability but often missed implicit steps or contextual dependencies, limiting their reliability for end-to-end automation.

## **2.3 Large Language Models (LLMs) and Generative AI**

The emergence of transformer-based LLMs has transformed process-model generation. A recent systematic review by [20] provides a comprehensive overview of how Generative AI is being adopted across Business Process Management. Their analysis shows a growing interest in using LLMs for process modelling, documentation and optimisation, but also highlights that current approaches are predominantly text-based and lack multimodal capabilities.

For example, Ref. [10] proposed ProMoAI, a framework that employs LLM prompting strategies and human-in-the-loop validation to generate and refine BPMN models automatically. Ref. [21] investigated multimodal representations for BPMN analysis and showed that hybrid textual-visual encodings enhance interpretability. Other case study [22] confirmed that LLMs can produce executable model drafts with minimal manual engineering, though issues such as hallucinated control flows and incomplete step coverage persist.

A recent contribution in this direction is GAI4BM by [23], which uses generative AI to transform textual descriptions into business process models. Their work demonstrates the feasibility of LLM-based automation for BPM tasks; however, their approach is limited to text-only inputs, whereas real-world procedures (particularly in judicial or operational domains) are often verbal, multimodal, and embedded in contextual explanations. Ref. [24] investigates the applicability of local LLMs to business process modelling tasks, demonstrating that self-hosted models can effectively generate process structures without relying on cloud services. Their work provides evidence that LLM-based BPMN generation can be decentralised; however, their approach requires high computational resources and technical setup.

A very recent work, Ref. [25] introduced BPMN Assistant, an LLM-based tool hosted on Hugging Face that generates BPMN diagrams from textual instructions. While their approach demonstrates the feasibility of using LLMs for interactive BPMN creation, it remains strictly text-driven and does not incorporate multimodal inputs. Another recent contribution is BPMN-Chatbot++ by [26], which extends LLM-based BPMN generation to collaboration diagrams and adds support for data objects and message flows. This system illustrates how LLMs can handle more sophisticated BPMN structures beyond simple activity flows. However, the approach relies exclusively on structured textual prompts and does not incorporate multimodal

inputs such as audio or video descriptions, which limits its applicability in domains where processes are conveyed verbally rather than through formal documentation

Recent literature therefore portrays LLMs as powerful assistants for first-draft model generation, provided that outputs are validated through symbolic or human review.

## ***2.4 Industrial Adoption and AI-Enabled BPM Tools***

The practical influence of LLMs is also visible in commercial BPM platforms. Bonita BPM now includes an AI-based BPMN generator, Camunda integrates AI-enabled process orchestration, and Bizagi has introduced AI agents for process automation [27–29]. These examples illustrate a clear industry trend toward embedding generative capabilities into business-process-management suites.

However, vendor documentation is not peer-reviewed and should be regarded as evidence of industrial adoption rather than scientific validation. Independent academic studies provide systematic evaluations confirming the feasibility of AI-assisted BPMN generation while highlighting limitations related to model reliability and semantic completeness [10,21].

## ***2.5 Multimodal Models and Diagram Interpretation***

Recent multimodal systems [21] extend LLMs to handle both textual and visual information. Experiments with multimodal GPT-4 demonstrate that such models can interpret and explain existing BPMN diagrams, but their use for generating diagrams directly from raw video or audio inputs is still in the early stages of development.

Hybrid architectures that combine multimodal ingestion with symbolic post-processing are emerging as promising directions for improving robustness and transparency.

## ***2.6 Modelling Judicial Workflows and Legal Processes***

In the legal and justice domains, research has focused on formalising judicial workflows using traditional modelling methods. Ref. [30] developed a use-case-driven approach for describing judicial business processes, while [31] coupled BPMN with ontologies to capture legal constraints and exceptions. More recent works such as [32,33] confirm the utility of BPMN for representing legal procedures, though they still rely on manual modelling.

These studies demonstrate the relevance of BPMN in legal contexts but do not employ LLMs or multimodal AI. Meanwhile, advances in generative process modelling show that LLM-based pipelines can already produce structured BPMN outputs from textual descriptions [10,21]. Although not specific to the judicial field, such frameworks could be adapted to legal processes if complemented with domain ontologies and retrieval-based validation.

Therefore, although no work has yet been published on the complete automatic generation of judicial procedure models based on LLMs, general-purpose LLM frameworks now provide the methodological basis for extending these capabilities to justice and public administration contexts, provided the necessary domain knowledge and expert validation are included.

## ***2.7 Summary***

Overall, prior studies established a progression from rule-based systems and grammar-driven NLP toward LLM-based and multimodal approaches. The most recent research confirms the feasibility of using generative models for BPMN creation but also underlines the need for hybrid verification. Within this landscape, the present work contributes by introducing a two-stage multimodal pipeline, combining

NotebookLM (Gemini) and ChatGPT (GPT-4), to generate BPMN models directly from verbal video descriptions, filling a gap between existing text-based LLM studies and the yet-unexplored field of multi-modal process modelling in judicial environments.

### 3 Tools

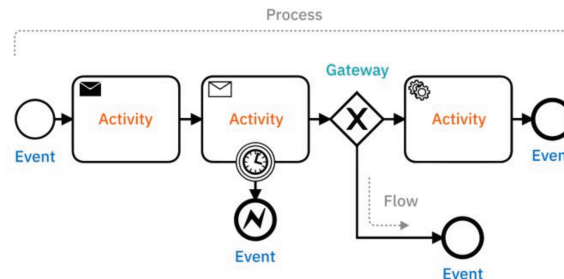
#### 3.1 Business Process Model and Notation (BPMN)

BPMN is widely used in business process management to translate high-level workflows into detailed, executable models that can be directly implemented in process-management systems. Its formal specification, maintained by the Object Management Group (OMG) [1], provides standardised semantics that enable interoperability across modelling tools and workflow engines. Numerous studies have confirmed BPMN's dominant role as a *de facto* standard in both academia and industry for bridging conceptual and executable representations of business processes [2,3,21].

The latest version of the standard, BPMN 2.0, incorporates additional elements that improve the interoperability and executability of modelled processes. These elements are represented by specific shapes to make the diagrams easier to interpret. The main ones are as follows:

1. Event: This defines a trigger that initiates, modifies or terminates a process.
2. Activity: A task (action) performed by technological or human means.
3. Gateway: This describes a decision point (branch) in a workflow. It can be parallel (AND), inclusive (OR), or exclusive (XOR).
4. Flow: This establishes a connection between elements.

Fig. 1 illustrates these elements graphically.



**Figure 1:** Example of BPMN diagram. Image extracted from Ref. [34]

#### 3.2 ChatGPT

ChatGPT, acronym for Chat Generative Pre-trained Transformer is a chat system developed by OpenAI based on its GPT models [7]. It started as a language model, and is, at present, a revolutionary tool in terms of functionality, accuracy and versatility. GPT can provide increasingly complex, reliable and useful responses and generate texts, summaries and translations with great accuracy, simulating human language. However, ChatGPT also has significant limitations. It can give incorrect answers, be overly sensitive to differences in the wording of questions (the prompt) and may not consistently clarify ambiguous questions, as indicated by the authors themselves [7].

