

---

# IA para la **PARTICIPACIÓN**

---

El uso de técnicas de IA para resolver problemáticas comunes en procesos participativos



**POLITICAL  
WATCH**



red.es



Mayo 2026. Informe emitido en el marco del Convenio que la entidad pública empresarial [Red.es](#) ha suscrito con la agrupación de entidades formada por [Political Watch](#) (Fundación Salvador Soler), la Fundación Hay Derecho y la Fundación Haz, para impulsar la implementación de la Carta de Derechos Digitales en el ámbito de los derechos de participación [c-038/23-ot].

La información y las opiniones expresadas en este estudio son de los autores/as y no reflejan necesariamente la opinión oficial de las instituciones firmantes del convenio de colaboración en cuyo marco se ha realizado este documento. Las instituciones firmantes del convenio no garantizan la exactitud de los datos incluidos en este documento. Ni estas instituciones ni ninguna persona que actúe en su nombre pueden ser considerados responsables del uso que pueda hacerse de la información contenida en el mismo.

Se autoriza la reproducción para fines docentes o sin ánimo de lucro, siempre que se cite la fuente.

Copyright: [CC BY-SA 4.0](#) 



DERECHOS  
DIGITALES

POLITICAL  
WATCH

## CASOS DE USO EN LOS PROCESOS PARTICIPATIVOS / DELIBERATIVOS

PREVIO a la participación			DURANTE	POSTERIOR a la participación		
CASO 1 Convocatoria	CASO 2 Dilemas	CASO 3 Dinamización	CASO 4 Moderación	CASO 5 Rendición de cuentas	CASO 6 Socialización	CASO 7 Vinculación

<b>CASO 1</b>	Convocatoria automatizada a procesos participativos
<b>CASO 2</b>	Selección de dilemas deliberativos a partir del análisis de insumos previos
<b>CASO 3</b>	Generación de materiales dinamizadores a partir de insumos previos
<b>CASO 4</b>	Análisis automatizado para la moderación y la detección de consensos
<b>CASO 5</b>	Elaboración automatizada de materiales de rendición de cuentas
<b>CASO 6</b>	Producción de materiales para la socialización de resultados
<b>CASO 7</b>	Vinculación entre insumos participativos y producción normativa

# Índice

El uso de técnicas de IA para resolver problemáticas comunes en procesos participativos // 5

Glosario de términos // 7

Los 10 criterios para seleccionar un modelo del lenguaje // 15

Cómo tener el control de lo que está pasando // 22

Cómo calcular los costes de un proyecto // 26

Soporte para lenguas co-oficiales // 30

Consideraciones éticas y de transparencia (código abierto, política de tratamiento de datos, transparencia algorítmica,...) // 32

Recomendaciones para su uso ético y responsable en entornos seguros de participación juvenil // 34

Introducción a los estudios de caso // 36

Convocatoria automatizada a procesos participativos // 41

Selección de dilemas deliberativos a partir del análisis de insumos previos // 54

Generación de materiales dinamizadores a partir de insumos previos // 68

Análisis automatizado para la moderación y la detección de consensos // 82

Elaboración automatizada de materiales de rendición de cuentas // 98

Producción de materiales para la socialización de resultados // 113

Vinculación entre insumos participativos y producción normativa // 126

# El uso de técnicas de IA para resolver problemáticas comunes en procesos participativos

La inteligencia colectiva puede definirse como la capacidad de un grupo de personas de tomar buenas decisiones, elegir qué hay que hacer, el mejor camino a recorrer para hacerlo, y de quién deberíamos rodearnos para hacerlo, todo esto a través de una comunicación humana donde también pueden participar herramientas tecnológicas (Shin et al., 2024)

Promover entornos que faciliten estas comunicaciones y habiliten el camino para una mejor inteligencia colectiva es el espíritu de este documento, en el cual, a través de una mezcla de metodologías de investigación y conocimiento técnico, se ofrecen **soluciones basadas en IA para 7 grandes retos vinculados con las diversas etapas del proceso participativo y deliberativo.**

Las soluciones que se plantean tienen en cuenta una serie de consideraciones previas, que se orientan a facilitar la labor de la Administración pública en el difícil reto de aterrizar una tecnología en constante expansión de manera responsable y asegurando que se protege a la ciudadanía en el proceso. Para ello se establecen una serie de criterios basados en la ética, la responsabilidad y la promoción del bien común a la hora de definir las orientaciones para cada uno de los retos definidos. Estos criterios son amplios, y van desde las capacidades actuales que ofrece la tecnología, los niveles de propiedad que se puede tener sobre el código, la transparencia y la rendición de cuentas, la protección de datos, el respeto por la multiculturalidad lingüística en un país como España, entre otras.

Por otro lado, al momento de analizar en detalle cada uno de los retos planteados y establecer criterios para encontrar posibles soluciones basadas en técnicas de IA, se toma como base la propuesta teórica de Fung (2006), que define tres elementos imprescindibles a fomentar en cualquier parte del proceso participativo, que son la inclusión, la deliberación de calidad y el empoderamiento de la ciudadanía en la toma de decisiones. De esta manera, las soluciones en ningún caso son completamente autónomas, sino que cuentan con elementos de revisión humana constantes a la vez que han sido diseñadas para impulsar y mejorar la calidad de mecanismos ya existentes, no para sustituirlos.

El documento tiene una estructura basada en módulos, en primer lugar se definen una serie de consideraciones generales, comunes a los 7 casos de análisis. En dichas consideraciones el lector encontrará desde definiciones clave hasta potenciales riesgos y salvaguardas a tener en cuenta al momento de aplicar las diferentes soluciones. Su objetivo es guiar los procesos de implementación que puedan surgir al momento de resolver cada uno de los presentes retos. A continuación se incluyen 7 módulos, uno por cada caso de uso, que se encuentran interconectados entre sí porque se ubican cada uno en un momento diferente del proceso participativo, antes, durante y después. Pero que al mismo tiempo son independientes entre sí, permitiendo al lector y potencial usuario de estas soluciones trabajar en cada reto de manera individual aprovechando el conocimiento técnico que aporta el documento.

Por último, cada caso de uso cuenta con un apartado común de consideraciones éticas, donde se intenta dar claves a tener en cuenta para asegurar que los procesos de implementación sean seguros y respetuosos de la diversidad humana y donde la protección de sus derechos fundamentales esté siempre en el centro del debate.

## Glosario de términos

### ■ Determinismo:

Propiedad de un sistema cuyo comportamiento es completamente predecible: dada la misma entrada, produce siempre la misma salida. Los LLMs no son deterministas por naturaleza (su salida puede variar entre ejecuciones incluso con el mismo prompt), lo que tiene implicaciones directas en la reproducibilidad, la evaluación y la auditabilidad de los sistemas. Las técnicas clásicas de aprendizaje automático y los sistemas basados en reglas sí suelen ser deterministas, y esa propiedad es una de sus principales ventajas en contextos institucionales.

### ■ Dataset:

Conjunto estructurado de datos utilizado para entrenar o evaluar un modelo de IA. En el contexto de estos casos de uso, el término aparece sobre todo asociado al dataset de evaluación: un conjunto de entradas representativas con criterios o respuestas de referencia que permite medir de forma sistemática la calidad del sistema y detectar regresiones cuando algún componente cambia.

### ■ Regresión:

Deterioro no intencionado del comportamiento de un sistema que antes funcionaba correctamente, provocado por un cambio en alguno de sus componentes. En los sistemas basados en LLMs, una regresión puede aparecer tras una actualización del modelo por parte del proveedor, una modificación del prompt, un cambio en las herramientas o en los datos de entrada. A diferencia de un fallo nuevo, la regresión afecta a algo que ya se había validado como correcto, lo que la hace especialmente difícil de detectar sin un dataset de evaluación que se revise periódicamente.

■ Traza:

Registro completo de una ejecución del sistema: la entrada recibida, los prompts enviados al modelo, las llamadas a herramientas, los resultados intermedios, la salida final, los tokens consumidos, la latencia y los errores encontrados. Las trazas son el elemento fundamental de la observabilidad y permiten diagnosticar problemas, auditar el comportamiento del sistema y reconstruir el proceso que llevó a una salida concreta.

■ Embedding:

Representación numérica de un fragmento de texto (palabra, frase o documento) en forma de vector de alta dimensionalidad, generada por un modelo especializado. Los embeddings capturan el significado semántico del texto de manera que fragmentos con significado similar tienen vectores cercanos en el espacio. Son la base de los sistemas de búsqueda semántica y de los componentes de recuperación (RAG): permiten encontrar documentos relevantes no por coincidencia de palabras clave sino por proximidad de significado.

■ Latencia:

La latencia es el tiempo que tarda el modelo en devolver una respuesta desde que se le envía la petición.

■ LLM (Large Language Model):

Modelo de IA entrenado con grandes cantidades de texto para comprender y generar lenguaje natural. Es el componente fundamental que utilizan los sistemas propuestos en estos casos de uso para analizar aportaciones, generar materiales o sintetizar información.

■ Reasoning LLM:

Existen variantes especializadas como los llamados reasoning LLMs (o modelos de razonamiento), que dedican tokens adicionales a procesos

internos de "pensamiento" antes de emitir una respuesta, lo que mejora su desempeño en tareas complejas de análisis y planificación a cambio de mayor latencia y coste.

- IA multimodal:

Categoría de modelos capaces de procesar y/o generar información en varios formatos simultáneamente (texto, imagen, audio, vídeo). En estos casos de uso es especialmente relevante para la generación de materiales gráficos a partir de insumos textuales y para la comprensión de documentos con contenido mixto.

- IA generativa:

Conjunto de técnicas de IA cuya finalidad principal es producir contenido nuevo (texto, imagen, audio, código) a partir de instrucciones del usuario y su entrenamiento previo, en lugar de limitarse a clasificar o predecir.

- Token:

Unidad mínima de texto que procesa un LLM. Un token suele equivaler aproximadamente a 3-4 caracteres o a un fragmento de palabra. Los modelos facturan habitualmente por número de tokens procesados (entrada) y generados (salida), por lo que el conteo de tokens determina directamente el coste variable de operación del sistema.

- Ventana de contexto:

Cantidad máxima de tokens que un LLM puede procesar de una sola vez (entrada más salida). Es una limitación técnica clave: determina cuánta información puede "ver" el modelo simultáneamente y obliga a diseñar estrategias como la compresión progresiva de contexto o la segmentación en fragmentos cuando el volumen de información supera ese límite.

#### ■ Entrenamiento:

El entrenamiento es el proceso por el que un LLM aprende patrones del lenguaje a partir de grandes volúmenes de datos.

#### ■ Fine-tuning:

El fine-tuning (o ajuste fino) es un entrenamiento adicional sobre un modelo base ya entrenado, con datos específicos de un dominio o tarea, para especializarlo sin necesidad de entrenar un modelo desde cero.

#### ■ Inferencia:

Proceso de ejecución de un modelo ya entrenado para obtener una respuesta a partir de una entrada concreta. En la práctica, es el "uso en producción" del modelo, y es donde se concentra el coste variable de los sistemas basados en IA.

#### ■ Memoria:

Capacidad del sistema para retener información entre interacciones o iteraciones. Los LLMs no tienen memoria propia (cada inferencia es independiente), por lo que la memoria se implementa a través de mecanismos externos: incluir el historial previo en la entrada, almacenar resúmenes persistentes, o recuperar información relevante desde bases de datos auxiliares.

#### ■ Prompt (y prompting):

Instrucción o conjunto de instrucciones que se proporciona al LLM para obtener una respuesta. El diseño del prompt (prompting) incluye el rol asignado al modelo, el contexto, los ejemplos y el formato esperado de la salida. En los sistemas propuestos, los prompts del agente incorporan criterios explícitos como las directrices del proceso, los objetivos de la fase o los elementos que deben preservarse al sintetizar información.

#### ■ Agente (y arquitectura agéntica):

Sistema basado en uno o varios LLMs que opera de forma autónoma para alcanzar un objetivo, decidiendo qué pasos dar a partir del contexto disponible. A diferencia de un pipeline determinista, un agente puede planificar, razonar sobre resultados intermedios y reaccionar adaptativamente. La arquitectura agéntica describe cómo se organizan estos componentes: desde agentes simples con una única función, hasta sistemas multi-agente jerárquicos con un orquestador que coordina agentes especializados, cada uno con acceso a sus propias herramientas.

#### ■ RAG (Retrieval-Augmented Generation):

Técnica que combina un LLM con un sistema de recuperación de información, de modo que antes de generar una respuesta el sistema busca y añade al contexto fragmentos relevantes de una base documental. Permite que el modelo trabaje con información actualizada o específica del dominio sin necesidad de reentrenarlo, y mejora la trazabilidad al poder citar las fuentes recuperadas.

#### ■ Tool calling:

Capacidad de un LLM para invocar funciones o herramientas externas (búsquedas, cálculos, llamadas a APIs, consultas a bases de datos) durante su razonamiento. Es el mecanismo que permite a los agentes interactuar con el entorno en lugar de limitarse a generar texto, y es clave en las soluciones propuestas para dotar a los agentes de capacidades específicas sin complicar la arquitectura.

#### ■ Salida estructurada:

Patrón por el que se obliga al LLM a producir su salida siguiendo un esquema predefinido (tipos de datos, campos obligatorios, valores permitidos) en lugar de texto libre. Permite que la salida sea directamente procesable por el resto del sistema sin necesidad de parsearla.

Se implementa habitualmente con bibliotecas como Pydantic, que definen el esquema mediante clases de Python y validan automáticamente que la salida del modelo se ajuste a él.

- Framework de orquestación:

Biblioteca o conjunto de herramientas que facilita la construcción de sistemas basados en agentes, gestionando el flujo de control, la invocación de herramientas, el manejo del estado y la comunicación entre componentes. En estos casos de uso se mencionan LangChain y su evolución LangGraph (que permite modelar el flujo del agente como un grafo dirigido con bucles, condicionales y estados persistentes), así como alternativas de bajo código como n8n para flujos más sencillos.

- Pipeline:

Secuencia determinista de pasos por los que pasa un dato hasta obtener un resultado. A diferencia de un agente, un pipeline tiene un flujo fijo y predecible. En las propuestas se utilizan tanto pipelines deterministas (para fases de ingesta, preparación o postprocesado) como agentes dinámicos (para fases que requieren adaptabilidad, como el análisis o la personalización), combinando ambos enfoques según la naturaleza de cada tarea.

- Guardrails (salvaguardas):

Mecanismos que restringen, orientan o filtran el comportamiento de un sistema basado en LLM para garantizar que sus salidas cumplan determinadas condiciones (evitar contenido inadecuado, ceñirse a un dominio, respetar formatos o políticas). Pueden aplicarse antes de la inferencia (en el prompt), durante (con comprobaciones en tiempo real) o después (validando la salida antes de emitirla).

- Human-in-the-loop:

Principio de diseño por el que las decisiones relevantes del sistema automatizado requieren validación, revisión o confirmación humana antes de

ser efectivas. En todos los casos de uso planteados, el papel de los LLMs es asistir al personal técnico, no sustituirlo: las salidas de los agentes son siempre insumos que una persona profesional revisa, contrasta con su propio criterio y decide cómo aplicar.

- **Speech-to-text (STT):**

Tecnología que transforma audio hablado en texto escrito. Es la entrada habitual para sistemas que procesan grabaciones de sesiones deliberativas, entrevistas o eventos participativos. Su calidad depende del modelo utilizado, del idioma, del ruido ambiente y del número de hablantes simultáneos.

- **Diarización:**

Técnica complementaria al speech-to-text que identifica y separa automáticamente a los distintos hablantes en una grabación. Permite asignar cada intervención a un hablante distinto sin necesidad de conocer su identidad real, lo que facilita análisis más finos como el seguimiento de posiciones individuales a lo largo de una sesión o el estudio del equilibrio de participación entre grupos.

- **Text-to-speech (TTS):**

Tecnología inversa al STT, que convierte texto escrito en audio hablado con voz sintetizada. Es útil para producir versiones audibles de materiales (resúmenes, relatorías, contenidos accesibles) o para sistemas de interacción oral.

- **Evaluación:**

Proceso sistemático de medir la calidad de las salidas de un sistema basado en LLM frente a criterios predefinidos (corrección, fidelidad a las fuentes, consistencia, ausencia de alucinaciones). En las propuestas, la evaluación es clave para iterar sobre los prompts y las arquitecturas antes de desplegar el sistema en producción. Existen bibliotecas especializadas como Ragas (enfocada en sistemas de recuperación) o Deepeval (para evaluación general de LLMs) que facilitan este proceso.

■ Observabilidad:

Capacidad de inspeccionar, registrar y auditar el comportamiento de un sistema basado en LLMs en producción: qué prompts se enviaron, qué respondió el modelo, cuánto tiempo tardó, qué coste tuvo y qué errores ocurrieron. Es imprescindible para diagnosticar problemas, controlar costes y mantener la trazabilidad necesaria en contextos institucionales. Herramientas como Langfuse (open source, auto-hospedable) o LangSmith (propietaria) proporcionan esta capacidad.

## Los 10 criterios para seleccionar un modelo del lenguaje

### 1. Evaluación de la calidad del modelo

La primera pregunta es siempre la más básica: ¿es el modelo lo bastante bueno haciendo lo que le vamos a pedir? Como primer nivel de orientación cabe apoyarse en las tarjetas de modelo (que detallan las especificaciones técnicas), en baterías de benchmarks públicos (MMLU y GPQA para razonamiento general, HumanEval o SWE-Bench para código, MT-Bench para seguimiento de instrucciones, entre otras) y en las leaderboards independientes que ordenan los modelos por capacidades concretas, teniendo presente que son una señal orientativa y no un sustituto de la prueba real. El mejor criterio de calidad siempre es el dataset de evaluación específico del caso de uso que se construye en la fase de preparación, porque captura exactamente lo que importa para el proyecto en lugar de una media sobre tareas genéricas. En los casos de uso que impliquen análisis, planificación o síntesis compleja conviene además considerar los reasoning LLMs como categoría aparte: suelen ofrecer mejor desempeño en tareas de razonamiento a costa de mayor latencia y mayor consumo de tokens.

### 2. Las tres capacidades que pueden descartar un modelo antes de empezar

Más allá de la calidad general, hay tres capacidades técnicas que condicionan de forma directa lo que se puede construir con un modelo. La multimodalidad (capacidad de procesar o generar información en formatos distintos del texto, como imagen, audio o vídeo) es especialmente relevante en los casos de uso que generan materiales gráficos a partir de insumos textuales. El tool calling es la base de cualquier arquitectura agéntica: si el modelo no invoca herramientas con precisión, el agente no puede interactuar con su entorno. Y la salida estructurada determina si lo que el modelo nos devuelve se puede integrar directamente con el resto del sistema o exige un procesamiento posterior. No todos los modelos soportan estas capacidades con la misma madurez y una limitación clara en cualquiera de ellas puede descartar un modelo antes incluso de entrar a comparar otros aspectos.

### 3. Ventana de contexto: alcance y limitaciones

La ventana de contexto determina cuánta información puede procesar el modelo en una sola interacción. Para los casos de uso planteados, con sesiones deliberativas largas, actas extensas o numerosos documentos de referencia, es una restricción práctica importante. Conviene, eso sí, compartir una advertencia: una ventana de contexto muy amplia no garantiza que el modelo la use bien. Existe una degradación de atención conocida cuando el contexto está saturado, ya que los modelos tienden a prestar más atención al principio y al final que al centro del material proporcionado, por lo que disponer de contextos grandes no sustituye a estrategias como la comprensión progresiva de contexto, el RAG o la segmentación en fragmentos más manejables.

### 4. Modelos propietarios o abiertos: implicaciones estratégicas

La primera decisión es si se trabaja con un modelo propietario (de pesos cerrados, accesible únicamente a través de la API del proveedor) o con un modelo open source (de pesos abiertos, que puede inspeccionarse, modificarse y desplegarse en cualquier infraestructura). Los modelos propietarios de última generación (GPT-5 de OpenAI, Claude de Anthropic, Gemini de Google) suelen marcar la frontera de calidad, pero generan una dependencia directa del proveedor en cuanto a disponibilidad, precios, condiciones de uso y tratamiento de los datos. Los modelos *open source* (Mistral, Gemma, Qwen, entre otros) han cerrado buena parte de la brecha de calidad, ofrecen transparencia total sobre los pesos y la arquitectura, y pueden consumirse de múltiples formas: desplegados en infraestructura propia, a través de servicios gestionados o mediante APIs de terceros que los sirven como servicio.

En los últimos años ha ganado tracción la arquitectura MoE (Mixture of Experts, o “mezcla de expertos”), que permite construir modelos con un número total de parámetros muy elevado pero activando solo una fracción en cada inferencia, lo que mejora la relación entre capacidad y coste de ejecución. Conviene, además, no quedarse anclado en comparaciones por tamaño absoluto: los modelos de generación más reciente, aun con menos

parámetros, suelen rendir mejor que modelos más grandes de hace uno o dos años, porque los avances en datos de entrenamiento y en técnicas de post-entrenamiento han sido especialmente rápidos. Una buena práctica es revisar periódicamente el panorama de modelos en lugar de dar por sentada la elección hecha al inicio del proyecto.