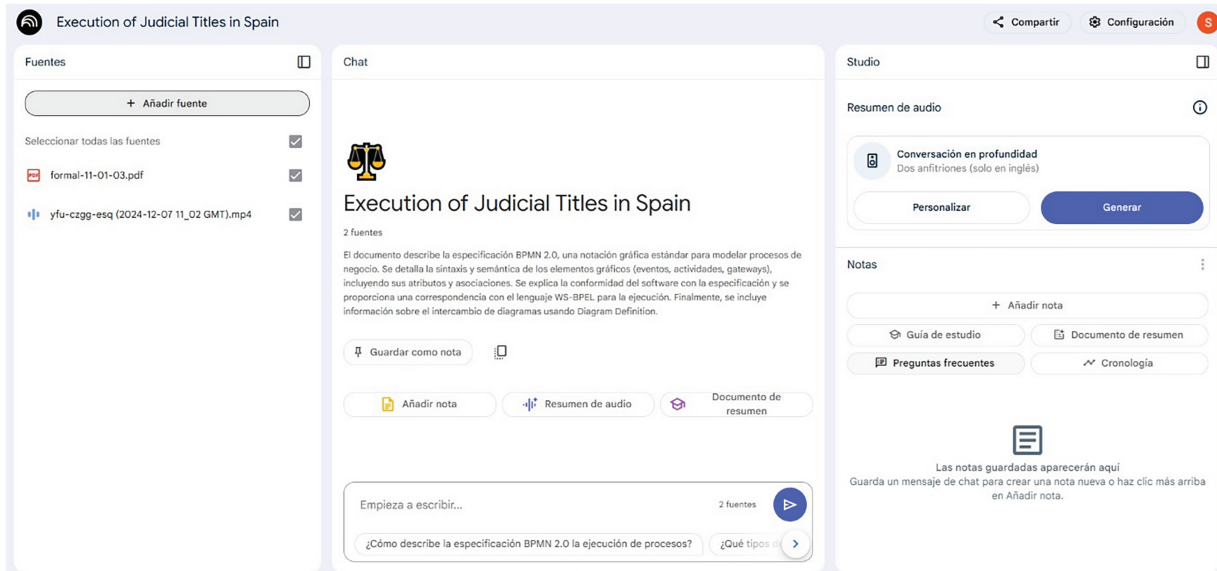
In this work, we used the GPT4 version [35]. ChatGPT-4 is a significant advancement in AI compared to previous models, offering improvements in areas such as context understanding, personalisation, text generation, multimodal integration (text and images), security and ethics, and linguistic capabilities.

### 3.3 NotebookLM

Developed by Google, NotebookLM [36] represents a significant advance in user-centred AI. Its name indicates its purpose: to become an AI-powered notebook. This tool enables users to learn and gain a better understanding of a wide range of topics from the documents and sources they enter. NotebookLM can generate personalised answers and assist with tasks such as organising ideas, generating summaries, elaborating proposals and even creating podcasts, all of which are adjusted to the user's particular context.

NotebookLM uses Google's Gemini model [8]. The system enables users to upload various documents, ranging from scientific papers to personal notes, in order to train the AI with specific information. Files uploaded to the tool can be images, videos or audio files, making the tool multimodal. Unlike chatbots such as ChatGPT or Gemini, NotebookLM only works with the documents supplied by the user, making it more difficult for it to fabricate false information or make mistakes. It is confined to the information with which the user supplies it, and every time it provides an answer, it shows the source from which it obtained that answer. This makes its answers highly reliable, since users control the sources on which they rely.

Fig. 2 shows an example of NotebookLM in use. The left side shows the sources (in this example, a PDF document and a video), the right side shows where audio summaries (podcasts) of uploaded content can be generated, and the centre shows a brief description of the source content, suggestions for related questions, and the prompt editing field.



**Figure 2:** Example of the use of NotebookLM. Screenshot of one of the tests performed in this work

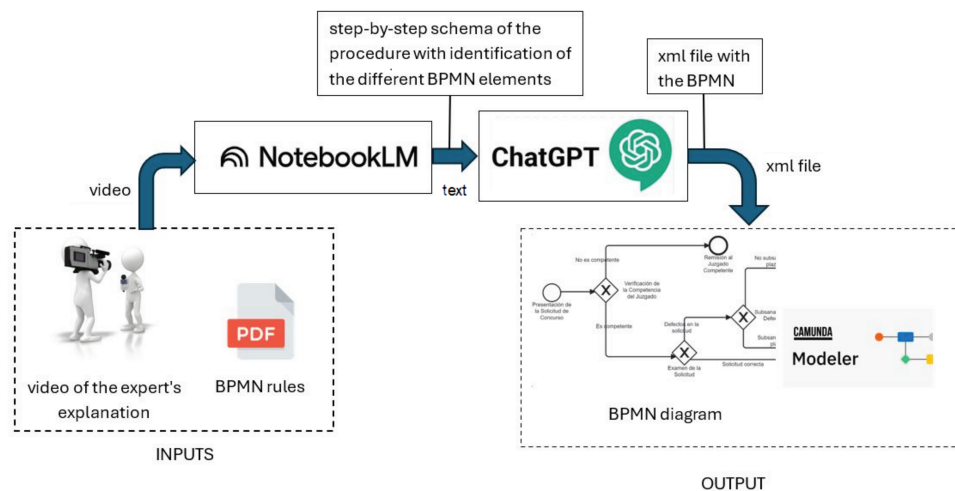
## 4 Methodology

This section outlines the proposed methodology for automatically generating BPMN process models from expert verbal descriptions captured on video. This approach combines large language models (LLMs) with process modelling standards to produce consistent and verifiable results. To ensure methodological transparency and reproducibility, the full pipeline, model configurations, prompt design and output validation steps are described.

#### 4.1 System Architecture

Fig. 3 shows the architecture of the proposed system. The workflow is organised into five sequential stages:

1. Video input acquisition: domain experts verbally describe a legal procedure, which is captured as a video recording.
2. Multimodal analysis with NotebookLM (Gemini): the raw video is ingested without prior transcription alongside an explanatory document on the BPMN 2.0 standard, and the model extracts a structured representation of the procedure.
3. Intermediate schema generation: NotebookLM generates a step-by-step schematic description of the process, explicitly identifying activities, events, and decision points in a format aligned with BPMN 2.0 constructs. This textual schema constitutes the intermediate representation used as input for the second LLM (ChatGPT).
4. BPMN XML generation with ChatGPT (GPT-4): The schema generated by NotebookLM is converted into a BPMN-compliant XML file that includes coordinates and layout information.
5. Diagram rendering: The XML file is imported into Camunda Modeler to generate the final BPMN diagram.



**Figure 3:** Proposed methodology

Preliminary experiments showed that a single model could not handle multimodal video input and XML generation with sufficient accuracy simultaneously, so this two-stage LLM pipeline was selected. Furthermore, this approach minimises the number of queries to each LLM and reduces the complexity of the prompts.

#### 4.2 Data Acquisition

Expert knowledge is captured through video recordings of detailed explanations of judicial procedures by experts in the field. The context of application is the Commercial Courts in Spain, whose powers are defined in Article 86 bis of the 1985 Organic Law on the Judiciary.

In collaboration with a magistrate specialising in commercial matters, 20 internal procedures were identified and the five most frequent were selected for this study. These procedures are:

1. Insolvency of a natural person who is not a business owner and has no assets (video length: 32 min).

2. Payment order procedure (video length: 18 min).
3. Mortgage foreclosure (video length: 15 min).
4. Ordinary trial (video length: 14 min).
5. Enforcement of judicial order (video length: 7 min).

The videos differ in length because each judicial procedure varies in complexity and therefore requires more or less time for the expert to describe it thoroughly. The original recordings were captured in high resolution, resulting in file sizes between approximately 113 and 565 MB. To reduce processing time in NotebookLM, the videos were later down-scaled, resulting in files of between approximately 5 and 12 MB. This adjustment did not involve altering content, segmenting the recordings, or modifying the audio track; it affected resolution only.

In each video recording, the expert described the full procedure verbally, using natural language and domain-specific legal terminology, speaking directly to the camera without using any additional visual aids (e.g., slides, whiteboards, documents, drawings), and emphasising the ‘who’, ‘when’ and ‘how’ of each step. This allows the expert to express himself freely without adapting their speech, testing the ability of LLMs to go beyond the level of specialisation required for knowledge extraction. The videos are fed into the selected LLMs without any prior pre-processing, such as transcription, segmentation or audio filtering.

Although the expert in our case study spoke directly to the camera without using visual or graphical elements (making audio alone sufficient) video recordings were chosen for two main reasons:

1. Methodology generality: the goal is to design a framework that works in complex scenarios where expert explanations may include gestures, graphics, or other visual cues. Video demonstrates that the methodology supports truly multimodal inputs and is ready for cases where visual information may be critical.
2. True multimodal processing: the approach operates directly on the full audio–visual video without relying on intermediate transcription, validating the system’s multimodal capabilities and showcasing a more comprehensive and versatile framework beyond audio-only automatic speech recognition (ASR) processing.

Therefore, the added value of using video is not related to improving transcription quality for a verbal explanation, but rather to demonstrating that the methodology can operate directly on multimodal video data, thereby providing a more general and extensible solution.

#### **4.3 Model Selection and Configuration**

This methodology was implemented using two LLM-based tools, NotebookLM and ChatGPT, which were selected for their complementary capabilities:

- NotebookLM (Gemini): Developed by Google [36], it uses the Gemini model. Its main strengths are its multimodal nature and its ability to ingest user data sources directly, including videos, images, and audio. This makes it ideal for the contextual analysis of specific documentation and multimodal content. Unlike other chatbots, NotebookLM adheres to the information provided by the user, reducing the likelihood of inventions or errors and allowing the sources of its responses to be traced. In our preliminary tests, NotebookLM was capable of generating a step-by-step diagram of complete procedures from a video in a single query and could even identify BPMN elements if the BPMN standard was provided as an additional source.
- ChatGPT (GPT-4): This chat system from OpenAI is based on the GPT family of models. GPT-4 [35] was chosen because of its ability to generate complex, reliable and useful responses with high accuracy, producing texts, summaries and translations that simulate human language. ChatGPT excels at

generating XML files and following complex instructions. Initial testing revealed that, while ChatGPT-4 cannot accept video directly and requires audio transcription, and while it needed multiple queries to generate BPMN diagrams with initial failures, it was effective in generating syntactically correct XML code from a textual schema.

No fine-tuning of the models was performed explicitly for this application; instead, their pre-trained capabilities and prompt engineering were relied upon to guide their behaviour.

#### 4.4 Prompt Engineering

To optimise interaction with the tools and ensure the efficient generation of BPMN models, two specific prompts were designed, one for each LLM. These prompts are simple and universal, meaning they can be applied to all judicial procedures, regardless of complexity, to obtain effective responses.

The exact prompts used are as follows:

1. Query to NotebookLM: *“Please create a TD graph that represents all the stages that make up the procedure described in the video XXXX from its beginning to its end and explicitly indicate what decides the transitions between nodes. Identify well the different elements of the process so that the graph diagram you present to me follows the BPMN 2.0 format whose rules you can consult in formal-11-01-03.pdf”.*
2. Query to ChatGPT: *“I am going to provide you with a procedure represented in a TD network. Please generate an XML file that can be converted into a BPMN 2.0-formatted graph in Camunda Modeler. You need to assign physical coordinates to each element and provide a text explaining what happens in each branch of the graph. The TD graph is as follows: XXXX” (the schema that NotebookLM generated in the first query is introduced here).*

NotebookLM’s response to the first query (the procedure outline) is inserted into the ‘XXXX’ field of the second ChatGPT query. These prompts are the result of several experiments and are designed to maximise the capabilities of each tool. Using a standardised NotebookLM prompt ensured consistent extraction of events, activities, gateways and flows, and using a ChatGPT prompt enabled the generation of reproducible XML with layout information. The XML files produced by ChatGPT were imported into Camunda Modeler for visualisation.

#### 4.5 Output Processing and Validation

The XML files produced by ChatGPT were imported into Camunda Modeler for visualisation and validation purposes. The resulting BPMN diagrams were assessed according to two criteria:

1. Syntactic validity: The XML files had to comply with the BPMN 2.0 schema definitions.
2. Semantic correctness: the diagrams were compared with the domain expert’s knowledge of the procedure to confirm that all steps, actors, and decision points were accurately represented.

Any corrections suggested by the domain expert were recorded for evaluation purposes (see [Section 6](#)).

### 5 Experimental Development

This section outlines the experimental design and evaluation of the proposed methodology. The aim was to establish whether a combination of multimodal AI and LLMs could produce BPMN diagrams that are accurate, complete and comparable to those generated manually by experts.

### 5.1 Experimental Design

The experimental design was based on applying the methodology to a representative set of judicial procedures, as well as validation by an expert:

- Judicial procedures: The five most frequent procedures in Spain's commercial courts were selected from twenty identified procedures in total. These procedures vary in complexity and duration.
- Model generation: For each of the five procedures, the NotebookLM + ChatGPT methodology was applied to generate the corresponding BPMN diagram from the expert's video recordings.
- Expert evaluation: The five BPMN models generated were thoroughly reviewed and approved by a magistrate specialising in commercial matters and collaborating on this research. The expert assessed whether the models were correct and adequately represented the procedures. They checked that all the steps in the generated BPMN graphs corresponded to the actual steps in the analysed process, and that the transitions between steps and the semantic fidelity were accurate.

### 5.2 Evaluation Metrics

To ensure rigour, the evaluation included the following metrics:

- Step Accuracy (SA): the percentage of process steps in the generated diagram that correctly match those in the expert reference.
- Step Recall (SR): the proportion of reference steps that were correctly captured in the generated diagram.
- F1 score: harmonic mean of SA and SR, providing a balanced measure of accuracy and completeness.
- Structural correctness (SC): percentage of correctly generated BPMN elements (e.g., events, gateways and flows).
- Time efficiency (TE): the average time taken to produce a complete BPMN diagram.

These metrics provide a quantitative assessment of the efficiency, accuracy and structural correctness of the generated models.

## 6 Results and Validation

The proposed methodology was applied to the five selected judicial procedures in the experiments carried out. For each procedure, NotebookLM and ChatGPT were queried using the prompts outlined in [Section 4.4](#).

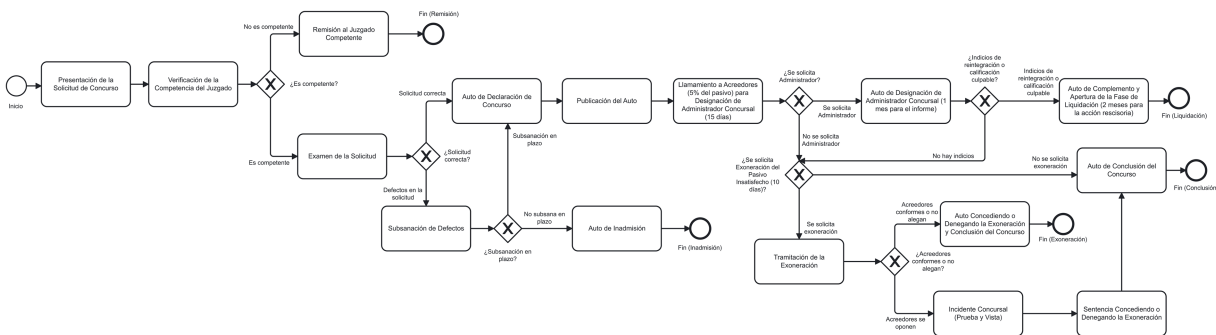
An example of the response obtained using these prompts is shown below for the first procedure considered: insolvency of a natural person who is not a business owner and has no assets. NotebookLM's response is shown in [Table 1](#). This response was entered into the ChatGPT prompt to obtain the procedure diagram in XML format (shown in [Table 2](#)), which can be displayed in a BPMN viewer. The resulting BPMN diagram is shown in [Fig. 4](#).

**Table 1:** Question entered and answer generated by NotebookLM for the first procedure. For space reasons, the answer is shown cropped where indicated by (. . .). As the query was conducted in Spanish, the answer has been translated into English for this article

| Prompt                                                                                                                                                                                                                                                                                                                                                                                                 | Answer                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Please create a TD graph that represents all the stages that make up the procedure described in the video procl.mp4 from its beginning to its end and explicitly indicate what decides the transitions between nodes. Identify well the different elements of the process so that the graph diagram you present to me follows the BPMN 2.0 format whose rules you can consult in formal-11-01-03.pdf.” | <p>Below is a BPMN 2.0 diagram representing the procedure described in the video:<br/> A[Submission of Request for Competition] --&gt; B{ Court Jurisdiction Verification}<br/> B -- Not Competent --&gt; C[Referral to Competent Court] . . .<br/> Taking into account that the graph must meet the following requirements:</p> <p>Events:</p> <ul style="list-style-type: none"> <li>• Initiating event: “A[Submission of Request for Competition]”. It is represented by a single circle.</li> <li>• Intermediate events: These are represented by a double-contour circle. In this case, they can be used to model decisions or bifurcation points of the process: “B{ Court Jurisdiction Verification}” . . .</li> <li>• End events: they are represented by a circle with a thick outline. In this case: “C[Referral to Competent Court]” (If the court is not competent) . . .</li> </ul> <p>Activities:</p> <ul style="list-style-type: none"> <li>• Tasks: they are represented as rectangles with rounded corners. In this process, the tasks would be: “F{ Auto de Declaración de Concurso}” . . .</li> </ul> <p>Gateways:</p> <ul style="list-style-type: none"> <li>• Exclusive (or XOR) gates: they are represented as diamonds with an “X” inside. They are used to model decisions with mutually exclusive paths. In this process, all decisions that are represented as bifurcations in the network would be modeled with XOR gates.</li> </ul> <p>Sequence flow:</p> <ul style="list-style-type: none"> <li>• Sequence flows: these are represented by arrows connecting the elements of the process. The sequence flow must follow the order described in the network, connecting events, gates and activities.</li> </ul> <p>Connections between the elements: . . .</p> |