## 5. ¿Dónde vive el modelo y quién lo opera?

Independientemente de si el modelo es propietario u *open source*, la segunda decisión es cómo se despliega y se consume. Se articula habitualmente en tres opciones. La primera es consumir el modelo a través de la API directa del proveedor (OpenAI, Anthropic, Google): no requiere infraestructura, la puesta en marcha es rápida, el coste es casi íntegramente variable y existe una dependencia clara del proveedor en cuanto a disponibilidad y condiciones. La segunda es desplegar un modelo (normalmente *open source*) sobre infraestructura propia: ofrece control total sobre los datos y el entorno, convierte el coste del modelo en coste fijo y traslada a la organización la responsabilidad de operarlo. La tercera son los servicios gestionados como Amazon Bedrock, Azure AI Foundry o Vertex AI, que sirven tanto modelos propietarios como *open source* y ofrecen un reparto intermedio, con infraestructura gestionada por el proveedor pero mayor flexibilidad en la selección del modelo.

En el caso del despliegue propio, el tamaño y arquitectura del modelo se convierten en una decisión clave, porque determina tanto el *hardware* necesario (memoria de GPU, número de aceleradores) como la calidad alcanzable. Servir estos modelos requiere un motor de inferencia especializado: vLLM es la opción de referencia en producción por su alto rendimiento y amplio soporte de modelos; SGLang es una alternativa competitiva; Ollama simplifica enormemente la configuración y es adecuado para despliegues ligeros o entornos de desarrollo; y llama.cpp es una opción especialmente versátil, capaz de ejecutar modelos cuantizados tanto en CPU como con aceleración por GPU, lo que permite arrancar sin hardware especializado y escalar después.

## 6. Costes variables, fijos y ocultos

Más allá del precio por token publicado por cada proveedor (con diferencias de uno o dos órdenes de magnitud entre los modelos más baratos y los más capaces de una misma familia), hay varios elementos a tener presentes al comparar costes: la ventana de contexto y la longitud de las respuestas multiplican directamente el coste por interacción, los reasoning LLMs facturan tokens adicionales por su razonamiento interno aunque no formen parte de la respuesta visible, y los modelos más capaces suelen tener además mayor latencia. Para el detalle del reparto entre coste de preparación, coste fijo y coste variable puede consultarse el apartado específico sobre cómo calcular los costes de un proyecto.

Una buena práctica es que no todo el sistema tiene por qué usar el mismo modelo. En una arquitectura agéntica con varios subagentes o con un orquestador que coordina tareas heterogéneas es habitual (y muy recomendable) asignar modelos distintos a distintas funciones: reservar el modelo más capaz y caro para las tareas que realmente lo exigen (razonamiento complejo, síntesis delicada, decisiones de calidad crítica) y apoyarse en modelos más ligeros y económicos para todo lo demás (clasificación sencilla, extracción, reformulación, generación de texto rutinario). Este enrutamiento por tarea puede reducir el coste total del sistema de forma muy significativa sin afectar a la calidad percibida.

## 7. ¿Con qué urgencia necesitas la respuesta?

La latencia depende directamente del tamaño del modelo, de la longitud del contexto de entrada, de la longitud de la respuesta esperada y del modo de consumo (con una API propietaria se suma además el coste de red). En los casos de uso planteados en esta guía, todos ellos inter-sesión y ninguno interactivo en tiempo real, la latencia no es un criterio determinante: no hay personas esperando delante de una pantalla a que el sistema responda, y los tiempos de ejecución de minutos e incluso horas para un análisis completo de una sesión son perfectamente aceptables. Conviene tenerlo presente únicamente si en el futuro se añaden casos de uso interactivos, porque en

ese momento la latencia pasaría a ser un criterio de primer nivel y podría llegar a condicionar la elección del modelo.

## **8. ¿Quién controla los datos y bajo qué reglas?**

Para una administración pública europea, el encaje del modelo con el marco regulatorio y con los valores europeos es una prioridad de primer nivel. Por un lado, hay que considerar el alineamiento con el AI Act (niveles de riesgo aplicables al caso de uso, obligaciones de transparencia y de evaluación), el respeto del RGPD y, en general, los valores que la Unión Europea viene impulsando en materia de IA confiable y centrada en las personas.

Por otro, es imprescindible prestar atención a la soberanía de los datos como dimensión complementaria pero distinta: dónde vive físicamente la información que se procesa, dónde se ejecuta la inferencia, quién puede acceder a los registros, qué condiciones establecen los términos de servicio sobre el uso de los datos para reentrenamiento del modelo y qué garantías ofrece el proveedor respecto a la localización y al tratamiento. Un modelo puede ser técnicamente compatible con el marco europeo y sin embargo no cumplir los requisitos de soberanía de una administración concreta si los datos salen del territorio o pasan por jurisdicciones con garantías distintas. Ambas dimensiones deben evaluarse juntas pero de forma explícita, y tienen un peso suficiente como para condicionar las otras decisiones.

## **9. El coste ambiental que casi nadie tiene en cuenta**

El impacto ambiental de un sistema basado en LLMs tiene dos componentes. El primero es el entrenamiento del modelo, que consume cantidades ingentes de energía eléctrica y de agua para la refrigeración de los centros de datos. Aunque ese coste ya se ha producido antes de que nadie adopte el modelo, elegirlo no es una decisión neutra: seleccionar un modelo es un acto de compra que beneficia al proveedor frente a competidores que pueden haber optado por prácticas de entrenamiento más sostenibles y una administración pública comprometida con la sostenibilidad debería tener en cuenta esa dimensión. El segundo componente es el consumo de inferencia:

cada consulta al modelo tiene un coste energético e hídrico que, individualmente es pequeño pero que se acumula con el volumen de uso. El problema es que la mayoría de los proveedores no publican datos desglosados de consumo por consulta ni por modelo, lo que dificulta cualquier estimación rigurosa. Las cifras más fiables proceden de investigadores independientes y de estudios académicos, y conviene contrastarlas con las declaraciones oficiales de los proveedores que, cuando existen, tienden a ser incompletas o a reflejar escenarios optimistas. Evaluar este criterio exige, por tanto, un ejercicio activo de búsqueda y verificación de fuentes en lugar de confiar en una única referencia.

Para una administración pública europea, el consumo energético y de agua de los sistemas de IA no es solo una cuestión técnica sino también ética e institucional. El Pacto Verde Europeo, los Objetivos de Desarrollo Sostenible (en particular el ODS 7 sobre energía asequible y limpia, el ODS 12 sobre producción y consumo responsables y el ODS 13 sobre acción por el clima) y la propia Estrategia Europea de IA establecen un marco de compromisos que una administración debe poder demostrar que respeta. En la práctica, esto se traduce en decisiones concretas de diseño: elegir modelos más pequeños cuando la calidad es suficiente, enrutar tareas sencillas a modelos ligeros en lugar de usar siempre el más capaz, preferir infraestructura alimentada con energías renovables y documentar las estimaciones de consumo como parte de la memoria del proyecto cuando sea posible.

## **10. ¿De veras necesitas una IA generativa?**

La IA generativa es una herramienta muy potente, pero no siempre es la respuesta adecuada. Existen técnicas de aprendizaje automático e IA tradicional (clasificadores, sistemas basados en reglas, métodos de búsqueda y recuperación, entre otras) que para tareas bien definidas son más baratas, más rápidas, más predecibles y más fáciles de auditar. Su coste computacional y energético es órdenes de magnitud inferior, su comportamiento es determinista o estadísticamente estable, y pueden ejecutarse sobre infraestructura modesta sin dependencia de proveedores externos. Sin embargo,

muchas de estas técnicas (en particular las supervisadas) requieren datasets etiquetados de calidad y en volumen suficiente, un recurso habitualmente escaso en la administración pública: los datos existen, pero pocas veces con las etiquetas que una tarea concreta necesita, y producirlas implica un esfuerzo manual considerable. Los sistemas basados en reglas, por su parte, se vuelven inmanejables a medida que crece la casuística. Y ninguna de estas alternativas resuelve bien las tareas propiamente generativas: comprender texto no estructurado en toda su riqueza, sintetizar, reformular o producir lenguaje natural coherente a partir de insumos heterogéneos.

En la práctica, la arquitectura más sensata suele ser híbrida: usar técnicas clásicas para aquellos subproblemas en los que resultan suficientes y recurrir al LLM solo cuando la complejidad del lenguaje o la naturaleza generativa de la tarea lo justifican. Los propios casos de uso propuestos en este documento incorporan ya este enfoque: la búsqueda vectorial que sustenta los sistemas de RAG no es IA generativa, la deduplicación y el agrupamiento de aportaciones se puede realizar con técnicas clásicas, y solo cuando se requiere interpretar, resumir o redactar entra en juego el modelo del lenguaje. La regla práctica es simple: usar la herramienta más simple, más barata y más predecible que resuelva el problema con la calidad requerida, y reservar los LLMs para aquello en lo que aportan un valor que ninguna otra técnica puede ofrecer.

## Cómo tener el control de lo que está pasando

Mantener el control de un sistema basado en IA generativa no consiste en tratarlo como una aplicación determinista (no lo es), sino en disponer de los mecanismos necesarios para observar cómo se comporta, medir si lo hace como se espera y reaccionar con rapidez cuando algo cambia. En este tipo de sistemas, el comportamiento depende de múltiples componentes (el modelo subyacente, el prompt del agente, las herramientas que tiene a su disposición, los insumos documentales, las versiones de las APIs de terceros) y cualquiera de ellas puede evolucionar, algunas de forma consciente y otras sin previo aviso. Por eso el control se articula sobre dos prácticas complementarias: la evaluación, que permite medir la calidad del sistema frente a criterios predefinidos, y la observabilidad, que permite ver qué está ocurriendo en producción a medida que ocurre.

### Evaluación

La evaluación es el proceso sistemático de medir la calidad de las salidas del sistema frente a un conjunto de criterios definidos por adelantado. En los casos de uso planteados, las dimensiones más relevantes suelen ser la fidelidad a las fuentes (que el sistema no añada información que no esté en los insumos), la corrección, la consistencia entre ejecuciones sobre entradas similares, la ausencia de alucinaciones y el respeto a los guardrails definidos (tono, longitud, formato, temas excluidos).

Para medirlas se recurre a un dataset de evaluación representativo (un conjunto de entradas con respuestas o criterios de referencia) sobre el que se ejecuta el sistema de forma automatizada, habitualmente con el apoyo de bibliotecas especializadas como Ragas (orientada a sistemas que usan RAG) o Deepeval (para evaluación más general de LLMs), y complementado con revisión humana periódica sobre una muestra de las salidas reales.

Este dataset no sirve únicamente para validar el sistema antes de un primer despliegue: es, sobre todo, la red de seguridad que permite detectar

regresiones cuando algo cambia. Los sistemas basados en LLM son especialmente sensibles a variaciones aparentemente menores y cualquier cambio, por ejemplo, una actualización del modelo por parte del proveedor (éstos pueden retirar una versión, introducir un nuevo comportamiento por defecto o aplicar ajustes de seguridad que alteran las salidas), un cambio de modelo motivado por costes, una modificación del prompt del agente o incluso una versión nueva de una herramienta que el agente invoca. Todo esto puede provocar caídas de calidad que de otro modo pasarían inadvertidas hasta que una persona usuaria las notase. Ejecutar la evaluación automática antes y después de cualquier cambio, y bloquear el despliegue si las métricas empeoran por encima de un umbral, es la única manera realista de mantener el sistema estable a lo largo del tiempo.

## Observabilidad

La observabilidad, por su parte, es la capacidad de inspeccionar, registrar y auditar lo que el sistema hace en producción a medida que lo hace. A diferencia de la evaluación, que se ejecuta sobre un conjunto controlado de casos, la observabilidad se alimenta del uso real. Implica capturar de forma estructurada los elementos relevantes de cada ejecución: la entrada recibida, los prompts efectivamente enviados al modelo, las llamadas a herramientas realizadas por el agente, los resultados intermedios, la salida final, los tokens consumidos, la latencia de cada paso y los errores encontrados. Existen plataformas específicas para esta tarea, como Langfuse (open source y auto-hospedable) o LangSmith (en su versión SaaS), que permiten inspeccionar trazas completas, filtrar por criterios, agregar métricas, etiquetar y detectar anomalías de forma interactiva.

La observabilidad es imprescindible por varios motivos complementarios. En primer lugar, permite diagnosticar incidentes cuando una ejecución concreta produce una salida inesperada: sin trazas detalladas es prácticamente imposible reconstruir por qué el agente tomó una decisión determinada.

En segundo lugar, hace visibles los usos imprevistos del sistema por parte de las personas usuarias: en la práctica, cuando un sistema se pone en manos de gente, aparecen consultas, flujos y escenarios que el equipo de diseño no había anticipado (entradas con formatos inesperados, preguntas fuera del alcance previsto, intentos de utilizarlo para tareas adyacentes), y detectarlos a tiempo permite ajustar los guardrails, ampliar el dataset de evaluación o incluso replantear el caso de uso antes de que el problema escale. En tercer lugar, es la vía más rápida para detectar regresiones silenciosas que hayan escapado a la evaluación automática, especialmente las provocadas por actualizaciones del proveedor del modelo: una caída súbita de la tasa de éxito, un cambio en la longitud media de las respuestas o un aumento inusual del consumo de tokens son señales que solo emergen si se monitoriza de forma continuada. Finalmente, facilita el control de costes en tiempo real y proporciona el rastro de auditoría que puede ser necesario para rendir cuentas de decisiones apoyadas en IA.

Ahora bien, en un contexto de administración pública la observabilidad debe convivir con las exigencias de privacidad y de protección de datos, sin entrar en conflicto con ellas. Por su propia naturaleza, las trazas que se capturan contienen información sobre lo que el sistema ha procesado, y en función del caso de uso esa información puede incluir datos personales o material sensible. Diseñar la capa de observabilidad implica, por tanto, decidir desde el principio qué se registra y qué no, aplicar anonimización en origen cuando proceda, establecer políticas de retención acotadas, restringir el acceso a las trazas mediante controles de autorización auditables y documentar cómo se tratan estos registros ante las personas afectadas y ante la propia autoridad de protección de datos. En los casos de uso planteados en esta guía, el riesgo se mitiga en buena medida por dos circunstancias: el sistema opera en un entorno controlado, al ser manejado por personal técnico de la administración y no expuesto directamente a la ciudadanía, y los insumos más sensibles se anonimizan antes de entrar en el pipeline. Aun así, conviene tratar la privacidad como un requisito de diseño en todo el sistema, incluyendo la capa de observabilidad, no como un ajuste posterior, porque

en el ámbito público una fuga o un uso indebido de las trazas puede tener un gran impacto sobre la confianza en el propio proceso participativo.

### Visión de conjunto

Evaluación y observabilidad funcionan mejor cuando se conciben como un ciclo cerrado: los casos interesantes detectados en producción a través de la observabilidad (tanto fallos como usos imprevistos de valor) se incorporan al dataset de evaluación, que así crece y se hace más representativo con el tiempo; y las métricas que la observabilidad calcula sobre el tráfico real se contrastan periódicamente con las de la evaluación controlada para detectar desviaciones. Como buenas prácticas generales, conviene instrumentar el sistema desde el primer día en lugar de como un añadido posterior, establecer umbrales y alertas que activen una revisión humana ante desviaciones, revisar manualmente una muestra de trazas con regularidad en lugar de confiar solo en métricas agregadas, y definir por adelantado un plan de respuesta para el momento en que se detecte una regresión (*rollback* al modelo o *prompt* anterior, comunicación a las personas usuarias, ajuste del *dataset* de evaluación).

### El factor humano

Junto a la evaluación y la observabilidad, la revisión humana constituye el tercer pilar del control del sistema y conecta esta sección con un principio de diseño que atraviesa toda la guía: las salidas del agente son siempre insumos para una persona, nunca decisiones finales. En la práctica, esta revisión humana desempeña tres funciones distintas y complementarias.

En primer lugar, actúa como control de despliegue ante cualquier cambio relevante (de modelo, de *prompt*, de herramientas o de versión del proveedor), revisando tanto las métricas de la evaluación automática como una muestra cualitativa de las salidas del sistema, porque los números por sí solos pueden ocultar regresiones sutiles que solo se aprecian leyendo las respuestas.

En segundo lugar, opera como muestreo periódico sobre producción, auditando con regularidad una selección de trazas reales capturadas por la observabilidad (no solo cuando se dispara una alerta), y es donde suelen aflorar sesgos emergentes, pérdidas de fidelidad difícilmente cuantificables y usos imprevistos del sistema.

En tercer lugar, mediante revisión de los materiales generados por parte de personal especializado y uso del patrón *human-in-the-loop* en los casos aplicables, se integra en el propio flujo operativo como modo de uso por defecto del sistema: cada salida producida por el agente es revisada, contrastada y, en su caso, ajustada por el personal técnico antes de cualquier aplicación. Esta tercera capa es cualitativamente distinta de las otras dos porque no se ejecuta sobre datos ni sobre versiones, sino sobre la propia interacción en producción, y es precisamente lo que permite convivir con el carácter no determinista de los LLMs sin asumir riesgos desproporcionados.

## Cómo calcular los costes de un proyecto

Calcular con un grado razonable de precisión el coste de un proyecto basado en IA generativa requiere distinguir entre tres categorías que responden a dinámicas muy distintas: el coste de preparación (lo que se invierte una sola vez, antes de poner el sistema en funcionamiento), el coste fijo del servicio y la infraestructura (lo que se paga de forma recurrente exista o no uso real) y el coste variable por consumo (lo que escala con la intensidad de uso del sistema). Cada categoría se comporta de forma diferente a lo largo del ciclo de vida del proyecto, y la elección de arquitectura puede desplazar peso entre unas y otras: invertir más en preparación (mejores *prompts*, una evaluación rigurosa, *guardrails* bien definidos) suele reducir el coste variable al evitar iteraciones innecesarias; optar por un modelo auto-hospedado traslada coste del variable al fijo; y un alcance bien acotado limita tanto la inversión inicial como el consumo recurrente.

## Coste de preparación

Bajo este epígrafe se agrupan los costes no recurrentes necesarios para llevar el sistema desde una idea hasta un servicio operativo. Incluye el análisis y diseño del caso de uso (alcance, criterios de éxito, definición de las entradas y salidas esperadas), el diseño técnico del sistema (selección del modelo, arquitectura del agente, definición de las herramientas y de los esquemas de salida estructurada, redacción y ajuste de los *prompts*) y la preparación de los insumos documentales que el sistema consultará: ingesta, limpieza, estructuración y, cuando proceda, anonimización de datos en documentación o aportaciones de procesos anteriores.

A esto se añaden dos partidas que con frecuencia se subestiman pero que son determinantes para la calidad final: por un lado, la construcción de un pequeño dataset de evaluación representativo del caso de uso y la definición de las métricas con las que se medirá el sistema (fidelidad a las fuentes, corrección, consistencia, ausencia de alucinaciones); por otro, la fase iterativa de pruebas sobre ese dataset, que es donde se ajustan *prompts*, *guardrails* y umbrales antes del despliegue. También corresponden a esta categoría las integraciones necesarias con los sistemas existentes (plataforma de participación, almacenamiento, autenticación) y la documentación para el equipo que operará el sistema.

En los casos de uso planteados, no se requiere el uso un *fine-tuning*, que suele formar parte de este coste: los modelos base actuales ofrecen calidad suficiente si se combinan con un *prompting* cuidadoso y con RAG cuando hace falta aportar información específica del proceso. Esto permite concentrar la inversión inicial en el diseño del sistema y en su evaluación, en lugar de en el entrenamiento extra del modelo.

## Coste fijo del servicio y la infraestructura

Bajo esta categoría se agrupan los costes que la organización asume de forma recurrente (habitualmente mensual o anual) con independencia del

volumen real de uso del sistema. Incluye el hospedaje de la aplicación (ya sea sobre servidores propios, contenedores gestionados o servicios cloud), el almacenamiento persistente que exige el caso de uso (bases de datos relacionales para el estado operativo, bases de datos vectoriales si se implementa RAG, almacenamiento de logs y de los artefactos auditables que respalden las decisiones) y los servicios de monitorización y observabilidad del stack de IA, como Langfuse (auto-hospedable) o LangSmith en su versión SaaS. A ello se suman el mantenimiento y soporte continuados (actualizaciones, parches de seguridad, evolución frente a cambios en las APIs de los proveedores o al ritmo de aparición de nuevos modelos) y las licencias o suscripciones de herramientas auxiliares.

Es también en esta categoría donde cobra importancia la decisión sobre cómo se consume el modelo del lenguaje, una cuestión que se aborda con más detalle en el apartado de criterios para seleccionar un modelo. Como regla general, un modelo open source desplegado sobre infraestructura propia (habitualmente sobre GPUs, on-premise o en cloud privado) convierte casi todo el coste del modelo en coste fijo: se paga la capacidad disponible exista o no tráfico, a cambio de un coste marginal por petición muy bajo y de un mayor control sobre los datos. En cambio, los modelos servidos a través de una API, ya sean propietarios (OpenAI, Anthropic, Google) o modelos *open source* ofrecidos por servicios gestionados como Amazon Bedrock, Azure AI Foundry, Vertex AI u otros similares, desplazan casi todo el coste del modelo a la categoría variable: el coste fijo se limita entonces a la infraestructura que rodea al modelo, y el consumo del propio modelo se factura por tokens efectivamente procesados.