**Table 2:** XML file generated by ChatGPT using the TD graph offered by NotebookLM. For space reasons, the answer is shown cropped where indicated by (...). As the query was conducted in Spanish, the answer has been translated into English for this article

| Prompt                                                                                                                                                                                                                                                                                                                                                                                 | Answer                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| I am going to provide you with a procedure represented in a TD network. Please generate an xml file to become a BPMN 2.0 formatted graph that can be represented in Camunda Modeler. I need you to assign physical coordinates to each element, and a text explaining what happens in each branch of the graph. The TD graph is as follows: (TD graph shown in Table 1 inserted here). | <pre> &lt;?xml version="1.0" encoding="UTF-8"?&gt; &lt;definitions xmlns="http://www.omg.org/spec/BPMN/20100524/MODEL" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:bpmndi="http://www.omg.org/spec/BPMN/20100524/DI" xmlns:omgdc="http://www.omg.org/spec/DD/20100524/DC" xmlns:omgdi="http://www.omg.org/spec/DD/20100524/DI" targetNamespace="http://bpmn.io/schema/bpmn" exporter="Camunda Modeler" exporterVersion="5.30.0"&gt; &lt;process id="Process_1" isExecutable="true"&gt; &lt;startEvent id="StartEvent_A" name=" Submission of Request for Competition "&gt; &lt;outgoing&gt;Flow_A_B&lt;/outgoing&gt; &lt;/startEvent&gt; &lt;exclusiveGateway id="Gateway_B" name=" Court Jurisdiction Verification "&gt; &lt;incoming&gt;Flow_A_B&lt;/incoming&gt; &lt;outgoing&gt;Flow_B_C&lt;/outgoing&gt; &lt;outgoing&gt;Flow_B_D&lt;/outgoing&gt; &lt;/exclusiveGateway&gt; &lt;endEvent id="EndEvent_C" name=" Referral to Competent Court "&gt; &lt;incoming&gt;Flow_B_C&lt;/incoming&gt; &lt;/endEvent&gt; &lt;/endEvent&gt; ... </pre> |



**Figure 4:** BPMN Model of the insolvency of a non-business person without assets procedure

The BPMN models obtained for each of the five legal proceedings analysed in this study are shown below: insolvency of a non-business person without assets (Fig. 4), payment order (Fig. 5), foreclosure (Fig. 6), ordinary trial (Fig. 7), and enforcement of judicial order (Fig. 8).

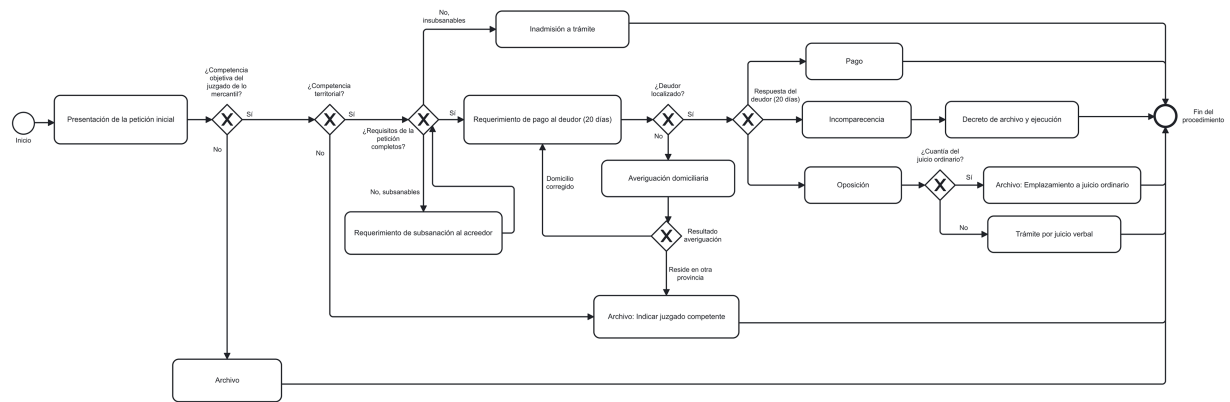


Figure 5: BPMN model of the payment order procedure

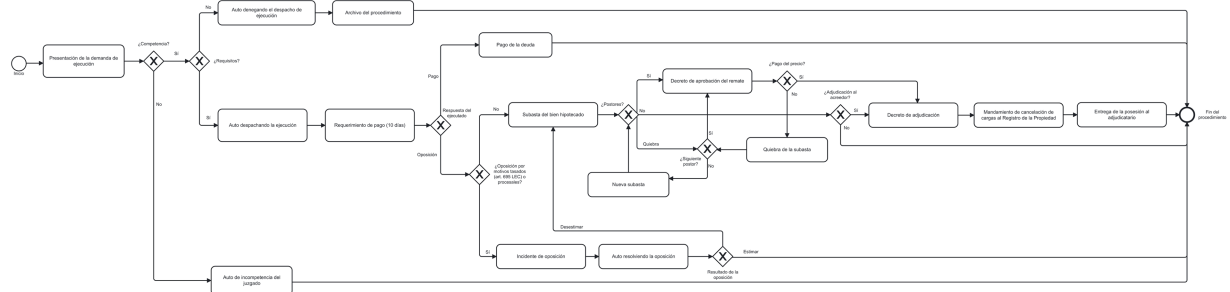


Figure 6: BPMN model of the foreclosure procedure

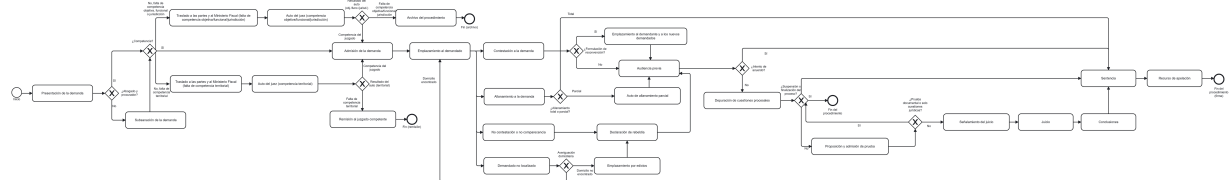


Figure 7: BPMN model of the ordinary trial procedure

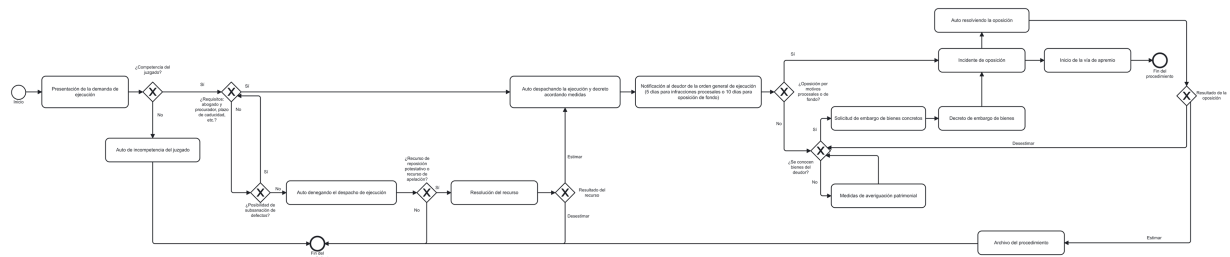


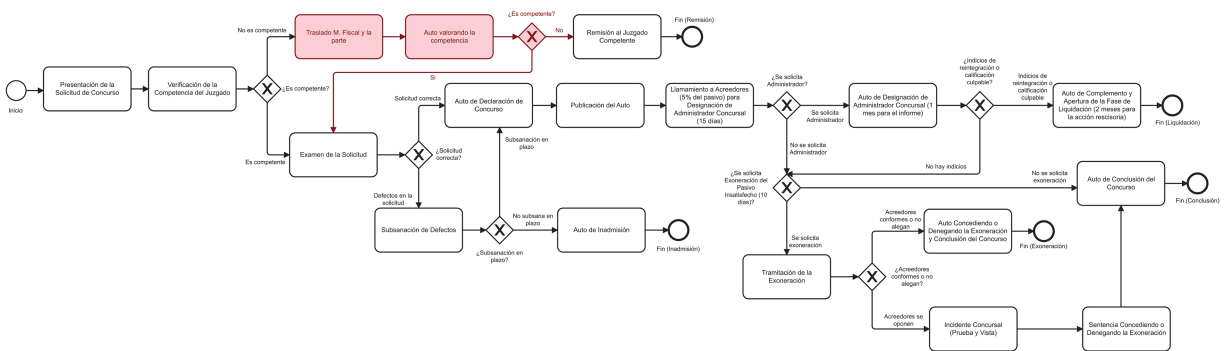
Figure 8: BPMN model of the enforcement of judicial order procedure