### Coste variable por consumo

El coste variable recoge todo aquello que escala directamente con el uso real del sistema. La partida principal suele ser el consumo del propio LLM, facturado en función de los *tokens* de entrada y de salida que procesa cada llamada al modelo: cuanto más largo el contexto que se le proporciona y más extensa la respuesta, mayor el coste por inferencia.

Conviene tener presente que los *reasoning LLMs* consumen, además, *tokens* adicionales para su razonamiento interno que también se facturan aunque no formen parte de la respuesta visible, y que tecnologías como RAG introducen llamadas adicionales de recuperación (*embeddings*) que contribuyen igualmente a esta partida.

A este coste se suman los servicios auxiliares: APIs de *speech-to-text* y diarización (facturadas habitualmente por minuto de audio procesado), APIs de *text-to-speech* (por carácter o por minuto generado), APIs de generación de imágenes multimodales (por imagen producida) y las llamadas a herramientas externas que el agente pueda invocar durante su razonamiento (búsquedas, consultas a APIs de pago, accesos a bases documentales de terceros). Cada una de estas partidas contribuye al coste por interacción y debe tenerse en cuenta al planificar el presupuesto global del servicio.

A la hora de estimar este coste es importante distinguir entre dos patrones de uso muy diferentes. En los escenarios de inferencia controlada (aquellos en los que las personas usuarias del sistema son el propio personal técnico de la administración, que lo opera en momentos y para tareas concretas) el volumen de peticiones es acotado y relativamente predecible, lo que permite estimar el coste a partir del número de sesiones o aportaciones previsto y mantener el presupuesto bajo control. En los escenarios de uso abierto a la ciudadanía, en cambio, la demanda es mucho más difícil de anticipar: el número de personas usuarias y la intensidad con que interactúen con el sistema pueden variar de forma significativa en función de la visibilidad del proceso, de su cobertura mediática o de los incentivos a participar, de modo que el coste variable puede oscilar notablemente y requiere instrumentar salvaguardas adicionales (límites por sesión, colas, tope de gasto, estrategias de caché, degradación controlada ante picos).

Los casos de uso planteados en esta guía encajan en el patrón de inferencia controlada, ya que el sistema está pensado para ser utilizado por personal de la administración sobre insumos concretos de cada proceso participativo, y no como un servicio expuesto directamente a la ciudadanía.

Esto facilita la previsión presupuestaria, pero conviene tener presente esta distinción al trasladar cualquiera de estas arquitecturas a otros escenarios donde la demanda no esté acotada. En todo caso, una buena práctica desde el diseño es instrumentar el sistema para medir el coste por unidad útil (por sesión analizada, por participante, por material generado) y no solo el coste agregado; ese indicador, imposible de afinar sin una fase piloto previa, es la única manera realista de trasladar el coste del proyecto a magnitudes comparables con las del resto de actuaciones de participación.

## Soporte para lenguas co-oficiales

Los sistemas propuestos deben funcionar en castellano, pero una administración pública española tiene la obligación de atender a la ciudadanía también en catalán, gallego y euskera. El soporte lingüístico condiciona tanto la elección del modelo del lenguaje como la del modelo de embeddings utilizado en los componentes de búsqueda y recuperación, y el nivel de soporte varía considerablemente entre lenguas.

Las últimas generaciones de modelos propietarios (la familia GPT-5 de OpenAI, Claude de Anthropic, Gemini de Google) ofrecen un soporte sólido en castellano y un rendimiento razonable en catalán y gallego. El Euskera es un caso distinto: su estructura aglutinante, su distancia tipológica con las lenguas romances y su menor presencia en los corpus de entrenamiento hacen que el rendimiento sea inferior. Según el *benchmark* [IberBench](#), solo el 13% de los LLMs evaluados incluyen el euskera en sus datos de entrenamiento o ajuste, frente al 35% para catalán y el 26% para gallego, lo que se traduce directamente en diferencias de calidad.

En el ámbito *open source*, los modelos generalistas (Mistral, Gemma de Google, Qwen de Alibaba, entre otros) presentan un soporte desigual que depende del corpus concreto de cada versión.

Existen, sin embargo, iniciativas específicas de gran valor. La más relevante es el proyecto ALIA, impulsado por el Gobierno de España y ejecutado por el Barcelona Supercomputing Center (BSC), que ha producido el [ALIA Kit](#): un conjunto abierto de modelos de texto, voz, traducción automática y multimodalidad entrenados en castellano y las lenguas co-oficiales. Su modelo de texto insignia es ALIA-40b (40 mil millones de parámetros, entrenado en 35 lenguas europeas sobre el supercomputador MareNostrum 5), cuya versión más reciente es ALIA-40b-instruct-2601 de enero de 2026; la familia Salamandra (2B y 7B), desarrollada también con financiación del proyecto AINA de la Generalitat de Catalunya y con alta representación del catalán en sus datos de entrenamiento, completa la oferta para despliegues con menores requisitos de hardware. Fuera del BSC, el centro HiTZ de la UPV/EHU ha producido [Latxa](#) (7B-70B), el primer LLM específico para euskera; y el CiTIUS de la USC ha creado [Carballo](#) para gallego. Estos modelos especializados son una alternativa concreta cuando los generalistas no alcanzan la calidad necesaria en una lengua determinada.

Para los componentes de búsqueda semántica y recuperación (RAG), el soporte lingüístico del modelo de embeddings es igual de determinante. Un modelo que no represente bien una lengua producirá búsquedas semánticas de baja calidad, impidiendo que documentos relevantes se recuperen correctamente. Entre las opciones actuales con soporte multilingüe (que incluye catalán, gallego y euskera) destacan: EmbeddingGemma de Google; Qwen3-Embedding de Alibaba (disponible en variantes de 0,6B a 8B parámetros); BGE-M3 de BAAI que ofrece recuperación densa, dispersa y multi-vector en un solo modelo; y Cohere Embed v4 y text-embedding-3-large de OpenAI proporcionan alternativas propietarias multilingües competitivas. En todos los casos, el rendimiento en catalán, gallego y especialmente en euskera debe verificarse explícitamente: los benchmarks multilingües suelen centrarse en lenguas mayoritarias y no garantizan calidad equivalente en lenguas con menor representación en los datos de entrenamiento.

La recomendación práctica es no asumir que el rendimiento en castellano se traslada automáticamente a las lenguas co-oficiales, y evaluar cada lengua

de forma independiente tanto en el LLM como en el modelo de *embeddings*, utilizando el marco de evaluación descrito en secciones anteriores. Conviene también considerar la detección automática de lengua y el enrutamiento a modelos o configuraciones optimizadas para cada una. El soporte lingüístico mejora con cada generación de modelos, lo que refuerza la importancia de la evaluación periódica como práctica permanente y no como una validación puntual.

## Consideraciones éticas y de transparencia

El uso de herramientas basadas en inteligencia artificial en procesos de participación ciudadana, ya sea para consultar insumos normativos, analizar aportaciones deliberativas, generar materiales de socialización, elaborar rendición de cuentas o ampliar la convocatoria, exige un marco sólido de garantías éticas, jurídicas e institucionales.

**Protección de datos y minimización.** Todo tratamiento de información derivada de procesos participativos debe ajustarse al Reglamento General de Protección de Datos (RGPD). Con carácter general, los insumos procesados por los sistemas automatizados deberán estar anonimizados o seudonimizados, limitarse a los datos estrictamente necesarios y utilizarse exclusivamente para los fines del proceso participativo. En el caso de comunicaciones personalizadas, los materiales generados no podrán revelar datos atribuibles a participantes concretos, salvo que dicha información sea pública o haya sido expresamente autorizada.

**Transparencia algorítmica y trazabilidad.** Los criterios mediante los cuales los sistemas seleccionan, agrupan, priorizan o vinculan contenidos deben ser públicos, comprensibles y auditables. Ello implica documentar las fases de análisis, los criterios de clasificación y la forma en que los resultados generados se relacionan con los insumos originales.

La capacidad de rastrear cada conclusión o material producido hasta sus fuentes ciudadanas es un requisito de legitimidad del proceso. Los sistemas no deben introducir jerarquizaciones opacas ni valoraciones normativas implícitas que condicionen la interpretación del contenido.

**Supervisión humana y neutralidad institucional.** Los resultados generados por la IA deben entenderse como insumos de apoyo, no como decisiones o conclusiones definitivas. La validación, interpretación y selección final de contenidos, incluidos los materiales de comunicación, los dilemas a deliberar o los destinatarios de la convocatoria, corresponde siempre a los equipos humanos responsables del proceso. La herramienta no debe generar contenido nuevo ni sugerir interpretaciones, recomendaciones o decisiones que sean competencia exclusiva de los participantes y de los órganos democráticos.

**Auditabilidad y código abierto.** En la medida de lo posible, se recomienda utilizar herramientas basadas en estándares abiertos o software auditable, lo que facilita la revisión independiente por parte de la comunidad técnica, académica e institucional, reduce la dependencia de proveedores privados y refuerza la confianza pública. Asimismo, deben preverse mecanismos periódicos de evaluación para detectar posibles sesgos en las fuentes utilizadas o en los perfiles identificados, dado que los datos disponibles no representan necesariamente a todos los sectores sociales por igual.

**Transparencia hacia la ciudadanía.** Las personas participantes deben ser informadas de forma clara cuando sus aportaciones puedan ser analizadas o utilizadas mediante herramientas de inteligencia artificial, indicando la finalidad del tratamiento y las garantías aplicadas. Del mismo modo, cuando los materiales difundidos hayan sido elaborados con apoyo de IA, esta circunstancia debe comunicarse de manera explícita. Esta transparencia no es solo un requisito normativo, sino un elemento constitutivo de la confianza en el proceso participativo.

## Recomendaciones para su uso ético y responsable en entornos seguros de participación juvenil

Cuando las herramientas basadas en inteligencia artificial se utilizan en procesos participativos dirigidos a población infantil y joven, resulta imprescindible adoptar garantías adicionales que protejan los derechos, la privacidad y la libertad de expresión de las personas menores de edad, y que preserven la integridad pedagógica y deliberativa del proceso.

**Protección reforzada de datos y prohibición de perfilado individual.** Debe evitarse la identificación o recopilación directa de información personal sobre menores a partir de fuentes abiertas. El contacto con colectivos juveniles debe realizarse preferentemente a través de organizaciones, asociaciones o centros educativos ya constituidos. Los sistemas automatizados deben trabajar sobre información agregada o anonimizada, centrándose en patrones colectivos y no en intervenciones individuales. En ningún caso la herramienta debe evaluar individualmente a las personas participantes ni generar perfiles de opinión.



**Transparencia activa y comunicación accesible.** Las personas jóvenes participantes deben ser informadas de forma clara, comprensible y adaptada a su edad sobre cómo han sido identificadas o invitadas, cómo se utilizarán sus aportaciones y qué papel desempeña la inteligencia artificial en el proceso. Los materiales generados, ya sean de dinamización, análisis o rendición de cuentas, deben emplear lenguaje claro, formatos accesibles (infografías, vídeos, resúmenes visuales) y evitar formulaciones que simplifiquen en exceso realidades complejas o induzcan respuestas predeterminadas. Debe comunicarse explícitamente que la herramienta no evalúa a los participantes ni influye en el resultado del proceso.

**Espacios seguros, inclusivos y no vigilantes.** El uso de herramientas automatizadas no debe generar en las personas jóvenes ninguna percepción de seguimiento individualizado o vigilancia. Las dinámicas participativas deben preservar espacios suficientes para la exploración libre de ideas,

el desacuerdo constructivo y la evolución de posiciones a lo largo del tiempo, evitando que los resultados del análisis automatizado cierren prematuramente el debate. Se recomienda realizar pruebas piloto antes de utilizar materiales generados con IA en procesos reales, especialmente con población joven, para verificar su comprensibilidad, pertinencia y ausencia de sesgos.

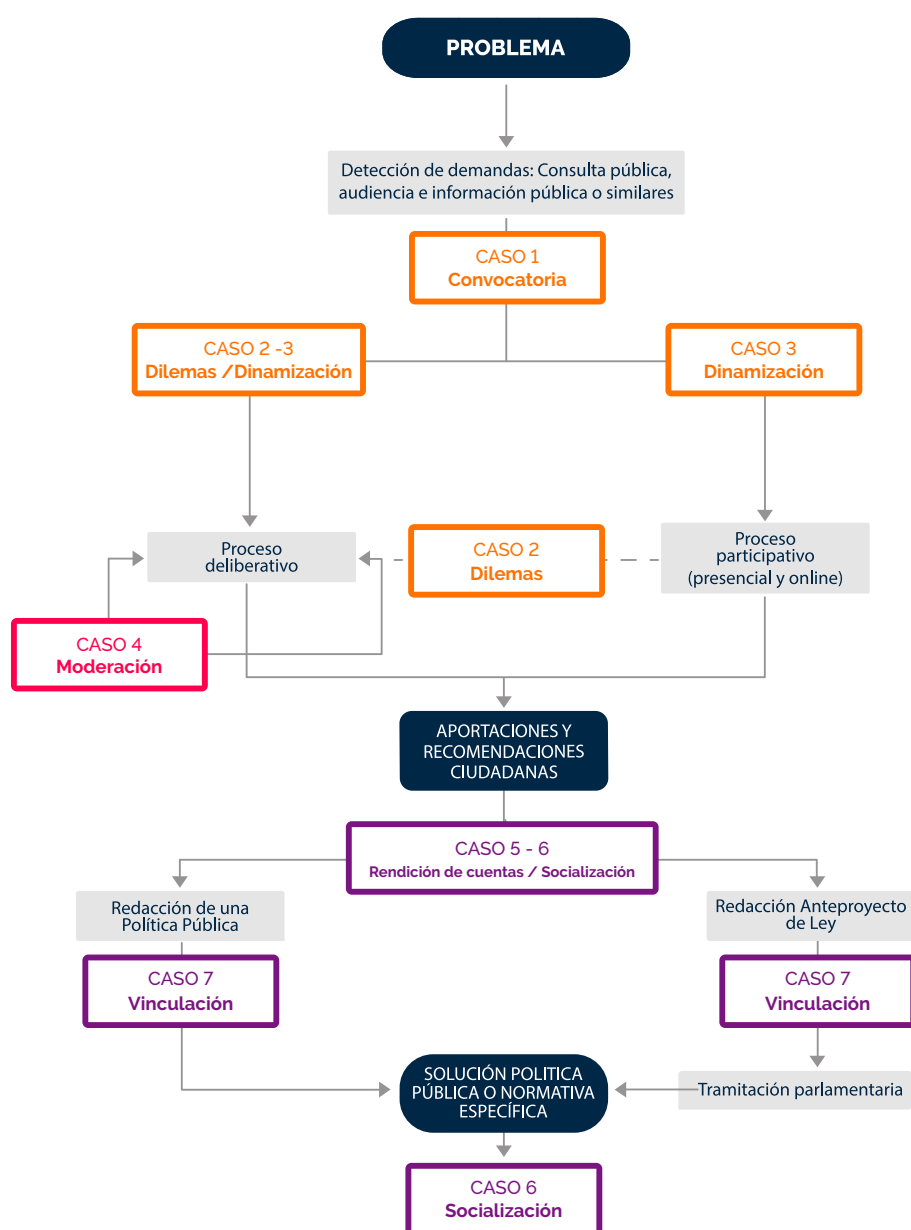
**Neutralidad, pluralidad y soberanía institucional.** La presentación de insumos y resultados debe reflejar la diversidad de enfoques existentes, sin métricas ni visualizaciones que orienten hacia interpretaciones concretas. Se recomienda el uso de infraestructuras públicas o soberanas, con componentes auditables y garantía de que los datos no se reutilizan con fines distintos al proceso participativo. El sistema debe estar bajo responsabilidad pública claramente identificada, con supervisión humana continua de su funcionamiento.

**Evaluación continua y retroalimentación.** Se recomienda establecer mecanismos periódicos de evaluación del impacto de estas herramientas en la percepción y la experiencia de las personas jóvenes participantes, incluyendo espacios para que puedan comentar, preguntar o reaccionar a los materiales recibidos. Esta retroalimentación permite ajustar progresivamente el uso de la tecnología y garantizar que contribuye a crear entornos participativos seguros, significativos e inclusivos para la infancia y la juventud.



# Introducción a los estudios de caso

A continuación, se presentan **7 estudios de caso** en diferentes fases del proceso participativo y deliberativo, cada uno de los casos presenta un reto común que enfrentan las instituciones a la hora de gestionar, dinamizar y sistematizar los resultados del proceso. En cada uno de ellos a su vez, se presenta una propuesta de solución basada en técnicas de IA, estas soluciones no son deterministas, sino que buscan ser fichas técnicas y orientadas a la acción, que permitan encontrar alternativas viables a quienes se encuentran ante uno de estos dilemas al organizar o gestionar un proceso participativo.



## PREVIO

La preparación y montaje de un proceso participativo o deliberativo es decisiva, de la claridad que se tenga al momento de definir participantes, reglas del proceso, insumos con los que se trabaja, dilemas a discutir, y materiales para la dinamización; dependerá el tipo de resultados que se obtengan, la fluidez del proceso en su conjunto, la comodidad de las personas que participan y la capacidad para obtener insumos de calidad que alimenten posteriormente procesos siguientes.

Se presentan por tanto en esta fase **soluciones para tres retos** concretos:

### ■ **Caso 1: Convocatoria automatizada a procesos participativos**

¿Cómo garantizar que las personas y colectivos que son convocados a ser parte de un proceso participativo sean realmente diversos y no los mismos de siempre? Este caso de uso busca dar respuesta a este interrogante, enfocándose en las posibilidades que ofrece la explotación de fuentes de información a las que la Administración tiene acceso, y cómo combinarlas con otras fuentes abiertas para detectar a perfiles y colectivos tradicionalmente marginados e invisibilizados.

### ■ **Caso 2: Selección de dilemas deliberativos a partir del análisis de insumos previos**

Un proceso participativo podría definirse en cierta manera como un espacio de debate libre donde escuchar las opiniones y donde recoger aportaciones de colectivos y personas diversas acerca de un tema. Sin embargo, ser capaces de definir el tema sobre el cual debatir, y saber plantear las preguntas correctas implica tener claridad acerca de cuáles son las principales necesidades y demandas ciudadanas. Lograr sistematizar y ordenar todos los insumos y recursos con los que cuenta una Administración para poder escoger un número limitado de dilemas sobre los cuáles debatir es el eje de este reto.

### ■ **Caso 3: Generación de materiales dinamizadores a partir de insumos previos**

Convocadas las personas que participarán del proceso y definidos los dilemas llega el momento de organizar el proceso en sí mismo, asegurar que fluya, que se cuenta con materiales que facilitan el debate y que todas las personas involucradas se sienten cómodas y escuchadas en cada uno de sus momentos. Esto requiere de entender por un lado los objetivos a conseguir a través del debate, pero también de materiales que hagan la tarea de quienes dinamizan algo más sencilla. Este reto se centra en cómo generar recursos y materiales que sean innovadores y al mismo tiempo faciliten el desarrollo de procesos dinámicos.

## **DURANTE**

El principal reto detectado durante el proceso participativo, y sobre todo en el caso de procesos deliberativos, es asegurar una dinamización eficiente, que sea capaz de procesar todos los recursos y aportaciones realizadas en cada día o jornada de debate, sin perder matices y asegurando que todas las voces son igualmente escuchadas.

### ■ **Caso 4: Análisis automatizado para la moderación y la detección de consensos**

El foco está puesto en la automatización del análisis del contenido de las sesiones y en la obtención de insumos que faciliten la fluidez en la moderación.

## **POSTERIOR**

La rendición de cuentas y la transparencia con los resultados de un proceso participativo son cuentas pendientes en la mayoría de las democracias, la capacidad de mejorar cómo se devuelve a la ciudadanía y demás actores involucrados el resultado e impacto de sus aportaciones e insumos,

y al mismo tiempo ofrecer canales para que el proceso se retroalimente y vuelva a comenzar requiere de una adecuada sistematización de resultados y de un seguimiento continuo de cómo cada aportación impacta en los resultados finales de una política o normativa concreta.

Se presentan por tanto en esta fase **tres casos**:

■ **Caso 5: Elaboración automatizada de materiales de rendición de cuentas**

¿Es posible hacer devoluciones personalizadas a cada persona o colectivo participante acerca del impacto que ha tenido su aportación en el proceso? En este reto se busca dar respuesta a este interrogante, cómo se sistematiza la información de manera adecuada y a la vez respetando la normativa vigente de protección de datos, cómo se le da seguimiento a cada aportación individual en el marco de un proceso colectivo y algunas cuestiones más.

■ **Caso 6: Producción de materiales para la socialización de resultados**

Socializar resultados implica ser capaces de comunicar de manera efectiva qué pasó durante el proceso, qué resultados se obtuvieron y qué va a pasar a continuación. Todo esto en un lenguaje claro, orientado a audiencias amplias y con el objetivo de mantener un vínculo cercano, o generar un vínculo antes inexistente, entre la Administración y la ciudadanía.