Table 3 shows the performance metrics of the proposed methodology for automatically generating BPMN models for the five judicial procedures considered. All these measurements were carried out under the supervision of the magistrate, who recorded the videos. It was his responsibility to verify that no steps were missing, that the system had not ‘invented’ any, and that the flows in the diagram were correct. Figs. 9–13

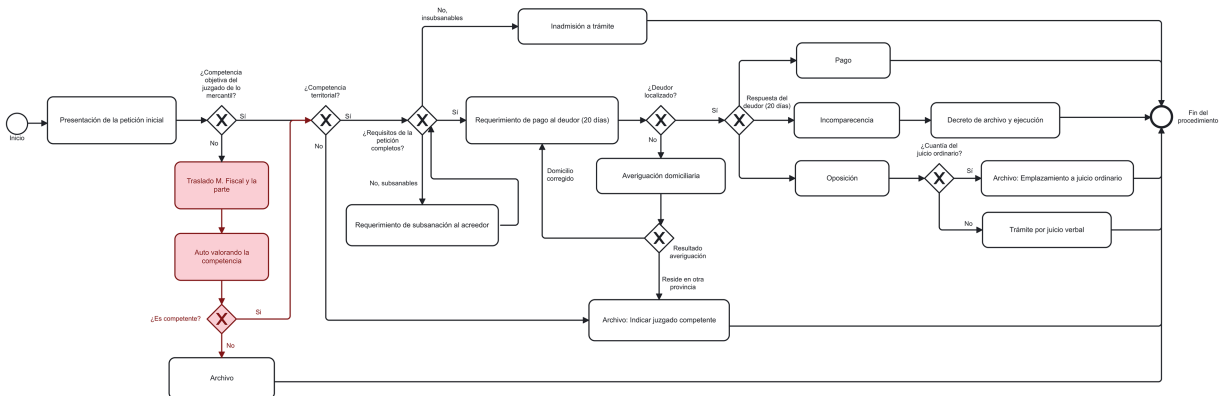
show the corrections that the expert made to the BPMN graphs generated by the proposed system. This made it possible to count the graphs' errors and successes and complete Table 3. For each procedure, the expert-corrected diagram shown in Figs. 9–13 constitutes the human baseline. All metrics in Table 3 are computed by comparing the generated model against this expert-validated reference.

**Table 3:** Performance metrics across five judicial procedures (SA: step accuracy, SR: step recall, F1: F1 score, SC: structural correctness, TE: time efficiency)

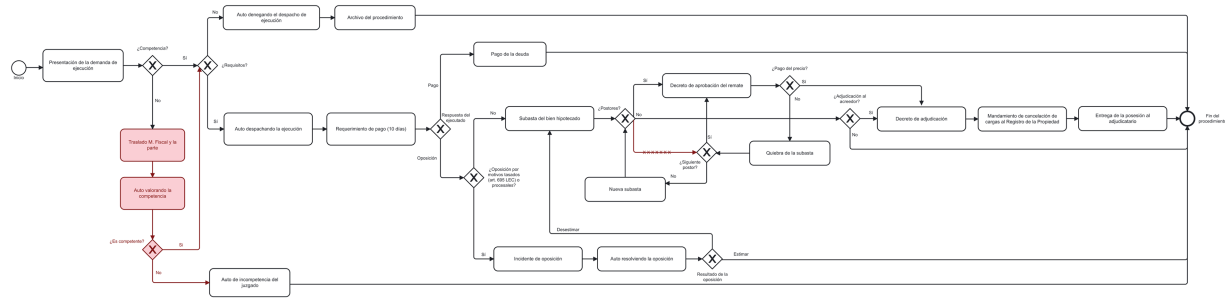
| Procedure                                       | SA (%) | SR (%) | F1 (%) | SC (%) | TE (min) |
|-------------------------------------------------|--------|--------|--------|--------|----------|
| Insolvency (non-business individual, no assets) | 100    | 84.21  | 91.38  | 95.00  | 5        |
| Payment order                                   | 100    | 86.66  | 92.89  | 100    | 3        |
| Foreclosure                                     | 100    | 88.88  | 93.99  | 95.00  | 4        |
| Ordinary trial                                  | 100    | 92,85  | 96,25  | 75.00  | 4        |
| Enforcement of judicial order                   | 100    | 86.66  | 92.89  | 76.92  | 4        |
| Average                                         | 100    | 87.85  | 93.48  | 88.38  | 4        |



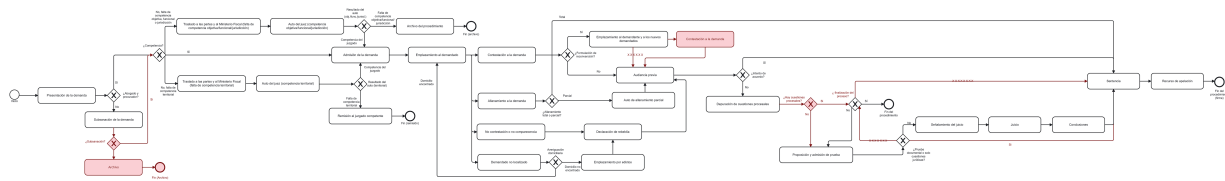
**Figure 9:** BPMN Model of the insolvency of a non-business person without assets procedure with the expert's corrections marked in red



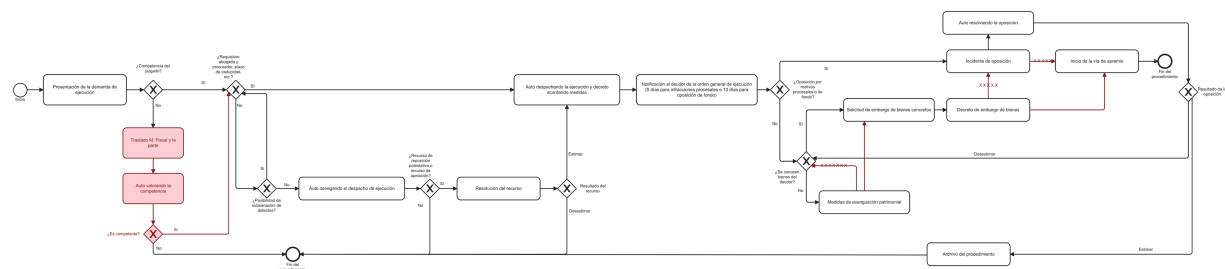
**Figure 10:** BPMN model of the payment order procedure with the expert's corrections marked in red



**Figure 11:** BPMN model of the foreclosure procedure with the expert's corrections marked in red



**Figure 12:** BPMN model of the ordinary trial procedure with the expert's corrections marked in red



**Figure 13:** BPMN model of the enforcement of judicial order procedure with the expert's corrections marked in red

To complement the descriptive metrics, we conducted a statistical analysis across the five judicial procedures. For each metric (precision, recall, F1-score, structural correctness), we computed the mean, standard deviation, and 95% confidence intervals using a t-distribution ( $df = 4$ ) and  $t(0.975, 4) \approx 2.776$ . The resulting intervals were narrow, indicating limited variability across procedures. Table 4 summarises the extended statistical analysis.

**Table 4:** Descriptive statistics and 95% CI ( $n = 5$ )

| Metric (%)                  | Mean   | SD    | 95% CI           |
|-----------------------------|--------|-------|------------------|
| Step Accuracy (SA)          | 100.00 | 0.00  | [100.00, 100.00] |
| Step Recall (SR)            | 87.85  | 3.25  | [83.82, 91.88]   |
| F1-score (F1)               | 93.48  | 1.81  | [91.24, 95.72]   |
| Structural correctness (SC) | 88.38  | 11.54 | [74.05, 102.72]  |

In addition, we applied a Wilcoxon signed-rank test to assess whether the deviations between the automatically generated models and the expert-validated baseline were statistically significant. The test

returned non-significant results ( $p > 0.05$ ) for all metrics, suggesting that errors are not systematic and supporting the robustness of the proposed methodology (SR:  $p = 0.0625$ , F1-score:  $p = 0.0625$ , SC:  $p = 0.0656$ ). This supports that the observed deviations are not systematic, the system performance is statistically stable, and the generated model does not show a consistent bias toward error in any metric.