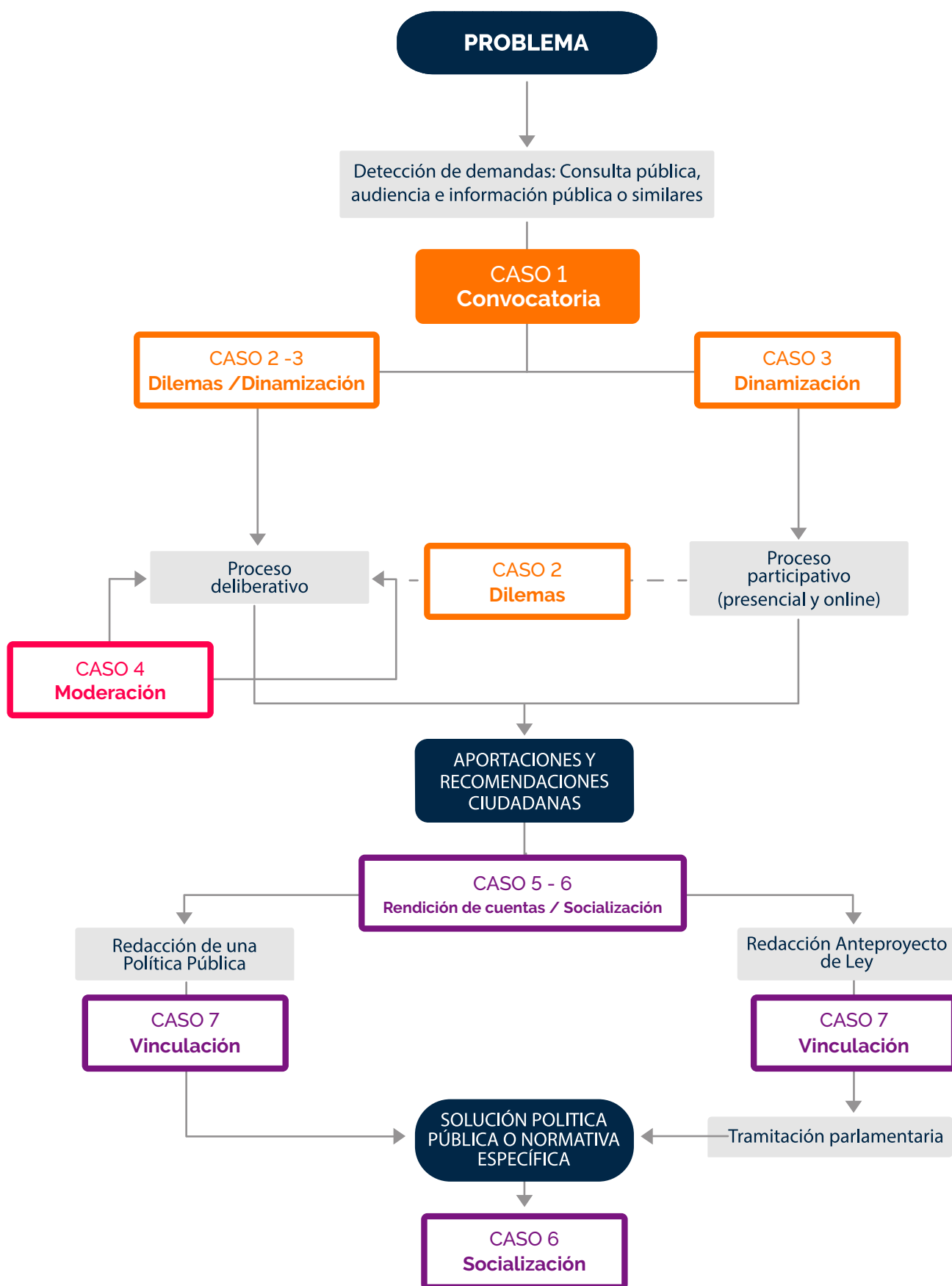
■ **Caso 7: Vinculación entre insumos participativos y producción normativa**

'Participar no cambia nada', 'los políticos no nos escuchan' y así un largo etcétera. ¿Cómo podemos vincular de manera directa las aportaciones ciudadanas con medidas específicas de una política pública o con artículos concretos de un Proyecto de Ley? Este caso aborda el reto de vincular de manera clara y directa las aportaciones recibidas con el resultado final, una manera innovadora de rendir cuentas, y de mostrar a quienes toman las decisiones que las voces ciudadanas han colaborado en la construcción de nuevas normativas.



Convocatoria  
automatizada  
a procesos  
participativos

# Diagrama



# 1.

## Convocatoria automatizada a procesos participativos

### Descripción del caso de uso

Se busca **garantizar que las personas y colectivos que son convocados a ser parte de un proceso participativo sean de perfiles y orígenes diversos y que vayan más allá de los que ya mantienen una relación previa con la Administración que lo convoca.** De las muchas posibles estrategias que existen para responder a este objetivo, el caso de uso se centrará en la explotación de fuentes de información que estén a disposición de las Administraciones públicas, ya sea porque sean datos originados por las propias administraciones públicas (bases de datos y registros públicos, procesos participativos previos...) o porque estén disponibles de algún modo (redes sociales, medios de comunicación, bases de datos generadas por la sociedad civil, etc).

### Actores involucrados

Se recomienda que, en las primeras fases de desarrollo de esta solución, se involucren en mayor o menor medida los siguientes actores

- **Personal técnico del Ministerio involucrado:** es clave que el personal a cargo de definir los objetivos y el impacto esperado del proceso, identifique de manera clara por un lado, a qué audiencia quiere llegar con los resultados, y por otro, los colectivos potencialmente

afectados por las posibles conclusiones y soluciones propuestas dentro del proceso participativo. En este sentido vale la pena preguntarse de manera amplia ¿qué tipo de soluciones se buscan conseguir?, ¿a qué tipo de necesidades se quiere dar respuesta? Responder a estas y otras cuestiones será clave para comenzar a ahondar en los posibles actores relevantes que requieren tener voz en el proceso.

Será clave en el mismo sentido, el contar con la base de datos de contactos histórica de actores con los cuales se suele colaborar o se ha colaborado en el pasado, para poder aprovechar esos datos como un punto de partida que permita iniciar y ampliar la búsqueda.

- **Personal técnico a cargo de diseñar proceso participativo/deliberativo:** más allá del personal técnico del Ministerio, es clave poder dialogar también con quienes estarán a cargo de la dinamización del proceso participativo en sí mismo, y poder valorar si poseen otros contactos o fuentes de información alternativas que permitan mapear a actores, colectivos o personas clave cuyas voces aporten miradas diferentes al proceso.
- **Stakeholders clave:** de manera opcional, se recomienda contar en las primeras fases de desarrollo de la solución con *stakeholders* pertenecientes a colectivos potencialmente afectados. Aportarán valor especialmente aquellos que pertenezcan a redes o alianzas grandes, o que formen parte de ecosistemas amplios de actores, ya que permitirán detectar aquellos puntos débiles del mapeo realizado. Además, permitirán definir aquellos criterios mínimos que deberá cumplir el mapeo en cuanto a colectivos que deberían ser representados para asegurar una diversidad real.

## Causas y consecuencias

Conseguir que un mayor número de personas participe en espacios participativos impulsados desde lo público ya es un reto, lograr que quienes participen pertenezcan a colectivos variados, que no sean siempre las mismas personas y que participen sobre todo quienes más reniegan a veces del sistema es una tarea cuando menos complicada. Las consecuencias de esta falta de representación amplia son variadas y conocidas, quitan legitimidad a los procesos, evitan que muchos colectivos se apropien de las conclusiones de estos procesos, y no permiten procesos de toma de decisiones amplios y con fuerte consenso.

La aplicación de soluciones basadas en IA probablemente no solucione este problema por completo, pero puede aportar ciertas facilidades al tener el potencial de detectar, a partir de un análisis masivo de datos, qué perfiles se están quedando fuera o cuáles están sobrerrepresentados, sobre todo en procesos masivos y con muchos participantes.

Sin embargo, un mal uso de la tecnología, o errores en su configuración, podrían tener ciertas consecuencias que afecten la legitimidad del proceso, o que impliquen una doble exclusión de ciertos colectivos, la humana y la de la máquina. Por ello resulta clave tener en cuenta ciertas cuestiones, una de ellas y tal como explica *Sammy McKinney (2024)* es el 'sesgo de selección' al analizar grandes volúmenes de datos, y es que resulta clave una adecuada configuración de los datos de entrenamiento de la IA para que pueda justamente localizar a aquellos colectivos que hasta la actualidad han estado infrarrepresentados, y no a la inversa. En esta primera 'configuración', resultará útil contar con la voz de colectivos variados y diversos, vinculados de manera directa, pero también indirecta, con el proceso participativo en cuestión. Las aportaciones de estos colectivos a la fase de 'afinación' de la herramienta favorecerá a la reducción de esos potenciales sesgos, un ejemplo de ello es por ejemplo contar con colectivos de jóvenes desde el inicio, que puedan detectar sesgos potenciales que excluyan a su propio colectivo en procesos similares.

Otro de los riesgos, se vincula a la contratación de algoritmos opacos para la selección de participantes, es lo que se denomina 'captura algorítmica' (Sieber, 2025) y que hace referencia al riesgo de no darnos cuenta de que, como usuarios, estamos siendo manipulados por los resultados. Esto tiene mucho que ver con entender cómo funciona un algoritmo y con realizar las suficientes iteraciones para asegurar que los resultados son todo lo óptimos posibles, añadiendo siempre una capa de contraste humano y sobre todo contar con voces potencialmente afectadas por esta captura. En este caso concreto, mapear actores para participar en un proceso participativo, ese contraste no debe hacerse solamente por personal del Ministerio involucrado en la definición del proceso, sino también por actores con suficiente agencia para poder ser críticos ante ciertas visiones de la administración, que aporten una mirada alternativa y aseguren que los resultados del mapeo son capaces realmente de localizar a voces infrarrepresentadas.

## Importancia de su resolución

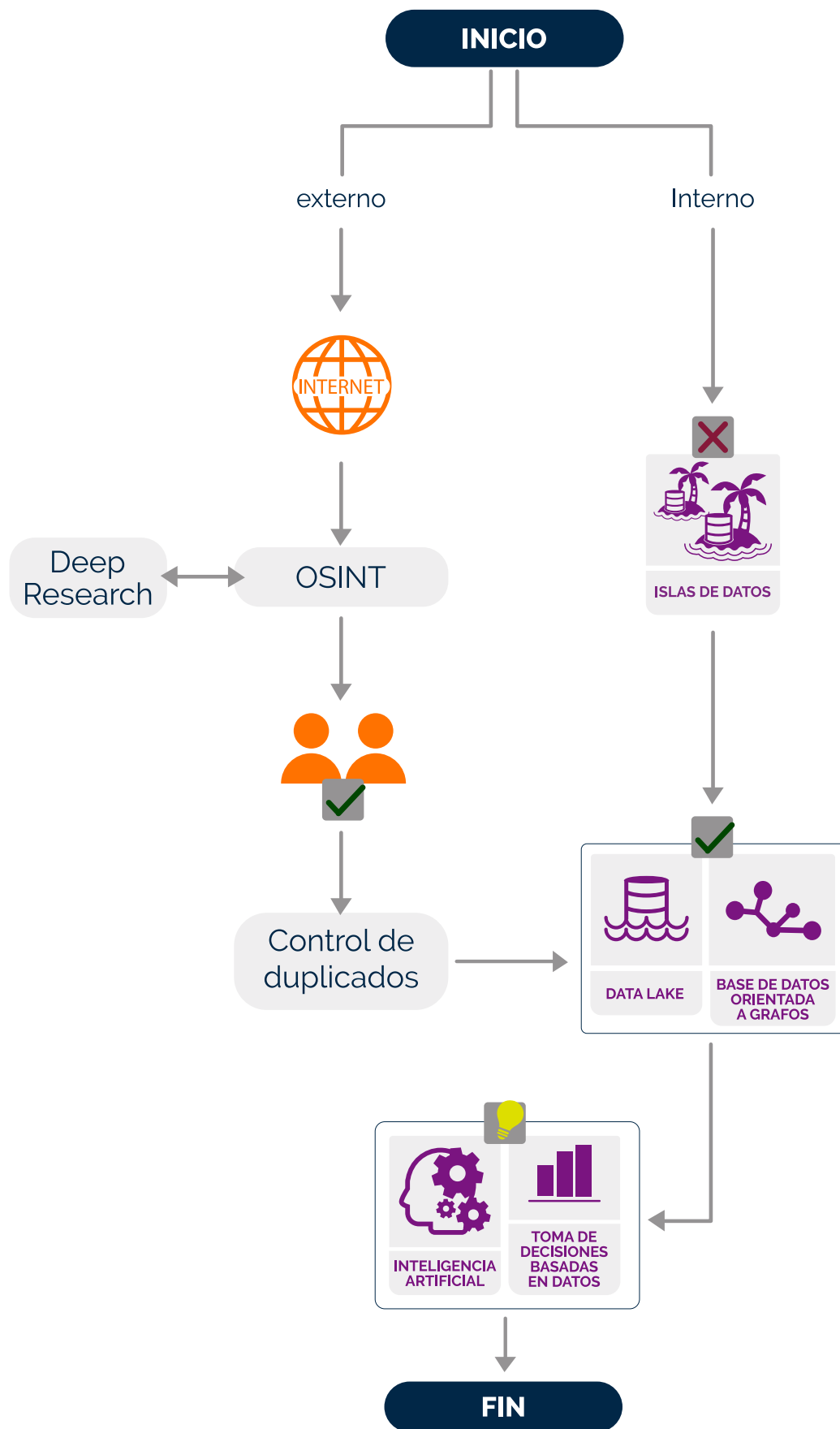
La participación ciudadana cobra sentido si logra reflejar de la manera más fiel e integral posible las voces de todos los actores involucrados en la problemática pública que se pretende abordar y si logra incidir en la propia toma de decisiones. De esa manera es cuando se consigue aprovechar la verdadera inteligencia colectiva que nutre y mejora la eficacia y el impacto de una política pública, un proyecto normativo o cualquier iniciativa que pretenda abordar un problema que afecta a una población o territorio concreto. Nadie mejor que las propias personas afectadas e involucradas para conocer bien el problema y plantear posibles soluciones.

Si una Administración Pública abre un proceso participativo y consigue que participe la mayor cantidad de personas y actores con las más diversas perspectivas, tiene muchas más probabilidades de comprender mejor y de manera más integral el problema a abordar y llegar a soluciones más efectivas y realistas, además de afianzar la confianza y la legitimidad que son la piedra angular de la democracia.



# Solución propuesta

# Diagrama



# Una mejor gestión de datos para democratizar la participación ciudadana

## Introducción

En la mayoría de los casos, al momento de convocar un proceso participativo, el problema no está en la falta de datos sino en el aislamiento de los mismos. Actualmente, las administraciones operan en islas de información que impiden ver el mapa completo de la sociedad. Al romper estos silos mediante una gobernanza robusta, no sólo optimizamos procesos internos, sino que habilitamos la capacidad de identificar y movilizar a sectores de la población que hoy por hoy son invisibles para las instituciones, democratizando así los procesos participativos.

Plantear una solución basada en herramientas tecnológicas para este problema requiere, sin embargo, tener en cuenta ciertas cuestiones previas:

### **a. El valor estratégico de la gobernanza de datos**

El principal obstáculo para una participación ciudadana real es la fragmentación de la información; los datos residen en bases desconectadas que limitan la visión institucional a los participantes habituales. Una gobernanza de datos centralizada y una categorización unificada de recursos digitales permiten cruzar variables socioeconómicas y demográficas de manera ética y segura. Este enfoque transforma el dato estático en un activo dinámico, facilitando que la administración salga de su burbuja tradicional y logre conectar con perfiles ciudadanos y actores infrarrepresentados que, hasta ahora, eran inalcanzables debido a la dispersión de la información.

### **b. Arquitecturas flexibles para un ecosistema conectado**

Para sostener esta visión, es necesario evolucionar de las rígidas estructuras relacionales hacia modelos más ágiles. La implementación de un *Data Lake* —un gran repositorio central que almacena datos en su formato original (desde censos y registros hasta interacciones sociales o encuestas)— permite conservar la riqueza del dato sin las limitaciones de las tablas tradicionales. Complementar esto con bases de datos orientadas a grafos, también llamado diagrama de relaciones es fundamental para el análisis de redes: mientras que una base de datos normal nos dice quién es una persona, el grafo nos revela cómo se conecta con otros actores, barrios y colectivos, si ya ha participado en otros espacios, en resumen nos permitiría conocer su rol y posición en el entramado social. Esta capacidad relacional es el motor que permite diseñar estrategias de participación mucho más granulares y efectivas.

### **c. Hacia una administración inteligente y proactiva**

Una estructura de datos bien gobernada es el combustible necesario para la Inteligencia Artificial. Sin esta base, cualquier solución de IA corre el riesgo de generar alucinaciones o sesgar los resultados a partir de datos incompletos. Al contar con un sistema de información limpio e integrado, la IA puede actuar como un puente, analizando patrones de comportamiento y sugiriendo qué grupos de población podrían verse más afectados por una política pública específica. De este modo, la tecnología no solo facilita la explotación de datos, sino que garantiza que la integración de soluciones futuras sea escalable, transparente y, sobre todo, orientada a una gobernanza más inclusiva y cercana al ciudadano.

### **d. Descubrimiento automatizado de actores no institucionalizados mediante agentes de investigación en fuentes abiertas**

Complementando la estrategia de gobernanza de datos internos descrita anteriormente, se propone incorporar un sistema de descubrimiento activo de organizaciones y personas relevantes que operan fuera de los canales

institucionales y que, por tanto, no figuran en las bases de datos de la Administración. El objetivo es ampliar el universo de perfiles convocables a los procesos participativos, incorporando voces que hoy permanecen invisibles para las instituciones.

### **Agente de investigación en fuentes abiertas (OSINT agent) con *deep research***

Se recomienda el uso de un agente de IA de tipo OSINT (*Open Source Intelligence*) complementado con un agente *deep research*. Mientras que un *deep research agent* está optimizado para sintetizar información y generar informes a partir de consultas abiertas, un agente OSINT está diseñado específicamente para la **búsqueda sistemática, la extracción estructurada de entidades (nombres, organizaciones, relaciones) y la verificación cruzada de fuentes**. Esto lo hace más adecuado para la tarea de descubrimiento de perfiles, donde lo que se necesita no es un informe narrativo, sino datos estructurados y verificables que puedan integrarse en una base de datos.

El agente operaría de forma autónoma sobre fuentes de información públicas: redes sociales (especialmente aquellas con orientación comunitaria o activista), medios de comunicación locales y digitales, registros de asociaciones y colectivos, plataformas de la sociedad civil (plataformas de movilización social tipo Change.org, de *crowdfunding* como Goteo.org, o incluso vecinales), directorios de ONGs y fundaciones, y publicaciones académicas o de divulgación. A partir de consultas temáticas definidas por el equipo técnico (por ejemplo: "colectivos juveniles activos en movilidad sostenible en [municipio]"), el agente rastrearía, identificaría y estructuraría perfiles candidatos con la información disponible: nombre, descripción de actividad, ámbito geográfico, datos de contacto públicos y fuentes de referencia.

## Control de duplicados

Antes de proponer la incorporación de un nuevo perfil, el agente debe ejecutar una fase de deduplicación automática. Esto implica comparar cada candidato descubierto contra los registros existentes en la base de datos mediante técnicas de *fuzzy matching* (comparación aproximada de nombres y denominaciones), normalización de identificadores (NIF/CIF, dominios web, identificadores de redes sociales) y resolución de entidades (*entity resolution*) para detectar si una misma organización aparece bajo nombres distintos o si una persona actúa en representación de una entidad ya registrada. Solo los perfiles que superen este filtro pasarán a la siguiente fase.

## Validación humana (*human-in-the-loop*)

Es imprescindible que ningún perfil descubierto por el agente se incorpore automáticamente a la base de datos. El flujo *human-in-the-loop* debe incluir un paso donde, tras descubrir al actor, el primer contacto sea para informarle (artículo 14 del RGPD) y solicitar su consentimiento explícito para estar en el censo participativo.

Se propone un flujo de validación en tres pasos:

1. El agente **genera un lote** de perfiles candidatos con su nivel de confianza y las fuentes que respaldan cada hallazgo;
2. Un técnico de participación **revisa el lote** a través de un panel de administración donde puede aceptar, rechazar o solicitar más información sobre cada perfil;
3. Los perfiles aprobados se **integran en la base de datos** con una etiqueta de origen ("descubrimiento OSINT – [fecha]") que permite trazar su procedencia y evaluar la calidad del proceso a lo largo del tiempo.

Esta fase de supervisión humana es crítica tanto por razones de calidad de datos como por cumplimiento normativo (RGPD), ya que la decisión de registrar a una persona u organización como potencial participante debe ser siempre una decisión revisada por un ser humano y autorizada por la organización o persona afectada

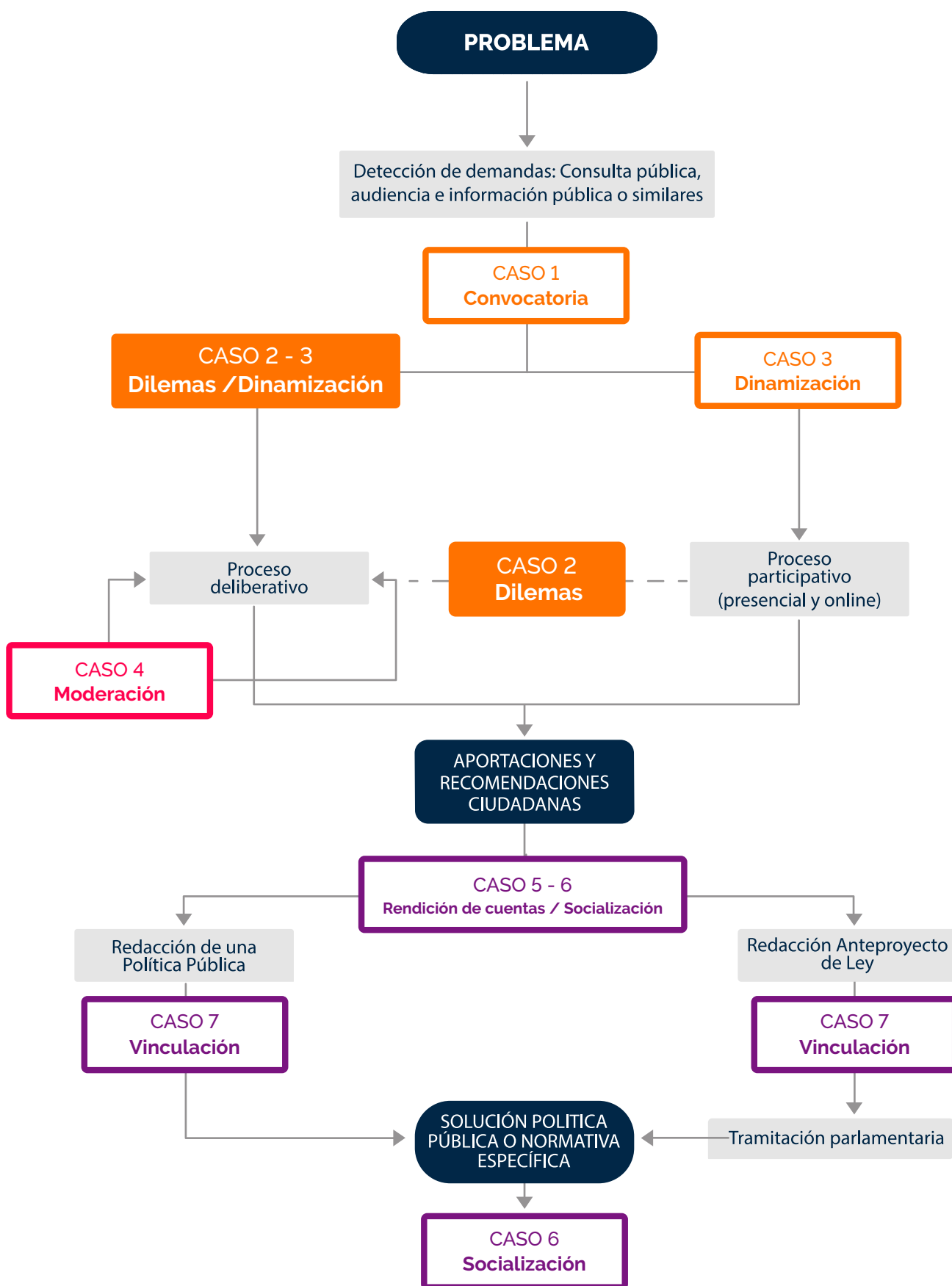
### ¿Por qué un agente OSINT y no simplemente un *deep research* genérico?

Un agente de *deep research* (como los ofrecidos por OpenAI, Perplexity o similares) es útil para responder preguntas complejas y generar informes, pero no está optimizado para la extracción sistemática de entidades ni para la integración estructurada en bases de datos. El agente OSINT, en cambio, se configura con pipelines específicos de recogida, normalización y validación de datos. Además, permite definir reglas de alcance (qué fuentes consultar, qué criterios de relevancia aplicar) y produce salidas directamente compatibles con el potencial *data lake* (u otra solución) descrito en la sección anterior. No obstante, ambos enfoques no son excluyentes: el *deep research* puede utilizarse como complemento para enriquecer los perfiles ya descubiertos (por ejemplo, generando un resumen de la actividad reciente de una organización), mientras que el agente OSINT se encarga de la tarea crítica de descubrimiento e identificación inicial.



Selección  
de dilemas  
deliberativos a  
partir del análisis  
de insumos  
previos

# Diagrama



# 2.

## Selección de dilemas deliberativos a partir del análisis de insumos previos

### Descripción del caso de uso

Para este supuesto existe un caso real del Ministerio de Juventud en el proceso vinculado con la Ley de Juventud y Justicia Intergeneracional: tras haber llevado a cabo una consulta ciudadana durante 2024 (con miles de respuestas recibidas) y varios espacios de diálogo con colectivos jóvenes, se plantea el reto de diseñar un proceso de democracia deliberativa (incluido como compromiso en el V Plan de Gobierno Abierto) y **tienen la necesidad de escoger las preguntas o dilemas de la deliberación** partiendo de las principales conclusiones, ideas, preocupaciones, etc. planteadas en los procesos anteriores. La digestión manual de esta información es demasiado lenta e intensiva en recursos humanos, por lo que se plantea la opción de automatizarlo, siempre que esta automatización sea fiable en cuanto a resultados y a la transparencia del proceso.

### Actores involucrados

En el análisis, revisión y valoración de los resultados de la automatización deberían involucrarse idealmente:

- **Personal técnico del Ministerio involucrado:** es clave que el personal que esté diseñando el proceso participativo, y que además haya estado en fases anteriores de detección de demandas y necesidades,

pueda revisar y valorar los resultados de la automatización, valorando los dilemas detectados por la tecnología y compararlos con los objetivos generales que se espera cumplir e incluso con las reflexiones propias obtenidas por quienes estuvieron en contacto con las personas participantes. Esto es clave ya que es necesario pulir e iterar con la tecnología para asegurar que se capten los matices necesarios a la hora de detectar posibles dilemas.

- **Personal técnico a cargo de diseñar proceso deliberativo:** resulta clave mantener un diálogo con las personas que tengan a su cargo el diseño del proceso deliberativo, ya que serán quienes mejor conozcan los antecedentes e insumos previos, y entiendan el contexto y el marco dentro del cual se quiere desarrollar el proceso. Su rol es doble, por un lado aportando esos insumos y sus propias reflexiones para que la automatización pueda mapear y resaltar aquellos dilemas clave, y por otro, retroalimentando los resultados y valorando su calidad, generando las iteraciones que sean necesarias para mejorar su funcionamiento.
- **Representantes de los principales actores vinculados con el proceso participativo:** se recomienda, sobre todo en los primeros ejercicios que se hagan para la identificación y selección de este tipo de dilemas y contenidos, contar con un pequeño grupo formado por representantes de los distintos grupos que se vayan a involucrar en el proceso deliberativo (por ejemplo sociedad civil, academia, sindicatos, etc), que puedan revisar, contrastar y aportar una mirada crítica a los resultados, esto resulta importante para seguir el principio de co-creación y aprovechar mejor la inteligencia colectiva, asegurando que estos colectivos se apropien desde un inicio del proceso a llevar adelante, pero también, resulta clave para evitar que, si un procesos deliberativo es impulsado desde la administración pública, pueda incurrir en sesgos al evitar poner sobre la mesa ciertos dilemas y debates, o matizarlos demasiado y hacer que pierdan peso y relevancia.

## Causas y consecuencias

La construcción y diseño de procesos deliberativos tiene entre sus principales y mayores retos la definición de los dilemas a abordar durante el proceso. Esta fase es clave, ya que una buena selección de dilemas puede asegurar que quienes participen se sientan escuchados e involucrados, pero además que el tema sobre el cual se encuentran debatiendo es actual y que las decisiones y recomendaciones que se hagan tendrán un impacto.

Por ello resulta tan importante que la selección de dilemas responda en primer lugar a un adecuado mapeo de necesidades y demandas ciudadanas y por otro que exista suficiente compromiso por parte de la administración pública para canalizar y rendir cuenta sobre los resultados del proceso. Si bien la tecnología no puede aportar en este segundo elemento, sí que puede hacerlo en la identificación de esos grandes dilemas a debatir.

En el apartado siguiente se describe en detalle el proceso técnico mediante el cual se realiza esa detección de dilemas, pero resulta interesante tener en cuenta algunas medidas de precaución para evitar que el uso de la tecnología pueda generar ciertas consecuencias que dañen la confianza ciudadana en el proceso, algunos autores como *González de la Garza (2025)* advierten del peligro de influir de manera sutil en la selección de los dilemas por parte de las administraciones o empresas a cargo de la dinamización de procesos participativos. Han acuñado el término *nudge* que vendría a traducirse como 'empujoncito' y hace referencia a que, en un entorno controlado por IA y con bajo nivel de transparencia, existe el riesgo de excesiva influencia por parte de la administración a cargo del proceso, haciendo una especie de preselección de los dilemas, o utilizando la información a su favor para seleccionar ciertas temáticas sobre otras.

Por ello, se recomienda en todo momento contar con diferentes actores que permitan validar los resultados de las automatizaciones, y llegar a un consenso a partir de los resultados otorgados por la tecnología.

Al final la tecnología nos debería permitir ser más eficientes en el procesamiento de datos, pero no en la selección final de los temas, donde resulta clave llegar a un consenso que otorgue a las necesidades de la ciudadanía la relevancia que merecen.

## Importancia de su resolución

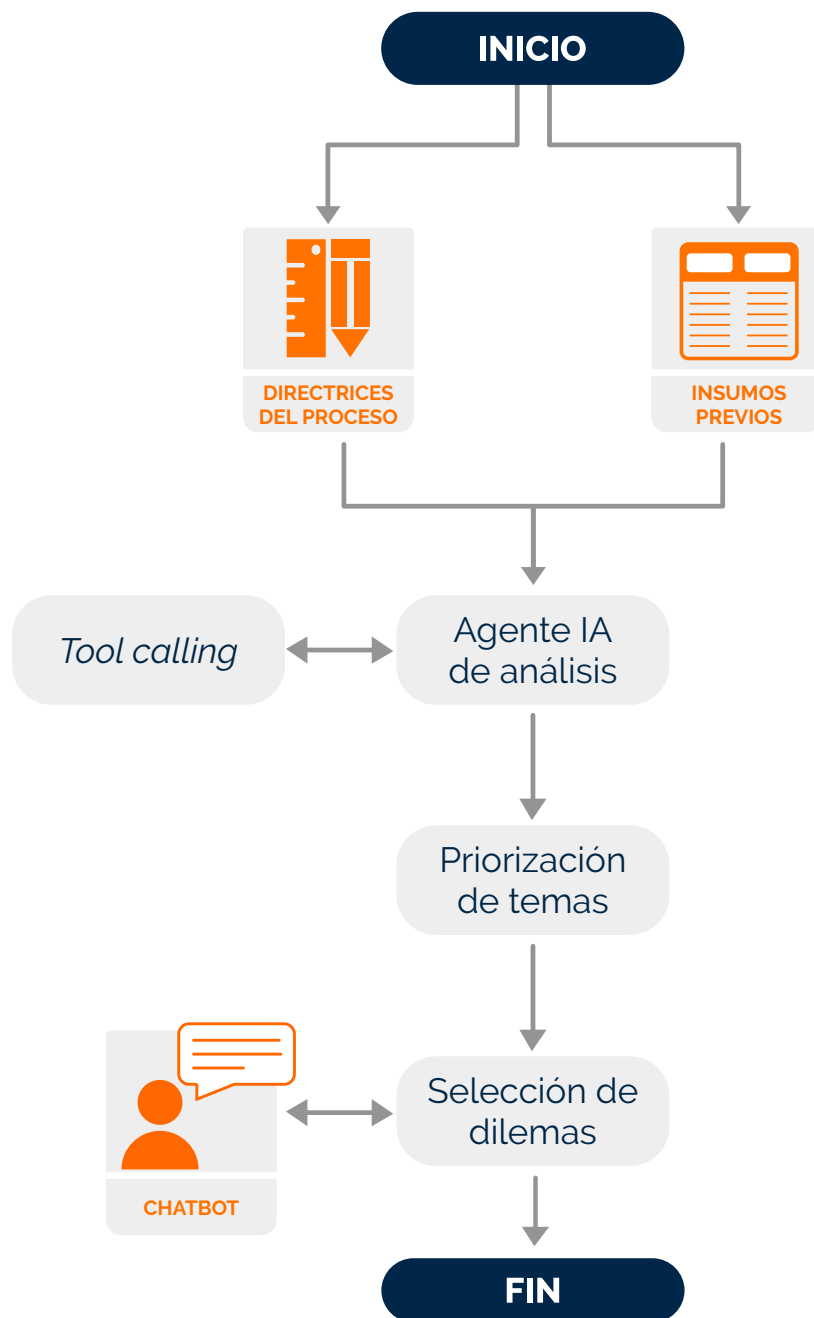
La resolución de este caso de uso permitiría ahorrar tiempo y recursos en el procesamiento, sistematización y análisis de la información recogida de los procesos participativos previos, y a su vez permitiría una definición más asertiva y realista de los dilemas a tratar en el proceso deliberativo, respondiendo a los propios intereses e inquietudes de los públicos objetivo.

Esto aportaría también una mayor coherencia a todo el ciclo del proceso participativo-deliberativo, y redundaría a medio y largo plazo en un abordaje más eficaz y legítimo de los dilemas discutidos y las acciones acordadas para abordarlos.



Solución  
propuesta

# Diagrama



# Agente de análisis de insumos previos con detección de alineamiento para la selección asistida de dilemas en procesos participativos

## Introducción

Se propone un sistema compuesto por un agente de análisis para la extracción de temáticas y dilemas, una fase de priorización y una interfaz conversacional que permiten a los equipos técnicos y políticos explorar, sintetizar y seleccionar los dilemas o preguntas clave para un proceso participativo a partir de los resultados de procesos anteriores (consultas ciudadanas, espacios de diálogo, encuestas, etc.). El agente de análisis, orquestado mediante un *framework* de ingeniería de agentes como LangGraph, procesa los insumos previos utilizando un patrón denominado *tool calling*. En lugar de seguir una secuencia fija de pasos, el agente dispone de un conjunto de herramientas especializadas y decide autónomamente cuáles utilizar en función del contenido que encuentra. Tras el análisis, una fase de priorización reordena los temas identificados equilibrando el interés ciudadano y el impacto social. Finalmente, una interfaz conversacional (*chatbot*) permite al equipo profundizar, reformular y validar los resultados antes de tomar decisiones.

Previo a detallar la solución propuesta, resulta clave comprender el tipo de entradas que tendría el sistema, aquí nos referimos principalmente a los insumos previos con los que se puede contar y a las directrices del proceso deliberativo, requeridas para que el agente cuente con el marco necesario para el análisis posterior.

El agente trabaja con dos tipos de entradas:

- **En primer lugar**, los insumos de procesos participativos previos: resultados de consultas ciudadanas, relatorías de espacios de diálogo, respuestas a encuestas, actas de reuniones y cualquier otro material generado en procesos realizados con anterioridad o, en caso de tratarse de un proceso recurrente o periódico, se puede aportar toda la información relativa a las sesiones previas. Estos documentos constituyen la materia prima sobre la que el agente identifica temas, preocupaciones, propuestas y tensiones.
- **En segundo lugar**, las directrices del proceso: documentos que definen los valores que guiarán el proceso, objetivos, marco normativo y alcance temático (si ya está definido) del proceso deliberativo que se está diseñando. Estas directrices cumplen la función de *guardrails* blandos (*soft guardrails*) para el análisis: a diferencia de los *guardrails* convencionales en sistemas de IA, que actúan como filtros duros bloqueando o eliminando contenido que no cumple ciertos criterios, aquí las directrices no restringen lo que el agente puede analizar ni producir, sino que proporcionan un marco de referencia contra el cual se evalúa y categoriza cada tema emergente. El agente analiza todo el contenido de los insumos sin excepción, pero etiqueta cada hallazgo con su grado de alineamiento respecto a las directrices, como se detalla más adelante.

## Una solución organizada en fases: la construcción del agente

### 1. Fase de extracción: Análisis con detección de alineamiento y señalización de tensiones

El agente de análisis procesa los insumos previos mediante *tool calling*: su caja de herramientas incluye la extracción de temas, detección de propuestas, identificación de consensos y disensos, clasificación de intervenciones

por tipo, y evaluación de alineamiento con directrices. El agente decide qué herramientas utilizar, en qué orden y cuántas veces según lo que encuentra en cada documento. Por ejemplificar: un insumo que contenga un debate con altos niveles de polarización activará herramientas de análisis de disenso que un documento de tipo acta informativa no necesitaría. Cada herramienta produce salidas estructuradas (definidas mediante esquemas como *Pydantic*) que garantizan un formato unificado y procesable por medios informáticos en las fases posteriores.

Un aspecto diferencial de este sistema es que, al tratarse de una fase preliminar de diseño, parte de las discusiones recogidas en los insumos pueden no estar alineadas con los valores u objetivos del proceso deliberativo que se pretende poner en marcha. Aquí es donde opera el mecanismo de *guardrails* blandos descrito anteriormente: el enfoque adoptado es deliberadamente no excluyente. En lugar de filtrar o descartar estos temas (como haría un *guardrail* duro), el sistema los señala mediante un indicador de alineamiento que clasifica cada tema o discusión emergente en relación con las directrices proporcionadas. Los *guardrails* blandos orientan la atención del equipo técnico del proceso sin censurar el análisis.

Esta decisión de diseño responde a un principio importante: en una fase de selección de dilemas, identificar un punto de conflicto o una tensión entre lo que la ciudadanía plantea y los objetivos institucionales puede ser más valioso que ignorarlo.

Un tema señalado como "parcialmente alineado" o "en tensión con las directrices" proporciona al equipo técnico información muy valiosa: puede revelar una preocupación ciudadana legítima que las directrices no han contemplado, una divergencia entre las prioridades institucionales y las de la población, o un área donde el propio marco del proceso necesita revisión. Eliminar automáticamente estos temas empobrecería el análisis y reproduciría sesgos institucionales.

## 2. Fase de priorización: reordenación por interés ciudadano e impacto social

Una vez que el agente de análisis ha identificado los temas, tensiones y propuestas presentes en los insumos, cada uno representado como una salida estructurada con campos tipados (tema, fuentes, indicador de alineamiento, citas de respaldo, metadatos de procedencia), el sistema ejecuta una fase de priorización que reordena estos elementos equilibrando dos dimensiones complementarias. La primera dimensión es el interés ciudadano: cuánta atención, debate o preocupación ha generado un tema entre los participantes de los procesos previos, medido a través de indicadores como la frecuencia de mención, la diversidad de fuentes que lo refieren (incluyendo la variedad de actores), la intensidad del debate (presencia de posiciones contrapuestas) y el número de propuestas concretas asociadas. La segunda dimensión es el impacto social: la relevancia estructural del tema para el bienestar de la población objetivo, evaluada a partir de las directrices del proceso, el marco normativo vigente y, cuando estén disponibles, indicadores socioeconómicos o datos de contexto proporcionados como entrada adicional al sistema.

Esta fase puede implementarse como un segundo agente con herramientas de apoyo o como un paso determinista más dentro del flujo de LangGraph, según la complejidad deseada.

El resultado de esta fase es una lista priorizada de temas candidatos a convertirse en dilemas deliberativos, donde cada tema incluye sus puntuaciones desglosadas en ambas dimensiones, su indicador de alineamiento (de la fase anterior) y una justificación textual de su posición en el ranking. Esta priorización no pretende ser definitiva sino orientativa: su valor reside en hacer explícitos los criterios de ordenación y en visibilizar casos donde las dos dimensiones divergen. Un tema con alto interés ciudadano pero bajo impacto social estimado puede indicar una preocupación social que merece ser abordada aunque no sea prioritaria desde una perspectiva técnica; inversamente, un tema con alto impacto social pero escasa presencia

en los insumos ciudadanos puede señalar un área donde la ciudadanía aún no ha tomado conciencia de un problema relevante. Ambas situaciones son información valiosa para el diseño del proceso deliberativo.

### 3. Fase opcional: Interfaz conversacional para la exploración y selección de dilemas

Tanto el informe del agente de análisis como la lista priorizada no se entregan como documentos cerrados sino como punto de partida para una exploración conversacional. El equipo de diseño puede interactuar con el *chatbot* para profundizar en temas concretos (“¿qué dijeron los jóvenes sobre vivienda en los espacios de diálogo?”), explorar conexiones entre temas (“¿hay relación entre las preocupaciones sobre empleo y las de salud mental?”), evaluar la viabilidad deliberativa de un tema (“¿este asunto tiene posiciones contrapuestas o hay un consenso claro?”), ajustar los pesos del ranking de priorización, o pedir al sistema que formule posibles dilemas a partir de las tensiones detectadas.

El *chatbot* tiene acceso a todos los insumos originales y a los resultados de ambas fases (análisis y priorización), lo que le permite citar fuentes concretas, distinguir entre lo que proviene de cada proceso previo y justificar la posición de cada tema en el ranking.