Table 3 shows a step accuracy of 100% (SA = 100%) for the five analysed judicial procedures. This excellent result indicates that all the steps identified and included in the generated models are correct. This finding is crucial as it shows that the methodology does not introduce erroneous steps into the diagram.

Step recall (SR) varies between 84.21% and 92.85% across the different procedures, averaging at 87.89%. This means that, on average, the methodology was able to capture almost 88% of all relevant steps described verbally in the videos. While not reaching 100%, this level of completeness is high, especially considering the complexity of verbal descriptions and natural language.

The F1 score, which measures the balance between precision and recall, ranges from 91.38% to 96.25%, averaging at 93.48%. An F1 score above 90% indicates that the system performs well overall, confirming that the methodology effectively identifies and captures most of the steps.

Structural correctness (SC) shows the greatest variability, ranging from 75.00% to 100% with an average of 88.38%. The payment order procedure achieved 100% structural correctness. However, the ordinary trial and the enforcement of judicial order procedures showed the lowest values, at 75.00% and 76.92% respectively. This metric is crucial for the validity of BPMN diagrams as it measures whether the generated model adheres to the rules and standard notation of BPMN 2.0 (e.g., events, activities, gates and flows). The results suggest that, while most models are structurally correct, there is scope to improve the translation of complex verbal descriptions into BPMN structures in certain cases.

On average, generating each model takes approximately 4 min, with times varying between 3 and 5 min per procedure. This is a significant result, given that one of the main limitations of traditional approaches is their slowness and time-consuming nature. The proposed automated methodology considerably speeds up the creation of diagrams.

In summary, Table 3 shows that the proposed methodology for converting verbal descriptions into BPMN models with multimodal AI is very effective. These experimental results confirm that the proposed methodology can overcome the limitations of traditional approaches, simplifying the adoption of modelling tools and improving the speed, accuracy and accessibility of process modelling. The resulting models were evaluated and approved by a domain expert.

## 7 Discussion

The methodology proposed in this work represents a significant contribution to the field of business process modelling, addressing persistent limitations of traditional and contemporary approaches. In particular, it directly compares with prior studies in terms of automation capability, accuracy, and processing time.

### 7.1 Comparison with Rule-Based and NLP-Based Approaches

Early research based on predefined business rules [11–14] enabled deterministic mappings between textual statements and workflow constructs but required extensive manual rule engineering. The proposed multimodal-AI methodology eliminates this dependency, achieving full model generation from verbal input without prior rule definition. Similarly, grammar-based and NLP pipelines [15–19] demonstrated partial automation by extracting tasks and gateways from structured text, but they suffered from ambiguity and incomplete recall. In contrast, our results show an average F1 score of 93.48% across five real judicial

procedures, exceeding the typical precision–recall values ( $\approx 70\text{--}85\%$ ) reported for text-based extraction systems such as those by [15,19].

## 7.2 Comparison with Recent LLM-Based Research

While previous frameworks such as [10,21] demonstrated the feasibility of LLM-based process-model generation, their robustness remains constrained by the need for structured text input and frequent human correction of incomplete or inconsistent outputs. In contrast, our methodology incorporates multimodal data fusion and staged validation, which improves robustness against incomplete or ambiguous inputs.

Specifically, NotebookLM's audio–visual summarisation stage filters redundant or noisy verbal elements before textual conversion, and ChatGPT subsequently enforces BPMN syntax consistency through structured prompting and XML schema validation. As a result, the system successfully produced syntactically valid and semantically coherent BPMN diagrams with an F1 of 93.48%, with only minor post-editing required, while reducing modelling time to an average of 4 min per process.

This contrasts with earlier text-only pipelines, which typically reported partial or inconsistent model generation (precision and recall between 70%–85%). The multimodal integration and double-stage verification thus enhance robustness not only to linguistic variability but also to domain-specific terminology, making the approach more reliable for real-world judicial and administrative settings.

## 7.3 Comparison with Judicial and Public-Sector Modelling

Previous studies on legal or judicial workflows [30–33] successfully formalised court procedures using traditional BPMN or ontology-enhanced frameworks but required manual diagram construction. Our approach complements these efforts by introducing automation while maintaining semantic fidelity through expert validation. The obtained F1 scores above 90% confirm that multimodal LLMs can accurately capture procedural logic even in highly specialised legal domains. These findings suggest that LLM-based modelling could significantly reduce documentation time and enhance transparency in justice administration.

## 7.4 Synthesis and Implications

Overall, the comparison with previous research demonstrates that the proposed multimodal AI methodology achieves state-of-the-art performance in accuracy and efficiency while pioneering a new data modality (verbal video input). Unlike rule-based, NLP, or single-modal LLM systems, it integrates multimodal comprehension and structured-output generation in a unified workflow. In terms of methodological innovation, there are two key contributions:

- **Direct modelling from video:** one of the most innovative aspects is using verbal descriptions of the process directly from a video recording of an expert as input for the LLM. Unlike all previous work, which uses verbal descriptions in text format (such as transcripts or documents), our multimodal approach eliminates the need for information pre-processing (e.g., manual transcription), saving time and enabling experts to express their knowledge naturally, using their own terminology. Introducing video into a multimodal model allows the model to focus on interpreting the procedural structure (identifying activities, actors, decision points, and logical flows) rather than generating a literal textual transcription of each statement. This approach avoids the considerable time required for transcription and eliminates potential transcription errors, allowing the model to operate at a higher level of abstraction.
- **Combination of two LLMs:** this methodology stands out due to its strategic combination of two different LLMs, NotebookLM and ChatGPT. This combination enables the generation of process models with the fewest possible queries and the simplest prompts. This combination capitalises on the complementary

strengths of each tool: NotebookLM is used for multimodal contextual analysis and detailed schema generation, while ChatGPT is used for accurate BPMN XML code creation. This optimises the number of queries and the design of prompts to generate efficient BPMN models regardless of process complexity. To date, no literature has been found that uses this combination of LLMs or explores NotebookLM's capabilities for this purpose.

In this work, "optimisation" refers not to a mathematically proven optimum, but to a practical, task-focused optimisation based on experiments and real constraints: generating a complete BPMN diagram from a verbal process description in a video using only publicly available, non-fine-tuned LLMs. The proposed two-stage approach is the "optimised combination" because it required only two queries, kept prompts simple and reusable, and delivered the highest-quality outputs (SA = 100%, average F1 = 93%, and strong structural correctness).

The importance of using a minimal number of queries is both technical and practical. Technically, fewer queries help avoid exceeding an LLM's fixed context window, preventing loss or summarisation of essential information that can degrade output quality. Practically, limiting the interaction to one or very few prompts makes the process much simpler for non-technical users, who benefit from straightforward, easy-to-follow steps. Overall, reducing the number of queries leads to a more robust, efficient, and user-friendly workflow.

On the other hand, an important goal is to provide only the minimal information needed for the model to generate a correct answer. Keeping prompts simple serves two key purposes: it preserves a focused context, avoiding issues caused by limited context windows, and it reduces the risk of hallucinations by preventing the model from drifting into irrelevant or incorrect information. In short, lowering prompt complexity improves the model's focus, accuracy, and efficiency.

Our research is the first methodology to explore the applicability of LLMs and multimodal AI in automatically generating business procedure/operation models directly from verbal descriptions in video, filling an important gap in the literature. While previous work with LLMs, such as that by [23,37], is seminal, it is based on text inputs. Even related work by [21], which uses the multimodal GPT-4 model, focuses on interpreting existing BPMN diagrams rather than generating them. Furthermore, modelling judicial procedures using LLMs for automatic generation is considered a novel application, as previous work in this area does not explore this combination.