Este diseño refuerza el carácter de herramienta de apoyo: la selección final de dilemas es una decisión política y técnica que corresponde a las personas responsables del proceso, no al sistema. La IA agiliza la digestión de grandes volúmenes de información y visibiliza patrones que serían difíciles de detectar manualmente, pero **el criterio humano determina qué temas merecen convertirse en preguntas deliberativas, cómo formularlas y cómo integrar las tensiones detectadas en un diseño de proceso inclusivo y productivo.**

## Riesgos y consideraciones de diseño

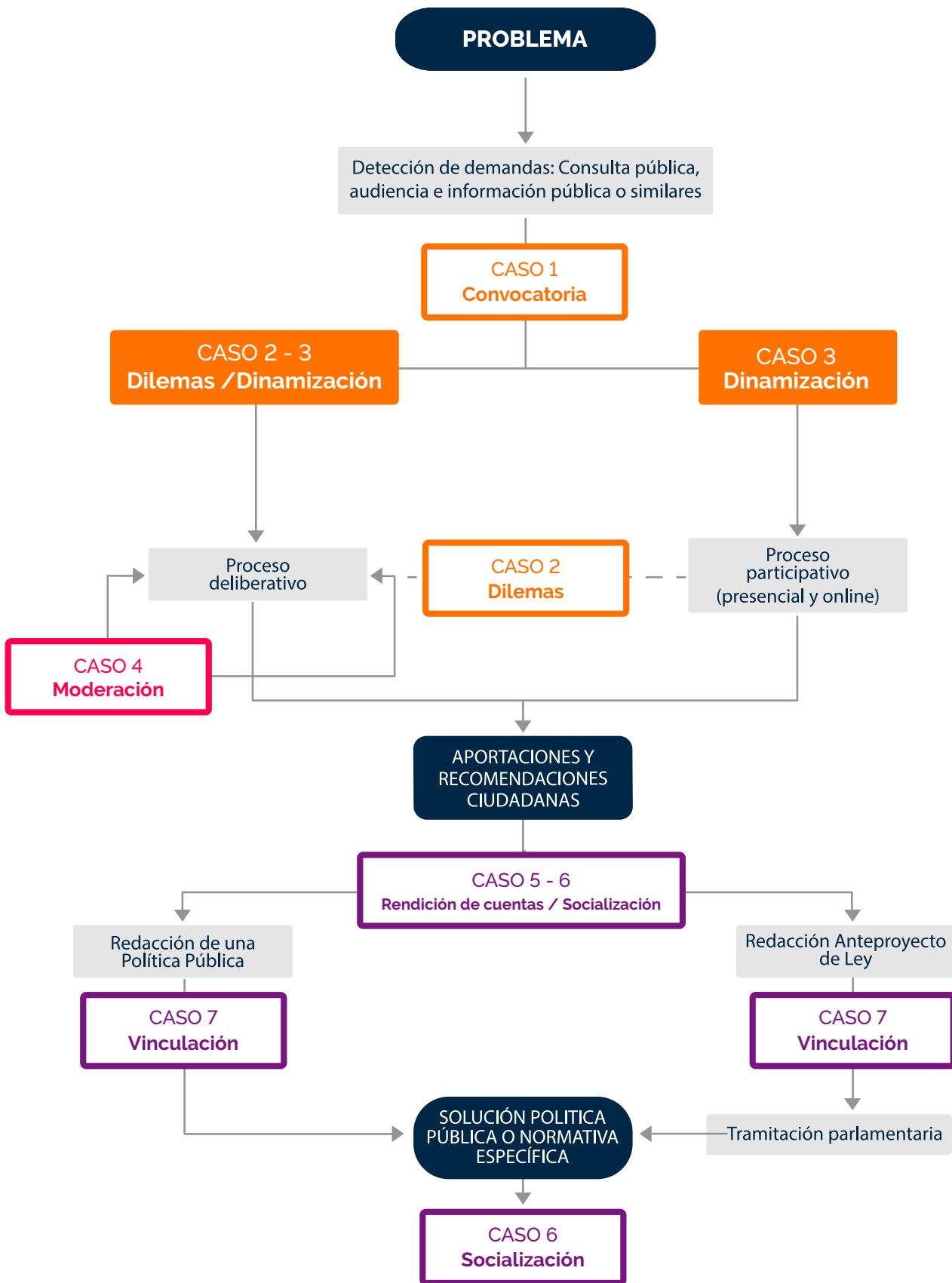
El principal riesgo del sistema de señalización de alineamiento es que las directrices del proceso sean demasiado restrictivas o estén formuladas de forma que el agente clasifique como "no alineados" temas que en realidad son pertinentes pero que simplemente no estaban previstos. Para mitigar este riesgo, se recomienda que las directrices incluyan no sólo los objetivos explícitos del proceso sino también una descripción de su espíritu y alcance, y que el indicador de alineamiento utilice una escala gradual (no una clasificación binaria alineado/no alineado) que distinga entre temas claramente fuera de alcance, temas en tensión con las directrices pero potencialmente valiosos, y temas plenamente alineados.

Adicionalmente, cuando los insumos provienen de procesos independientes (segundo escenario), existe el riesgo de que el agente equipare el peso de fuentes con representatividad muy diferente (una consulta con miles de respuestas frente a un espacio de diálogo con veinte personas). El sistema debe hacer explícita la procedencia y el volumen de cada fuente en sus informes, y el equipo de diseño debe poder filtrar y ponderar las fuentes a través de la interfaz conversacional. La transparencia sobre las fuentes es especialmente crítica cuando el sistema se utiliza para justificar la selección de dilemas ante la ciudadanía, ya que la trazabilidad del proceso es un requisito de legitimidad democrática.

# 3

Generación  
de materiales  
dinamizadores a  
partir de insumos  
previos

# Diagrama



# 3.

## Generación de materiales dinamizadores a partir de insumos previos

### Descripción del caso de uso

La información de los procesos previos se quiere usar específicamente para el **diseño de materiales dinamizadores de los procesos participativos**. En este caso, el ejemplo más clarificador es del desarrollo de una solución basada en IA que convierta los resultados de procesos participativos previos en tarjetas de un juego de mesa (por ejemplo con la estética del juego Dixit) con las que plantear dinámicas participativas.

### Actores involucrados

En el análisis, revisión y valoración de los resultados de la automatización deberían involucrarse:

- **Personal técnico del Ministerio involucrado:** es clave que el personal que esté diseñando el proceso participativo, incluyendo la definición de objetivos, pueda revisar y valorar los resultados de la automatización, valorando el material producido y compararlo con los objetivos generales del proceso. Permitiendo nuevas reflexiones y mejoras de la tecnología.
- **Personal técnico a cargo de dinamizar y/o gestionar el proceso participativo:** resulta clave mantener un diálogo con las personas que vayan a estar a cargo de la dinamización, ya que serán quienes, a partir

de los materiales y recursos elaborados, deberán asegurar el correcto flujo del proceso y el aprovechamiento de dichos materiales. Por otro lado, y en función de los objetivos del proceso, pueden aportar ideas nuevas y mejoras para asegurar que los resultados sean idóneos para el proceso a iniciar.

- **Representantes de los principales actores vinculados con el proceso participativo:** se recomienda, sobre todo en los primeros ejercicios que se hagan para el diseño de este tipo de materiales, contar con un pequeño grupo formado por representantes de los distintos grupos que se vayan a involucrar en el proceso (por ejemplo sociedad civil, academia, sindicatos, etc), que puedan revisar, contrastar y aportar una mirada crítica a los resultados, esto resulta importante para seguir el principio de co-creación y aprovechar mejor la inteligencia colectiva, asegurando que estos colectivos se apropien desde un inicio del proceso a llevar adelante. Por otro lado, contar con representantes de estos colectivos, ayudará también a dar contexto a los resultados, por ejemplo un colectivo de niños y niñas puede aportar información sobre el tipo de materiales con el que se sentirían más cómodos interactuando o que sienten más cercano a su día a día en comparación con lo que puedan decir sobre ello un grupo de adultos.

## Causas y consecuencias

Muchas veces los procesos participativos presentan fallas de diseño metodológico, es decir, no proponen consignas claras ni atractivas de trabajo para que las personas participantes puedan conocer en profundidad el problema a tratar, involucrarse en su reflexión y animarse a expresar su punto de vista, a debatir e incluso a disentir entre ellos. Estos errores radican generalmente en dificultades para presentar y exponer la información previa que ofrezca contexto e introduzca el tema a abordar, y más cuando se trata de sistematizar y recoger un volumen grande de aportaciones recibidas en un momento anterior.

En este sentido, el uso de la tecnología, y en especial de la IA generativa, ofrece ciertas ventajas orientadas a facilitar los procesos de diseño de los procesos participativos, estas ventajas, si se usan de manera adecuada pueden favorecer a la creatividad, mejorar ciertos procesos y aportar innovación que permita una mayor cercanía con la ciudadanía y dinámicas más cercanas. Las principales consecuencias de aplicar estas herramientas se vinculan con la aceleración de ciertos procesos de diseño, incorporando además recursos y aprendizajes de procesos anteriores sin los largos periodos de espera y procesamiento de la información que solían ser una de las principales barreras. Por otro lado, favorece a innovación y a una mayor amplitud de recursos disponibles para asegurar que quienes participan mantengan el interés y el compromiso en el proceso, evitando sesiones que pueden ser tediosas o utilizando metodologías anticuadas.

Sin embargo, resulta clave tener en cuenta dos cuestiones para evitar que el uso de la tecnología afecte de manera negativa a estos procesos, y que pueda generar efectos colaterales no deseados:

- **Garantizar la transparencia en todo el proceso:** comunicar a la ciudadanía de manera clara el uso que se ha hecho de la IA y con qué finalidad es un ejercicio importante para mantener la confianza de quienes participan (OCDE, 2023)
- **Evitar priorizar la eficiencia por encima de la calidad en el proceso:** aunque la IA permita acelerar los procesos de diseño y elaboración de materiales para el proceso participativo, es clave asegurar que existe personal técnico a cargo de revisar todos los resultados obtenidos mediante automatizaciones, su correspondiente testeo para verificar si son realmente útiles de acuerdo a los objetivos que se espera obtener del proceso participativo y testear si realmente favorecen las dinámicas entre quienes participan o si pueden generar otros retrasos inesperados (Aranzazulab, 2025). Es clave normalizar las iteraciones y 'curar' los resultados antes de ponerlos en marcha en vivo y en directo, recordando que quienes participan tienen unos tiempos limitados y ofrecen sus aportaciones de manera voluntaria, y por tanto deben protegerse sus intereses y favorecer entornos seguros y procesos de calidad.

## Importancia de su resolución

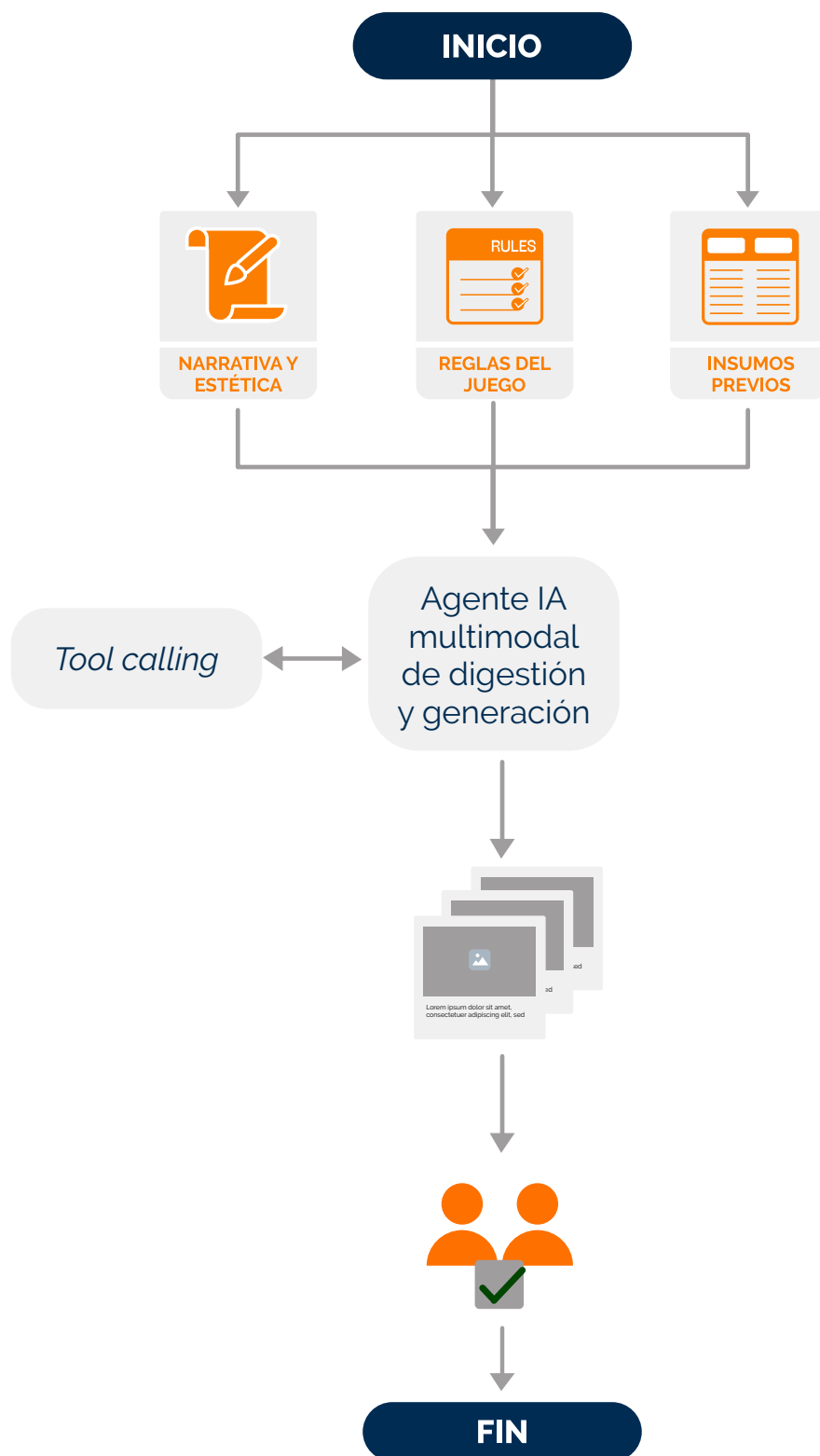
Una tecnología que permita la elaboración de materiales didácticos basados en información recogida de un proceso participativo previo, y para aplicar en la dinamización de espacios de co-creación o participación ciudadana, mejoraría considerablemente la eficiencia y eficacia de ese espacio y el arribo a conclusiones de trabajo más concretas y claras. Permitiría presentar e introducir las conclusiones del proceso previo de una manera atractiva, lúdica y amena para los nuevos participantes, ahorrando tiempo, sosteniendo su atención y logrando una participación más activa.

Esto redundaría en una mayor coherencia en todo el ciclo participativo, alentaría a los participantes a volver a involucrarse en futuras instancias y a multiplicar la participación por el efecto "contagio" o "boca a boca" entre sus contactos. A medio y largo plazo esto permitiría consolidar el espacio y por ende, mejorar las políticas públicas relacionadas y la confianza ciudadana en las instituciones y en la política.



Solución  
propuesta

# Diagrama



# Agente multimodal para la generación de materiales de dinamización

## Introducción

Se propone un sistema basado en agentes de IA que transforme los resultados de procesos participativos previos en materiales listos para usar en dinámicas participativas gamificadas (tarjetas, narrativas o historias, reglas adaptadas, elementos visuales). El sistema recibe tres tipos de entradas:

1. **Los insumos del proceso o procesos previos** (resultados de consultas, relatorías, conclusiones),
2. **las reglas del juego o dinámica elegida** (puede ser por ejemplo un juego basado en cartas, con una dinámica al estilo del juego Dixit con debates estructurados u otra basada en dinámicas de estrategia y prospectiva al estilo del Ciudadelas por ejemplo)
3. y una **dirección artística o lore que defina la estética y el universo narrativo de los materiales** (por ejemplo un escenario utópico, en un pasado o un futuro lejano o en un universo alternativo donde hay magia).

A partir de estos insumos, el agente genera el conjunto completo de materiales necesarios para llevar a cabo la dinámica. El sistema requiere un modelo de lenguaje multimodal con capacidad de generación de imágenes y textos, ya que la producción de tarjetas incluye tanto el diseño visual como el contenido textual.

Esta propuesta se inspira en proyectos como [Analogía](#) (MIT Media Lab), que utiliza un *pipeline* de IA generativa para transformar historias, reflexiones

y símbolos, recogidos de una comunidad en mazos de cartas personalizadas con imágenes surrealistas generadas por IA e inspirado en juegos de narrativa basados en imágenes como Dixit o el Mysterium. Analogía demuestra cómo la generación automática de materiales de juego a partir de insumos colectivos puede facilitar el diálogo empático y la reflexión compartida en dinámicas participativas. La propuesta que se describe a continuación, generaliza y amplía este enfoque para su aplicación en procesos de participación ciudadana institucional.

El sistema requiere, como condición previa, contar con una serie de entradas que le permitan enmarcar el contexto para poder realizar el proceso de análisis y producción de resultados. Así, el agente trabaja con tres tipos de entradas:

1. **En primer lugar**, los insumos participativos: resultados de consultas ciudadanas, relatorías de espacios de diálogo, conclusiones de procesos previos y cualquier otro material que recoja las voces y preocupaciones de la ciudadanía. Estos documentos proporcionan la base temática y conceptual que se transformará en contenido para las tarjetas y materiales del juego.
2. **En segundo lugar**, las reglas del juego: un documento que describe la mecánica de la dinámica participativa gamificada (tipos de tarjetas, número de cartas por categoría, turnos, condiciones de victoria o de avance, roles de los participantes, estructura de las rondas). Estas reglas actúan como restricciones para el agente generativo, que debe producir materiales compatibles con la mecánica definida.
3. **En tercer lugar**, la dirección artística y lore: una guía que establece el universo narrativo (ambientación, tono, vocabulario), la estética visual (paleta de colores, estilo de ilustración, referencias gráficas), y las narrativas marco que dan coherencia al juego como experiencia.

## Estructura de la solución propuesta

### a. Arquitectura del agente

El agente se orquesta mediante un *framework* como LangGraph y utiliza un patrón de *tool calling* para adaptar su comportamiento a la mecánica concreta de cada juego. Dispone de herramientas como: extracción de temas y tensiones de los insumos, síntesis de situaciones o dilemas para las tarjetas del juego, generación de textos narrativos coherentes con el *lore*, generación de imágenes a partir de *prompts* derivados de la dirección artística, validación de coherencia entre el contenido generado y las reglas del juego, y equilibrio con la dificultad o la distribución temática entre tarjetas. El agente decide qué herramientas invocar y en qué orden según la mecánica: un juego tipo Dixit basado en imágenes evocadoras requerirá un uso intensivo de la herramienta de generación de imágenes, mientras que un juego de debate estructurado como The Metagame se centrará más en la generación de textos provocadores y situaciones de dilema. Cada material generado utiliza salidas estructuradas (esquemas *Pydantic*) que garantizan que las tarjetas (textos e imágenes) se producen en formatos unificados y directamente utilizables.

### b. Materiales generados

El sistema produce el conjunto completo de materiales necesarios para ejecutar la dinámica participativa. Esto incluye: las imágenes de las tarjetas con su diseño visual (ilustraciones, fondos, composición), generadas mediante un modelo con capacidad de generación de imágenes; los textos de las tarjetas cuando la mecánica lo requiera (preguntas, situaciones, dilemas, datos, provocaciones para el debate); los materiales textuales de apoyo al juego, como las instrucciones para participantes, la introducción narrativa que contextualiza la dinámica, los elementos de *lore* que enriquecen la experiencia (descripciones de escenarios, trasfondos de personajes, contexto narrativo de cada ronda), y las guías para las personas dinamizadoras con indicaciones sobre cómo facilitar cada fase del juego. Adicionalmente,

cuando la mecánica del juego incluye componentes como tableros, fichas, marcadores de progreso o elementos de escenificación, el sistema puede generar los diseños correspondientes o proporcionar especificaciones detalladas para su producción.

### **C. Generación asistida de reglas de juego y dirección artística**

El sistema podría contemplar un paso adicional que cubre dos escenarios frecuentes en la práctica. El primero es cuando no existen reglas de juego ni dirección artística predefinidas: el equipo de diseño dispone únicamente de los insumos participativos y del objetivo general de la dinámica. En este caso, el agente puede generar una propuesta completa de mecánica de juego y dirección artística a partir de los propios insumos, analizando la naturaleza de los temas y respondiendo a una serie de preguntas guía para orientar el tipo de resultados (¿qué tipo de resultado se busca con el proceso?, ¿qué tipo de colectivos participarán?, ¿se admite un debate polarizado?, ¿requieren exploración creativa?, ¿hay componentes emocionales?), el perfil de la población participante y los objetivos del proceso para recomendar una mecánica adecuada y una estética coherente con el público y las temáticas.

El segundo escenario es cuando las reglas y la dirección artística existen pero no están suficientemente alineadas con los objetivos del proceso participativo. Por ejemplo, un juego de cartas diseñado originalmente para un contexto diferente que se quiere adaptar a un proceso de participación juvenil: la mecánica puede ser buena pero el tono, la complejidad del lenguaje o la estética pueden no encajar con el público.

En este caso, el agente analiza las reglas y la dirección artística proporcionadas, las contrasta con los objetivos del proceso y el perfil de participantes, y genera recomendaciones concretas de adaptación: simplificar ciertas reglas, ajustar el tono narrativo, modificar la paleta visual, añadir o eliminar categorías de tarjetas, o reformular las condiciones de las dinámicas para que favorezcan mejor la deliberación.

#### **d. Validación de alineamiento y coherencia**

Un aspecto crítico del sistema es garantizar que los materiales generados sean fieles a los insumos originales y coherentes con las reglas y la dirección artística. El agente incluye una herramienta de validación que utiliza un patrón *LLM-as-a-judge* que verifica: que el contenido de las tarjetas refleje temas, preocupaciones o propuestas realmente presentes en los insumos (evitando que la IA "alucine" contenido sin respaldo o que resulte poco útil para el proceso); que el número, tipo y formato de las tarjetas sea compatible con las reglas del juego; que las imágenes generadas sigan la dirección artística establecida (estilo, paleta, tono); que la distribución temática de los materiales sea equilibrada y representativa de la diversidad de voces en los insumos; y que el lenguaje y la complejidad sean adecuados al perfil de la población participante.

Esta validación automática no sustituye la revisión humana, que sigue siendo imprescindible antes de utilizar los materiales en una dinámica real, pero reduce significativamente el esfuerzo de curación necesario.

### **Riesgos y consideraciones de diseño**

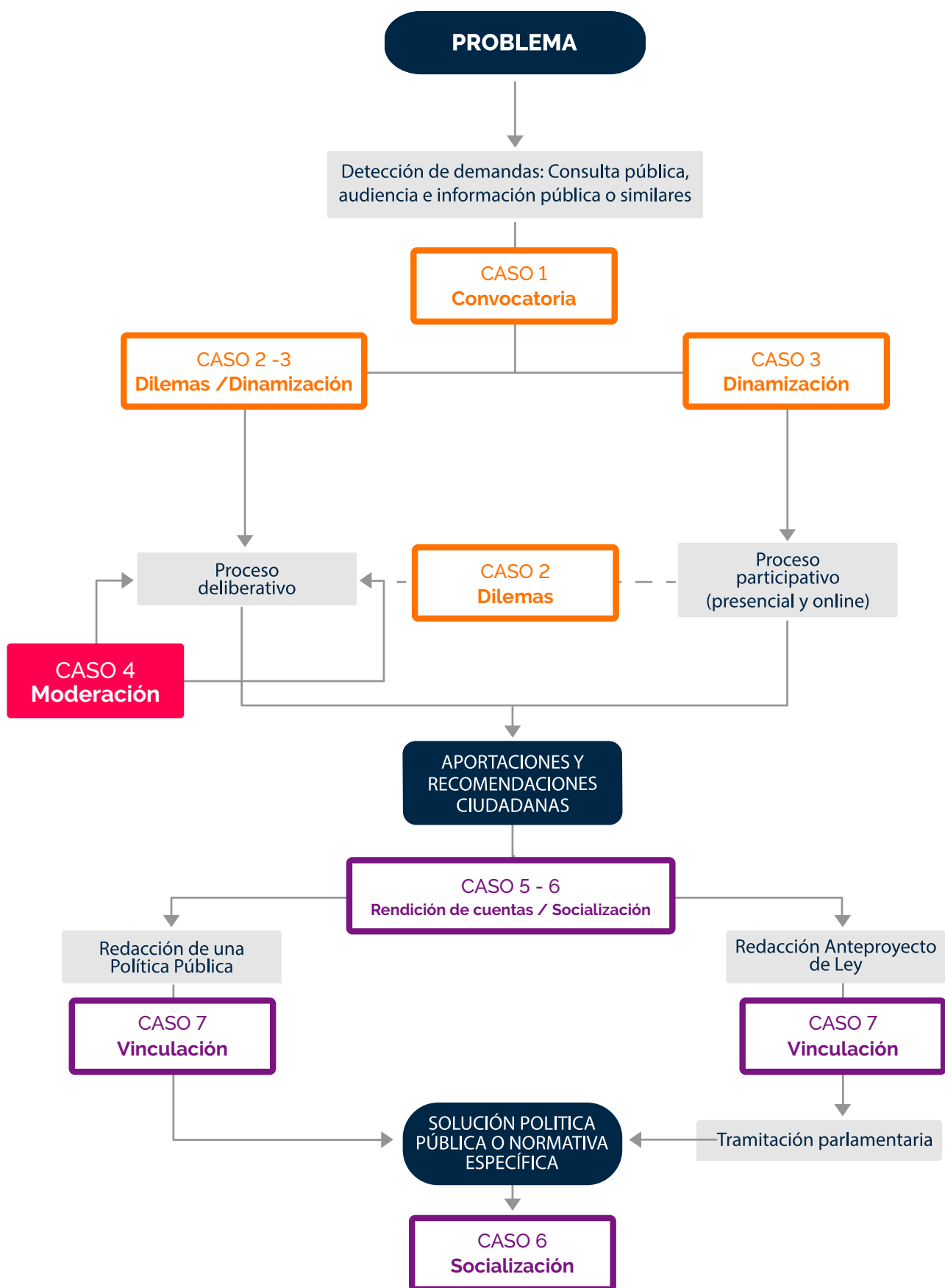
La generación de imágenes mediante IA introduce consideraciones específicas. Los modelos de generación de imágenes actuales pueden producir sesgos visuales (representaciones estereotipadas de género, edad, etnia o diversidad funcional) que resultan especialmente problemáticos en materiales destinados a procesos participativos inclusivos. Se recomienda una revisión humana cuidadosa de todas las imágenes generadas, con especial atención a la representación de la diversidad, y la inclusión de directivas explícitas de diversidad e inclusión en la dirección artística proporcionada al sistema.

Adicionalmente, la fidelidad a los insumos originales es un riesgo central: la IA generativa tiende a “embellecer” o reinterpretar el contenido, lo que podría distorsionar el sentido de las aportaciones ciudadanas, especialmente en estas mecánicas gamificadas que pueden incluir componentes de ficción. El sistema debe mantener trazabilidad entre cada material generado y los objetivos e insumos que lo inspiran, permitiendo al equipo de diseño verificar que las tarjetas representan fielmente las preocupaciones y propuestas de la ciudadanía. Por último, como se señala en secciones anteriores, es imprescindible realizar pruebas piloto con los materiales generados antes de su uso en dinámicas reales, especialmente cuando se trabaja con población joven o infancias.

# 4

Análisis  
automatizado para  
la moderación y  
la detección de  
consensos

# Diagrama



# 4.

## Análisis automatizado para la moderación y la detección de consensos

### Descripción del caso de uso

La moderación de procesos participativos deliberativos puede suponer un reto, especialmente cuando el número de participantes es elevado o cuando el proceso incorpora alguna característica innovadora (como el hecho de ser un proceso híbrido presencial - online, por ejemplo). Ante este reto, se plantea la posibilidad de **automatizar el análisis del contenido de las sesiones de deliberación y generar insumos para una mejor moderación**. De esta manera la persona moderadora contaría rápidamente con información útil para orientar su labor, como la detección de temas prioritarios, la ausencia en las discusiones de algún asunto que se considera estratégico, la deriva hacia posiciones de consenso o disenso, entre otras.

En este caso de uso la tecnología puede facilitar el encuentro genuino entre diferencias, y no simplemente exponerlas, creando arquitecturas de participación que promuevan la escucha activa y la construcción colectiva por encima del debate confrontacional.

## Actores involucrados

En el análisis, revisión y valoración de los resultados de la automatización deberían involucrarse:

- **Personal técnico del Ministerio involucrado:** es clave que el personal que haya estado involucrado en el diseño y la definición de objetivos del proceso participativo y de deliberación, pueda revisar y valorar los resultados de la automatización, valorando los elementos de consenso y disenso detectados y compararlos con los objetivos generales del proceso. Permitiendo nuevas reflexiones y mejoras de la tecnología.
- **Personal técnico a cargo de dinamizar y/o gestionar el proceso participativo:** resulta clave mantener un diálogo con las personas que hayan estado involucradas directamente en el proceso participativo, y que hayan mantenido diálogo o contacto con las personas participantes, ya que pueden comparar notas y reflexiones propias vinculadas al proceso con los resultados obtenidos con el uso de la IA.
- Se considera interesante además, en los casos de uso donde la tecnología se aplique para detectar elementos de consenso y disenso, la participación de **representantes de los principales actores vinculados con el proceso participativo**, sobre todo en los primeros ejercicios que se hagan para detectar elementos de consenso y disenso, contar con un pequeño grupo formado por representantes de los distintos grupos que se hayan visto involucrados en el proceso (por ejemplo sociedad civil, academia, sindicatos, etc), que puedan revisar, contrastar y aportar una mirada crítica a los resultados, sobre todo en contraste con sus propias notas o reflexiones, con el objetivo de mejorar la calidad de los resultados a través de diversas iteraciones.

## Causas y consecuencias

Uno de los principales retos que enfrentan los procesos participativos, y en especial los que utilizan metodologías de deliberación, es el tiempo de ejecución y las excesivas cargas para el personal a cargo de su dinamización. Cada fase del proceso tiene sus propios retos, y la dinamización no está exenta de ellos, ya que se requiere de cierta agilidad para asegurar que el proceso fluya adecuadamente, a la vez que sin perder información clave para asegurar que las personas participantes se sientan cómodas y que continúen enriqueciendo el debate mientras no pierden su nivel de compromiso con todo el proceso. En esta fase la carga de trabajo puede resultar excesiva e incluso generar demoras en el proceso, especialmente en aquellos donde el número de participantes es muy alto, como puede suceder en formatos online o híbridos.

Y al mismo tiempo, el poder contar con un mayor número de participantes, brinda legitimidad a los procesos de participación, permitiendo que más personas puedan involucrarse, favoreciendo a la representatividad y a la multiplicidad de voces dentro del proceso.

Por ello, mejorar la capacidad de analizar las aportaciones de los participantes, y de obtener resultados que no pierdan la riqueza y la inteligencia colectiva generada en esos espacios de confianza resulta clave, pero también requiere de una serie de salvaguardas para evitar perder reflexiones e ideas valiosas en el camino.

La tecnología ofrece por un lado el potencial de analizar un gran volumen de datos, identificando entre todas las voces y reflexiones de quienes participan ciertos patrones, agrupando ideas y opiniones similares, detectando incluso aquellas declaraciones que logran mayor apoyo por parte de grupos con mucha divergencia de opiniones (*Campagnucci, et al., 2025*), al mismo tiempo, la velocidad de procesamiento que permite la tecnología, facilita que este análisis de posibles 'puntos de consenso' o incluso aquellos que están generando más disenso, puedan ser detectados mientras el debate

ocurre, sin tener que esperar días a que el proceso termine, lo cual puede incluso cambiar el tono y el recorrido del debate en tiempo real (*Landemore, 2023*)

Sin embargo, la tecnología debe utilizarse siempre con precaución, ya que de lo contrario podría generar efectos colaterales no deseados en procesos participativos donde el objetivo es mejorar la relación de confianza entre la ciudadanía y la administración. En este sentido, la bibliografía advierte sobre ciertos elementos a tener en cuenta de manera preventiva:

- **La pérdida de los 'matices' y la riqueza lingüística:** se trata de una duda que tiene lógica, en el proceso de automatización, y con el objetivo de generar declaraciones que fomenten el mayor consenso posible, existe el riesgo de perder esos pequeños matices que enriquecen una postura sobre otra, o esa creatividad lingüística que permite que ciertas ideas calen más entre ciertas personas o entre grupos con demasiada divergencia (*Campagnucci, et al., 2025*). Aquí la principal recomendación es contar con un proceso de 'revisión' de los resultados por parte de los dinamizadores del proceso, que pueda aportar esos matices en función de la información que se haya recogido por su parte de manera manual.
- **Excesiva priorización de la eficiencia:** los procesos democráticos, para hacerse bien, llevan tiempo, y este es uno de los principales retos de nuestra era, se exigen unos niveles de eficiencia y de velocidad que muchas veces no cuadran con las necesidades de los procesos participativos, ni mucho menos con el cuidado que requiere recuperar la confianza y la conexión con la ciudadanía (*Aranzazulab, 2025*). Sin embargo, también es clave ser realistas y reconocer la necesidad de efficientar ciertos procesos que en la actualidad llevan demasiado tiempo, por ello se recomienda la hibridación de soluciones mezclando la automatización con el cuidado humano.

- **Dudas sobre la transparencia:** una cuestión clave a tener en cuenta es asegurar la absoluta transparencia sobre el uso de soluciones basadas en IA, explicando el para qué, el cuándo y el cómo se utilizan, ya que de lo contrario, su uso podría generar rechazo y dudas de parte de los participantes del proceso (OCDE, 2023).

Teniendo en cuenta estos elementos, y con el objetivo de promover que la aplicación de la tecnología aporte nuevas miradas y no afecte los resultados de los procesos de participación, se recomienda un enfoque híbrido, mediante el cual los resultados de la automatización sean siempre contrastados por parte de los dinamizadores de los procesos, e incluso contrastados por parte de participantes, para aportar mayor riqueza al debate.

## Importancia de su resolución

El análisis automatizado y simultáneo de las aportaciones generadas en un proceso deliberativo que facilite la moderación, es decir, que permita guiar y ayudar a los participantes a dar sentido a la conversación, identificar las tensiones clave que surjan y profundizar en su comprensión mutua y de los puntos de vista alternativos, es fundamental para recoger con la mayor fidelidad posible la diversidad de perspectivas y arribar a acuerdos constructivos que las contemplen.

Al mismo tiempo, evitar la polarización o radicalización del debate promueve un intercambio más rico, fructífero y amable, favoreciendo que el proceso fluya adecuadamente, que no se pierda información relevante en el camino, y por ende, contribuye a que las personas participantes se sientan cómodas y que continúen enriqueciendo el debate y sostengan su nivel de compromiso.

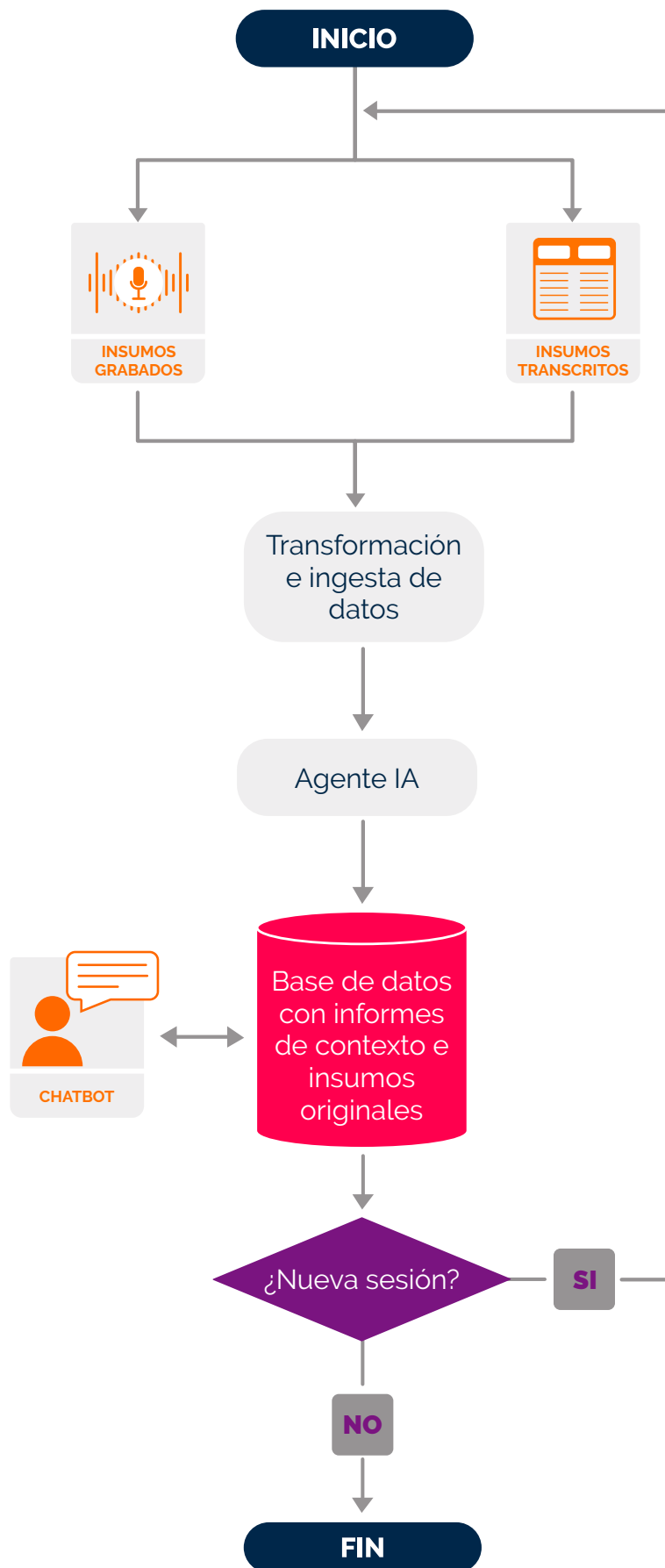
Además, lograr que ese análisis lo haga una tecnología reduce significativamente las cargas de trabajo, las demoras y los tiempos del proceso participativo, permite un número mucho mayor de participantes y que el foco del proceso se concentre en lo importante: **participar, debatir, deliberar y llegar a acuerdos colectivos.**

Todo esto redundará en una mayor diversidad y multiplicidad de voces dentro del proceso, mayor grado de representatividad y legitimidad de la instancia deliberativa, y en última instancia, mejor adaptabilidad y nivel de eficacia de los proyectos o iniciativas resultantes.



Solución  
propuesta

# Diagrama



# Agente iterativo de análisis deliberativo con comprensión progresiva de contexto para apoyo a la moderación

## Introducción

Se propone un sistema basado en un agente de IA que acompañe al mediador a lo largo de todo un proceso deliberativo, analizando de forma iterativa las sesiones de trabajo y ofreciendo información procesable en tiempo casi real. El sistema está diseñado para procesos participativos que se desarrollan a lo largo de múltiples jornadas (días, semanas o meses) con eventos periódicos (presenciales, digitales o híbridos) en los que las personas participantes debaten y formulan argumentos o propuestas sobre un determinado tema.

El sistema se compone de cuatro bloques:

4. un módulo de ingesta que transforma transcripciones textuales o grabaciones de audio en texto procesable;
5. un agente de análisis deliberativo orquestado con LangGraph que ejecuta el ciclo iterativo y genera las dos salidas del sistema;
6. un almacén persistente que conserva los informes de contexto acumulado y las transcripciones originales como respaldo auditable;
7. y una interfaz conversacional tipo *chatbot* donde el mediador recibe los informes y puede interactuar con el agente para explorar los resultados.

Cada uno de estos componentes se describe en detalle en los siguientes apartados, donde se aterriza en la estructura de la solución propuesta.

## Un ciclo iterativo con doble salida

Cada evento del proceso participativo genera una entrada al sistema, ya sea una transcripción textual (relatoría manual o generada por IA) o una grabación de audio que el sistema transforma en texto mediante herramientas de *speech-to-text*. Opcionalmente, esta transcripción puede enriquecerse con un proceso de diarización de hablantes, que permite la identificación automática de quién interviene en cada momento, lo que añade una capa extra y de gran utilidad para que el mediador pueda rastrear la evolución de posiciones individuales a lo largo del proceso, aunque no es imprescindible para el análisis de consenso y disenso a nivel temático.

**α. El núcleo del sistema es un agente de IA orquestado mediante un *framework* de ingeniería de agentes (como *LangGraph*) que, tras cada evento, recibe dos entradas: la transcripción de la sesión recién celebrada y el informe de contexto acumulado generado tras el evento anterior (o vacío si se trata de la primera sesión). A partir de ambas entradas, el agente produce dos salidas diferenciadas:**

La primera salida es un informe para el mediador ("informe humano"), diseñado para ser leído y utilizado directamente por la persona que coordina el proceso. El informe combina dos perspectivas complementarias: el análisis del estado de la deliberación y el apoyo a la preparación de la siguiente sesión.

En cuanto al estado de la deliberación, incluye: mapeo de posiciones de consenso y de disenso con indicación de su intensidad y evolución respecto a sesiones anteriores, detección de temas prioritarios surgidos en la sesión, identificación de asuntos estratégicos ausentes en la deliberación, y propuestas concretas formuladas por los participantes.

En cuanto al apoyo a la moderación, incorpora: análisis de la dinámica de participación (equilibrio entre voces, grupos o perfiles que no han tenido oportunidad de intervenir sobre ciertos temas), detección de bloqueos argumentativos o circularidad en el debate (temas que se repiten sin avanzar

hacia nuevas posiciones), señales de fatiga o desconexión temática (pérdida de profundidad o de diversidad argumental a lo largo de la sesión), y sugerencias de enfoque para la siguiente sesión: temas que podrían abrirse, preguntas que lanzar al grupo, o reformulaciones de propuestas que podrían desbloquear puntos enquistados.

Este informe se presenta junto a una interfaz de *chatbot* que permite al mediador explorar los resultados de forma conversacional: preguntar por las diferencias respecto a sesiones anteriores, profundizar en un tema concreto, solicitar sugerencias de abordaje para puntos de disenso, o pedir reformulaciones de propuestas que puedan facilitar el acuerdo.