The proposed multimodal AI-based methodology for automatically generating BPMN models from verbal process descriptions can also be positioned in relation to emerging immersive and collaborative process-modelling environments. In industrial and organisational contexts, extended reality (XR) and metaverse-driven process design are gaining attention as means to visualise and manipulate business processes in 3D, multi-user spaces. The automatic creation of BPMN diagrams from natural verbal input could serve as a foundational step for such environments, enabling experts to describe processes verbally and immediately visualise them in immersive, interactive settings for collaborative refinement. Exploring this connection between multimodal AI modelling and XR-based or industrial metaverse technologies represents a promising avenue for future research.

Regarding the applicability and generalisation of the proposed technique, it is important to note that, although the methodology has been validated in the specific context of Spain's Commercial Courts, it is presented as a generalisable computational methodology. This means it can be used to model operations and processes in other organisations and disciplines, such as public administration, industry, health and education. Its ability to capture expert knowledge without manual transcription or modification of natural language expands its potential use in environments where process modelling is crucial.

Despite the promising results, it is important to recognise the limitations of the methodology. The methodology's effectiveness continues to depend on the quality of the verbal description provided by the expert. While multimodal LLMs are robust, the clarity and comprehensiveness of the video explanation are essential for the accurate extraction of information. LLMs such as ChatGPT can be sensitive to differences in prompt wording and may occasionally provide incorrect answers or fail to clarify ambiguous questions. While NotebookLM reduces the likelihood of inventions by sticking to sources, experimenting with different prompts (prompt engineering) remains necessary for each domain or tool to achieve optimal results.

On the other hand, expanding the evaluation to multiple judicial experts poses practical and institutional challenges. Expert validation in the legal domain requires magistrates with formal jurisdiction, procedural authority, and direct experience with the specific case types under study. Access to such experts is inherently limited due to confidentiality regulations and workload constraints. For these reasons, the present study relied on a single magistrate who was both available and formally qualified to assess procedural correctness. Nevertheless, we recognise that multi-expert validation would reduce subjectivity and increase robustness, and we are currently coordinating with additional courts and judicial offices to enable broader expert participation in future studies.

An additional direction for future work involves exploring open-source multimodal LLMs such as LLaVA, Mistral-MM, and other emerging models. Although these systems are rapidly evolving and increasingly capable, their adoption in the present study was not feasible due to two constraints: (i) their multimodal capacities (particularly regarding audio-video understanding) remain heterogeneous and often require customised fine-tuning pipelines, and (ii) their deployment could demand high-end GPU hardware that is not accessible in non-technical or resource-constrained environments. Since one of the aims of this work is to provide a low-barrier, widely accessible methodology that can be used without programming skills or specialised infrastructure, we focused on readily available cloud-based tools. Nonetheless, as open-source multimodal models become more stable and lightweight, evaluating their suitability for BPMN generation will constitute a valuable extension of this research.

Finally, expert validation is an essential step in the methodology to ensure the correctness and applicability of the generated models. This highlights that full automation still requires human oversight.

## 8 Conclusions

Business process modelling is essential for improving, standardising, and automating organisational workflows, and BPMN remains the most widely recognised notation for this purpose. However, creating BPMN diagrams manually is time-consuming, prone to inconsistencies, and requires specialised expertise.

This study introduces an innovative methodology for automatically generating BPMN models from verbal expert descriptions using multimodal artificial intelligence. By combining NotebookLM (Gemini) and ChatGPT (GPT-4), the proposed pipeline produces complete and accurate BPMN diagrams directly from video recordings without the need for manual transcription or pre-processing. Validation conducted in Spain's Commercial Courts demonstrates high performance, with an average F1-score of 93.5% and perfect step accuracy (SA = 100%). Moreover, the modelling time was reduced to approximately 4 min per procedure (ranging from 3 to 5 min), illustrating the substantial efficiency gains achieved by the approach.

These results show that the methodology enhances the speed, accuracy, and accessibility of process modelling. Allowing experts to describe procedures naturally, while delegating the extraction and structuring tasks to multimodal LLMs, significantly reduces modelling effort and lowers the technical barrier for non-specialist users. The combined use of NotebookLM and ChatGPT proved particularly effective for resolving

semantic ambiguities, capturing contextual nuances, and maintaining coherence in the generated diagrams with a minimal number of prompts.

The methodology was designed to provide a simple, fast, and accessible solution for any user, rather than to conduct a comprehensive comparative study of all available Large Language Models (LLMs). Through an iterative process, ChatGPT was initially tested alone, but it struggled with long, unstructured video transcriptions, resulting in unreliable BPMN outputs due to context overload. This highlighted the need for an intermediate step. NotebookLM was chosen because it can directly process video and is well suited for synthesizing clear outlines from complex material. ChatGPT, in turn, performs effectively when provided with such a structured schema. The proposed two-stage setup is considered the most effective and user-friendly approach for achieving robust process modeling with minimal instructions, minimal steps, and high accuracy for non-technical users.

In conclusion, this work highlights the transformative potential of multimodal LLMs for process representation and knowledge capture. The strong performance and substantial reduction in modelling time indicate that the methodology can be extended to other application domains and may constitute a significant step forward in the continuous improvement and automation of organisational activities.

Despite these promising results, several limitations must be acknowledged:

1. Dependence on input quality: the accuracy of the generated BPMN models depends heavily on the clarity and completeness of the expert's verbal explanation. Ambiguous or fragmented descriptions may lead to missing or misplaced steps.
2. Limited dataset size: the evaluation was conducted on a small set of judicial procedures from a single institutional context. Wider validation across sectors, languages, and organisational environments is required to assess generalisability.
3. Closed-source models: both NotebookLM and ChatGPT are proprietary systems. Although they provide strong performance, their closed nature limits transparency, interpretability, and reproducibility across environments. Future research should evaluate open-source multimodal LLMs to improve explainability and adaptability.
4. Human-in-the-loop dependency: expert validation remains essential to ensure semantic correctness and legal compliance, particularly in sensitive judicial domains.
5. Single-expert validation: access to multiple judicial experts is restricted by confidentiality obligations, institutional regulations, and high workloads. For this reason, the validation relied on a single magistrate. Nonetheless, expanding evaluation to additional experts from different courts is planned to reduce subjectivity and strengthen generalisability.

Future research will focus on expanding the dataset, experimenting with open-source multimodal models, and incorporating automatic semantic validation through ontologies or simulation-based feedback. Additional efforts will explore the identification of inefficiencies and dispensable activities, the integration of timing and Key Performance Indicators (KPI)-related information, and the creation of guidelines to help experts describe processes in ways that maximise AI performance. Another promising direction is the integration of this methodology into XR and industrial-metaverse platforms to support immersive, collaborative process modelling.

Although the expert in our case study relied only on spoken explanations, using video and a multimodal LLM was essential to demonstrate that our two-stage framework can handle more complex scenarios. This design choice prepares the methodology for cases where expert explanations include visual cues, gestures, or graphical elements. It also opens the door for future research on how multimodal LLMs integrate verbal

and visual information to enhance the generated process schema, capturing non-verbal nuances conveyed through video.

In summary, although the present study demonstrates a substantial advance in multimodal LLM-based BPMN generation, these limitations underline that full automation and domain-robust reliability will require further investigation. Recognising these constraints strengthens the transparency of the methodology and provides a clear roadmap for subsequent research.

**Acknowledgement:** Not applicable.

**Funding Statement:** This research was funded by Fundación CajaCanarias and Fundación Bancaria “la Caixa”, grant number 2023DIG11.

**Author Contributions:** The authors confirm contribution to the paper as follows: Conceptualization: Rosa Aguilar; methodology: Silvia Alayón and Rosa Aguilar; software: Carlos Martín, Jesús Torres and Manuel Bacallado; validation: Guzmán Savirón; formal analysis: Silvia Alayón, Carlos Martín and Rosa Aguilar; writing—original draft preparation: Silvia Alayón, Jesús Torres and Manuel Bacallado; supervision: Silvia Alayón. All authors reviewed the results and approved the final version of the manuscript.

**Availability of Data and Materials:** The multimodal video recordings used in this study cannot be publicly released due to judicial confidentiality regulations and the explicit request of the domain expert. However, all BPMN models generated during the experiments in XML format can be provided upon reasonable request to the authors.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest to report regarding the present study.

## References

1. Object Management Group. Documents associated with business process model and notation™ (BPMN™) Version 2.0 [Internet]. 2011 [cited 2025 Sep 5]. Available from: <http://www.omg.org/spec/BPMN/2.0>.
2. Chinosi M, Trombetta A. BPMN: an introduction to the standard. *Comput Stand Interfaces*. 2012;34(1):124–34. doi:10.1016/j.csi.2011.06.002.
3. Gažová A, Papulová Z, Smolka D. Effect of business process management on level of automation and technologies connected to industry 4.0. *Procedia Comput Sci*. 2022;200:1498–507. doi:10.1016/j.procs.2022.01.351.
4. Eder J, Franceschetti M. Time and business process management: problems, achievements, challenges. In: *Proceedings of the 27th International Symposium on Temporal Representation and Reasoning (TIME)*; 2020 Sep 23–25; Bozen-Bolzano, Italy. p. 3:1–3:8.
5. Schummer J. Science communication across disciplines. In: Holliman R, Thomas J, Smidt S, Scanlon E, Whitelegg E, editors. *Practising science communication in the information age: theorising professional practices*. Oxford, UK: Oxford University Press; 2008. p. 53–66.
6. Almarie B, Teixeira PEP, Pacheco-Barrios K, Rossetti CA, Fregni F. Editorial—the use of large language models in science: opportunities and challenges. *Princ Pract Clin Res*. 2023;9(1):1–4. doi:10.21801/ppcrj.2023.91.1.
7. OpenAI. ChatGPT Website [Internet]. 2022 [cited 2025 Sep 5]. Available from: <https://chatgpt.com/es-ES/overview/>.
8. Google. Gemini Website [Internet]. 2023 [cited 2025 Sep 5]. Available from: <https://gemini.google.com>.
9. Anthropic. Claude Website [Internet]. 2025 [cited 2025 Sep 5]. Available from: <https://www.anthropic.com/claude>.
10. Kourani H, Berti A, Schuster D, van der Aalst WMP. Process modeling with large language models. *arXiv*: 2403.07541. 2024.
11. Knolmayer G, Endl R, Pfahrer M. Modeling processes and workflows by business rules. In: Van der Aalst W, Desel J, Oberweis A, editors. *Business process management*. Berlin/Heidelberg, Germany: Springer; 2000. p. 16–29. doi:10.1007/3-540-45594-9\_2.

12. zur Muehlen M, Indulska M. Modeling languages for business processes and business rules: a representational analysis. *Inf Syst.* 2010;35(4):379–90. doi:10.1016/j.is.2009.02.006.
13. Kluza K, Nalepa GJ. Automatic generation of business process models based on attribute relationship diagrams. In: Lohmann N, Song M, Wohed P, editors. *Business process management workshops*. Cham, Switzerland: Springer International Publishing; 2014. p. 185–97. doi:10.1007/978-3-319-06257-0\_15.
14. Kluza K, Nalepa GJ. A method for generation and design of business processes with business rules. *Inf Softw Technol.* 2017;91:123–41. doi:10.1016/j.infsof.2017.07.001.
15. Sintoris K, Vergidis K. Extracting business process models using natural language processing (NLP) techniques. In: *Proceedings of the 2017 IEEE 19th Conference on Business Informatics (CBI)*; 2017 Jul 24–27; Thessaloniki, Greece. p. 135–9. doi:10.1109/cbi.2017.41.
16. Leopold H, Mendling J, Polyvyanyy A. Supporting process model validation through natural language generation. *IEEE Trans Software Eng.* 2014;40(8):818–40. doi:10.1109/tse.2014.2327044.
17. Honkisz K, Kluza K, Wiśniewski P. A concept for generating business process models from natural language description. In: Liu W, Giunchiglia F, Yang B, editors. *Knowledge science, engineering and management*. Cham, Switzerland: Springer International Publishing; 2018. p. 91–103. doi:10.1007/978-3-319-99365-2\_8.
18. van der Aa H, Carmona J, Leopold H, Mendling J, Padró L. Challenges and opportunities of applying natural language processing in business process management. In: *Proceedings of the 27th International Conference on Computational Linguistics*; 2018 Aug 20–26; Santa Fe, NM, USA. p. 2791–801.
19. Nasiri S, Adadi A, Lahmer M. Automatic generation of business process models from user stories. *Int J Electr Comput Eng IJECE.* 2023;13(1):809. doi:10.11591/ijece.v13i1.pp809-822.
20. Hafner A, Wittges H, Rinderle-Ma S. GenAI in business process management: a systematic review of the current state. In: *Proceedings of the Americas Conference on Information Systems 2025*; 2025 Aug 14–16; Montréal, QC, Canada.
21. Dolha DN, Buchmann RA. Generative AI for BPMN process analysis: experiments with multi-modal process representations. In: Řepa V, Matulevičius R, Laurenzi E, editors. *Perspectives in business informatics research*. Cham, Switzerland: Springer Nature; 2024. p. 19–35. doi:10.1007/978-3-031-71333-0\_2.
22. Ziche C, Apruzzese L. LLM4PM: a case study on using large language models for process modeling in enterprise organizations. In: *Proceedings of the International Conference on Business Process Management*; 2024 Sep 1–6; Krakow, Poland. p. 472–83. doi:10.1007/978-3-031-70445-1\_35.
23. Daclin N, Mallek-Daclin S, Zacharewicz G. Generative AI for business model generation (GAI4BM): from textual description to business process model. In: *Proceedings of the 10th International Food Operations and Processing Simulation Workshop (FoodOPS 2024)*; 2024 Sep 18–20; Tenerife, Spain. Bari, Italy: CAL-TEK srl; 2024. p. 1–8. doi:10.46354/i3m.2024.foodops.013.
24. Apaydin K, Zisgen Y. Local large language models for business process modeling. In: Delgado A, Slaats T, editors. *Process mining workshops*. Cham, Switzerland: Springer Nature; 2025. p. 605–9. doi:10.1007/978-3-031-82225-4\_44.
25. Licardo JT, Tankovic N, Etinger D. BPMN assistant: an LLM-based approach to business process modeling. *arXiv: 2509.24592*. 2025.
26. Safan A, Köpke J. BPMN-chatbot++: LLM-based modeling of collaboration diagrams with data objects and message flows. In: *Proceedings of the 23rd International Conference on Business Process Management (BPM 2025)*; 2025 Aug 31–Sep 5; Seville, Spain.
27. Bonitasoft. Transform your business process directly into a BPMN Model with AI [Internet]. 2024 [cited 2025 Sep 5]. Available from: <https://www.bonitasoft.com/ai-bpmn-generator>.
28. Camunda. AI-Enabled Process Orchestration [Internet]. 2023 [cited 2025 Sep 5]. Available from: <https://camunda.com/platform/ai-enabled-process-orchestration/>.
29. Bizagi. Artificial Intelligence with Bizagi. 2024 [cited 2025 Sep 5]. Available from: <https://www.bizagi.com/en/solutions/artificial-intelligence>.

30. Zhang T, Zeng X, Liu Z. Modeling workflow for judicial business processes: a use case driven method. In: Proceedings of the 2021 7th International Conference on Information Management (ICIM); 2021 Mar 27–29; London, UK. p. 45–56. doi:10.1109/icim52229.2021.9417044.
31. Ciaghi A, Weldemariam K, Villafiorita A. Law modeling with ontological support and BPMN: a case study. In: Proceedings of the Second International Conference on Technical and Legal Aspects of the E-Society; 2011 Feb 23–28; Gosier, France. p. 29–34.
32. Audrito D, Ferraris AF. Legal design through business process model and notation (BPMN): a digital services act case-study. In: Proceedings of the 9th Joint Ontology Workshops (JOWO 2023); 2023 Jul 19–20; Sherbrooke, QC, Canada.
33. Mamrot S. Application of business process modelling and reengineering to law making process in Poland. *JeDEM*. 2023;15(1):144–68. doi:10.29379/jedem.v15i1.736.
34. Camunda. What is BPMN? Business Process Model and Notation [Internet]. 2013 [cited 2025 Sep 5]. Available from: <https://camunda.com/bpmn/>.
35. OpenAI. GPT-4 Website [Internet]. 2023 [cited 2025 Sep 5]. Available from: <https://openai.com/index/gpt-4/>.
36. Google. NotebookLM Website [Internet]. 2024 [cited 2025 Sep 5]. Available from: <https://notebooklm.google/>.
37. Kampik T, Warmuth C, Rebmann A, Agam R, Egger LNP, Gerber A, et al. Large process models: a vision for business process management in the age of generative AI. *Künstl Intell*. 2025;39(2):81–95. doi:10.1007/s13218-024-00863-8.