La segunda salida es un informe de contexto acumulado (“informe para el LLM”), que no está destinado a lectura humana sino a servir como memoria estructurada del proceso para el propio agente. Este informe sintetiza todo el conocimiento acumulado hasta el momento (posiciones, acuerdos, desacuerdos, evoluciones temáticas) de forma comprimida y optimizada para ser procesada por el modelo de lenguaje en la siguiente iteración. Su función es crítica: permite que el agente mantenga una visión completa del proceso sin necesidad de reprocesar todas las transcripciones anteriores.

#### **b. Compresión progresiva de contexto: por qué este patrón y no la acumulación directa**

El enfoque alternativo más obvio sería alimentar al modelo con todas las transcripciones anteriores concatenadas. Sin embargo, esta estrategia presenta un problema conocido como degradación en forma de U de la atención: los modelos de lenguaje actuales tienden a prestar más atención al contenido situado al principio y al final de su ventana de contexto, perdiendo capacidad de procesamiento sobre el material intermedio. En un proceso participativo de muchas jornadas, esto significaría que el agente daría un peso desproporcionado a las primeras y últimas sesiones, ignorando progresivamente las intermedias (precisamente donde suele producirse la evolución de posiciones y la construcción de consensos). Además, la acumulación directa multiplica el consumo de tokens con cada iteración, encareciendo exponencialmente el coste del sistema.

El patrón de compresión progresiva resuelve ambos problemas: el informe para el LLM actúa como una memoria comprimida que conserva la sustancia (posiciones, acuerdos, disensos, evoluciones) pero elimina la redundancia, el ruido y los detalles que ya no aportan valor analítico. De este modo, el agente trabaja siempre con un contexto de tamaño controlado, manteniendo una visión equilibrada de todo el proceso independientemente de cuántas sesiones hayan transcurrido.

### **c. Interfaz conversacional para el mediador**

El informe para el mediador no se entrega como un documento estático, sino que se presenta junto a una interfaz de chatbot que transforma la experiencia de un consumo pasivo de información a una exploración activa. El mediador puede conversar con el agente para profundizar en cualquier aspecto del análisis: preguntar qué ha cambiado respecto a la sesión anterior, explorar un tema concreto con mayor granularidad, solicitar sugerencias de abordaje para puntos de disenso, o pedir reformulaciones de propuestas que puedan facilitar el acuerdo entre posiciones enfrentadas.

Un aspecto clave de esta interfaz es que el chatbot tiene acceso al informe de contexto acumulado (el "informe para el LLM"), lo que significa que puede responder preguntas no solo sobre la sesión más reciente sino sobre todo el historial del proceso. Esto permite consultas longitudinales que serían imposibles con un informe aislado: "¿ha habido algún cambio de posición sobre el tema X desde la tercera sesión?", "¿qué temas se mencionaron al inicio del proceso y han desaparecido de la discusión?" o "¿cuáles son las propuestas que más apoyo transversal han recibido a lo largo de todas las sesiones?".

Este diseño refuerza el principio de que el sistema es una herramienta de apoyo, no un sustituto del criterio humano. Al permitir que el mediador formule preguntas activamente, se evita el riesgo de que conclusiones automáticas se tomen como verdades absolutas. La persona mediadora es quien conoce la dinámica humana de la sala y quien decide cómo actuar; el sistema le proporciona una visión analítica enriquecida sobre la que apoyar sus decisiones.

#### **d. Análisis inter-sesión frente a análisis intra-sesión en tiempo real**

El sistema propuesto opera exclusivamente en modo inter-sesión: el análisis se ejecuta una vez finalizada cada jornada de trabajo, y sus resultados están disponibles para que el mediador los consulte antes de la siguiente sesión. Se ha descartado deliberadamente un modo de análisis en tiempo real (intra-sesión) por varias razones:

- **En primer lugar**, el análisis en directo añadiría una capa significativa de complejidad técnica (procesamiento de audio en streaming, latencia de inferencia, gestión de contexto parcial) sin un beneficio proporcional, dado que durante la sesión ya existe una persona moderadora (o varias) cuya función es precisamente interpretar el devenir de la dinámica de la sala y tomar decisiones tácticas en el momento.
- **En segundo lugar**, presentar sugerencias automáticas a la persona encargada del proceso mientras modera podría generar una sobrecarga cognitiva contraproducente, o incluso condicionar su criterio profesional al introducir un sesgo de automatización en decisiones que requieren sensibilidad humana. El valor diferencial del sistema reside en lo que el moderador humano no puede hacer por sí solo: procesar de manera rápida el volumen acumulado de todas las sesiones anteriores, detectar patrones longitudinales en la evolución de posiciones, y ofrecer una visión analítica que integre la totalidad del proceso deliberativo. Este enfoque inter-sesión complementa la labor humana en lugar de intentar sustituirla o competir con ella durante la propia dinámica deliberativa.

## Riesgos y consideraciones de diseño

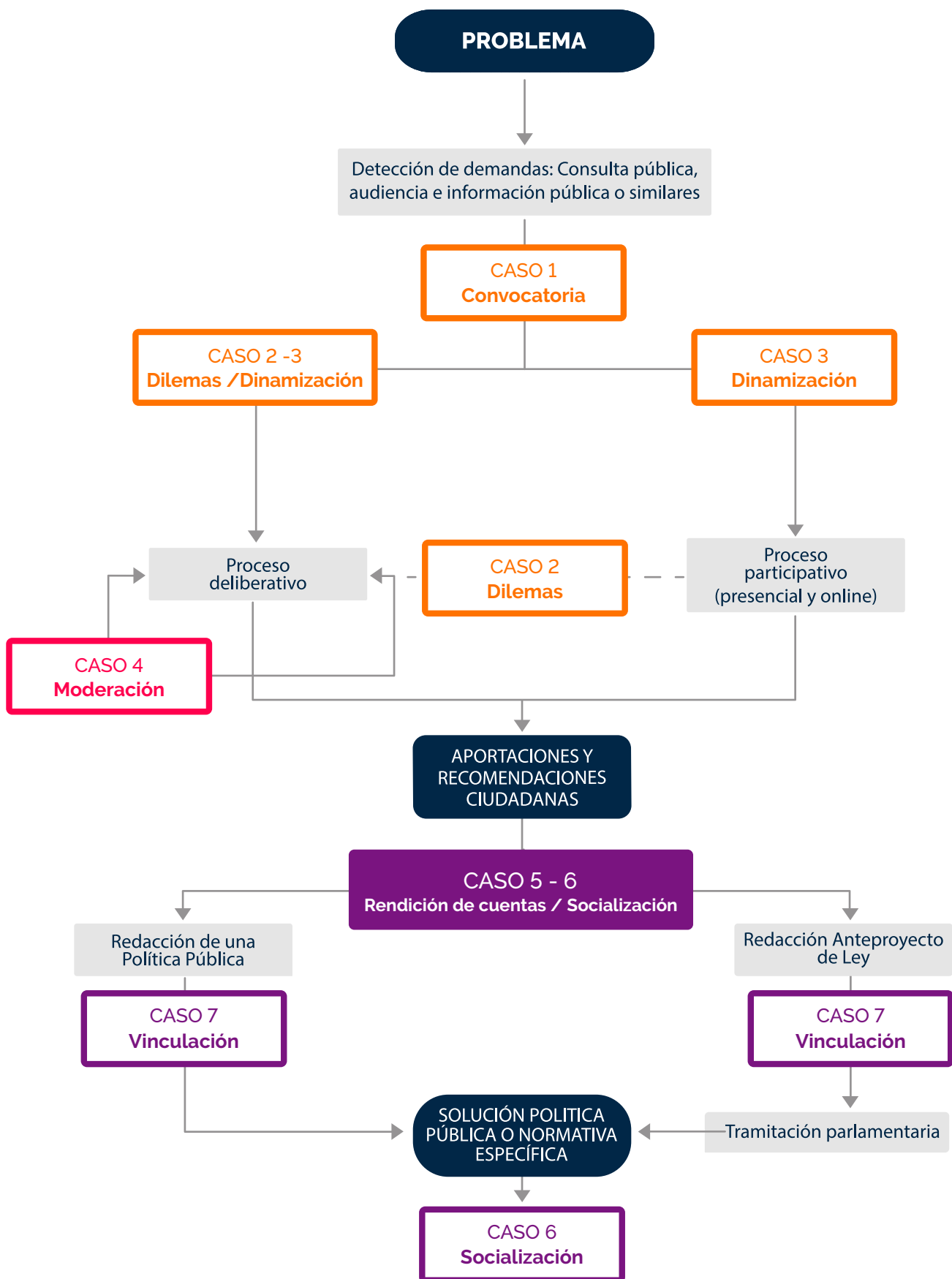
El patrón de compresión progresiva introduce un riesgo inherente: cada ciclo de síntesis implica una pérdida controlada de detalle. Tras muchas iteraciones, matices de las primeras sesiones podrían difuminarse si la compresión es demasiado agresiva. Para mitigar este riesgo se recomienda: conservar siempre las transcripciones originales como respaldo verificable, diseñar el prompt de compresión con criterios explícitos sobre qué información es eliminable (repeticiones, intervenciones procedimentales) y cuál debe preservarse siempre (posiciones sustantivas, propuestas concretas, cambios de posición), e incluir periódicamente un paso de verificación humana del informe de contexto acumulado para detectar posibles derivas o pérdidas significativas.

Adicionalmente, es importante considerar la gestión de expectativas: el sistema ofrece análisis probabilísticos, no certezas. Los informes deben presentarse siempre con indicadores de confianza y las fuentes (transcripciones concretas) que respalden cada conclusión, de modo que el mediador pueda verificar y contrastar cualquier hallazgo que le resulte sorprendente o contraintuitivo.

5

Elaboración  
automatizada  
de materiales  
de rendición de  
cuentas

# Diagrama



# 5.

## Elaboración automatizada de materiales de rendición de cuentas

### Descripción del caso de uso

Una de las asignaturas pendientes habituales de los procesos deliberativos es la comunicación de los resultados una vez el proceso ha concluido, ya sea para socializar los resultados ante el público general o para rendir cuentas ante las personas participantes.

Aunque ambas cuestiones son importantes, no plantean los mismos retos ni requieren el mismo grado de innovación. Mientras que la socialización de los resultados tiene más que ver con la comunicación institucional tradicional (aunque podría beneficiarse de innovaciones provenientes del campo del lenguaje fácil o de la generación de diversos contenidos de comunicación para públicos diferenciados), **la rendición de cuentas plantea el particular reto de personalizar lo máximo posible la respuesta al perfil y aportaciones de cada persona o colectivo que haya participado.**

En cualquier caso, la automatización de estos procesos permitiría a las administraciones efficientar la gestión de los procesos participativos y solventar una de sus tradicionales debilidades que es además fuente de desafección y escepticismo de la ciudadanía respecto de la participación.

## Actores involucrados

En la elaboración de materiales y su revisión previa publicación deberían involucrarse:

- **Personal técnico del Ministerio a cargo del proceso participativo:** son quienes mejor conocerán el proceso realizado y el contenido de sus resultados, y además quienes deberán determinar los requisitos mínimos a cumplir en materia de identidad y lenguaje oficial para las comunicaciones. Se recomienda especialmente contar con personal vinculado con el área de comunicación, que pueda determinar las posibles audiencias a las que se dirigirán los materiales.
- **Personal técnico con conocimientos en lenguaje fácil:** ya sea que se realicen iteraciones con entidades o actores expertos en lenguaje fácil o que alguien que conozca sobre estos temas participe en el proceso de revisión de los contenidos generados.
- **Personal técnico a cargo de dinamizar y/o gestionar el proceso participativo:** resulta clave mantener un diálogo con las personas que hayan estado involucradas directamente en el proceso participativo, y que hayan mantenido diálogo o contacto con las personas participantes, ya que pueden realizar aportaciones sobre el tono a utilizar o los elementos más destacados del debate y contrastar esas ideas con los resultados generados por la IA.
- **Representantes de los principales actores vinculados con el proceso participativo:** se recomienda, sobre todo en los primeros ejercicios que se hagan para elaborar material específico para la rendición de cuentas con actores estratégicos, contar con un pequeño grupo formado por representantes de los distintos grupos que se hayan visto involucrados en el proceso (por ejemplo sociedad civil, academia, sindicatos, etc), que puedan revisar, contrastar y aportar una mirada crítica a los materiales elaborados, con el objetivo de mejorar la calidad de los resultados a través de diversas iteraciones.

## Causas y consecuencias

La comunicación y rendición de cuentas con los resultados de un proceso participativo es una de las grandes tareas pendientes de parte de las administraciones públicas, y representa un elemento clave dentro del proceso de participación ciudadana, ya que brinda confianza en el proceso, facilita la evaluación de su impacto y resultados y permite mantener un vínculo estrecho con los participantes directos del proceso, garantizando su apropiación y además que puedan seguir aportando en las distintas fases siguientes (OCDE, 2022).

Este proceso de rendición de cuentas requiere necesariamente de contar con materiales que sean capaces de cumplir una serie de requisitos, desde la fidelidad de los resultados respecto al proceso llevado a cabo, hasta el respeto de los criterios vinculados con la comunicación clara y la accesibilidad para todas las personas según establece la normativa vigente, en concreto la Ley 6/2022 que establece la necesidad de tomar medidas para garantizar la accesibilidad universal en todas las relaciones con la administración pública, como así también para facilitar la participación en la vida pública y en procesos electorales por parte de todas las personas.

No cumplir con estos requisitos de partida, que requieren de un proceso de reflexión y definición de los elementos clave a tener en cuenta antes de la elaboración de cualquier tipo de material comunicativo, podría implicar que ni siquiera utilizando tecnología de vanguardia se consiga mejorar el vínculo entre los actores estratégicos y las administraciones. Teniendo esto en cuenta, se recomienda tener en cuenta los siguientes elementos para garantizar buenos resultados:

- **Cumplir con lo estipulado en el marco de la Ley 6/2022**, asegurando la accesibilidad universal en el ámbito de las telecomunicaciones, las relaciones con las administraciones públicas y la participación en la vida pública (BOE, 2022)

- **Definir criterios y protocolos de comunicación** clara que deban ser respetados por cualquier tipo de comunicación destinada a rendir cuentas sobre los resultados de un proceso participativo siguiendo las recomendaciones de entidades expertas en la materia (Prodigioso [Volcán](#), 2025).
- **Realizar iteraciones con actores provenientes de diversos contextos y colectivos vulnerables**, que puedan testear las primeras versiones de los contenidos elaborados de manera automatizada, para incorporar mejoras y ajustar el protocolo.
- **Definir las posibles audiencias que recibirán el contenido elaborado**, los canales oficiales a través de los cuales se difundirá la información y el formato de comunicación elegido. Todo esto permitirá construir soluciones basadas en IA que respeten aquellos requisitos formales a cumplir al tratarse de comunicación oficial a la vez que facilitar su lectura, comprensión y apropiación por parte de actores variados.

## Importancia de su resolución

Incorporar una tecnología que facilite la preparación de los materiales para la rendición de cuentas a las personas que formaron parte de un proceso participativo y/o deliberativo no sólo ahorraría tiempo y recursos sino que permitiría cumplir con la última instancia del ciclo de la participación y por ende, dar lugar a un nuevo comienzo. Es decir, la rendición de cuentas no sólo cierra el proceso de participación sino que permite abrir otro y consolidar ese espacio, aportando legitimidad, coherencia y confianza.

Las personas que participaron reciben un feedback de su intervención, conocen los productos y resultados que ha generado, incluso si se realiza un seguimiento en esa rendición, pueden saber los impactos reales y concretos que ha tenido y en qué medida ha alcanzado los objetivos propuestos al inicio.

Este acto de transparencia es fundamental para mejorar la experiencia de la participación, fortalecer la confianza de los participantes, tanto en la instancia y en el espacio como en la institución impulsora, para alentar su involucramiento en próximos procesos, y para que den a conocer entre las personas de su entorno su experiencia y de esa manera multiplicar la participación.

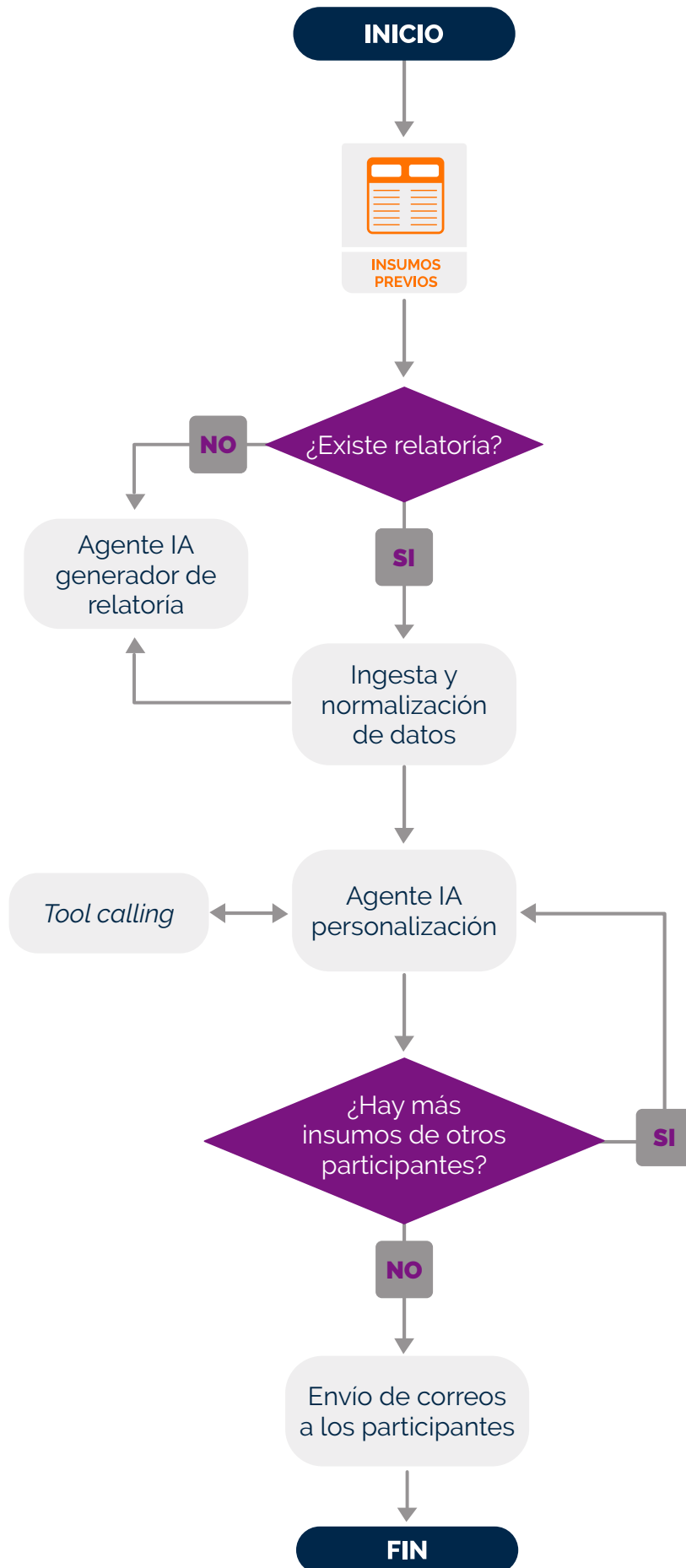
Asimismo, la rendición de cuentas se ve obligada a medir de alguna manera los resultados de la participación y por lo tanto aprender y realizar cambios para mejorar experiencias futuras.

Con todo esto, la resolución de este caso de uso se vuelve crucial en la consecución de un proceso participativo eficiente, cercano y significativo en sus resultados.



Solución  
propuesta

# Diagrama



# Flujo de análisis y rendición de cuentas personalizada basado en agentes para procesos participativos.

## Introducción

Se propone un sistema basado en agentes de IA que automatiza la generación de materiales de rendición de cuentas personalizados tras la finalización de un proceso participativo. A diferencia de la socialización general de resultados, la rendición de cuentas plantea el reto específico de vincular las aportaciones concretas de cada persona o colectivo participante con los resultados derivados del proceso, ofreciendo una respuesta individualizada que refuerce la percepción de escucha efectiva y el valor de la participación.

El sistema combina dos paradigmas de ejecución complementarios, coordinados por un orquestador central implementado con un *framework* como LangGraph. Los módulos de ingesta y preparación de datos siguen pipelines deterministas (secuencias de pasos predefinidos), ya que las tareas de extracción, transcripción y normalización son predecibles independientemente del proceso participativo. En cambio, los módulos de análisis y personalización se implementan como agentes autónomos con acceso a herramientas (lo que se conoce como *tool calling*), que permitiría adaptar su comportamiento dinámicamente a la estructura y contenido concretos de cada proceso.

El sistema se estructura en cuatro módulos: primero, un *pipeline* de ingesta y normalización de datos que extrae la materia prima del proceso participativo desde múltiples fuentes (mensajes de plataformas digitales como Decidim, Discord o foros web; transcripciones textuales de relatorías manuales;

o grabaciones de audio convertidas a texto mediante herramientas de *speech-to-text* con diarización opcional de hablantes). Segundo, un *pipeline* determinista de generación de relatorías que, cuando no exista una relatoría previa, sintetiza automáticamente los intercambios en narrativas estructuradas, incluyendo procesamiento de documentos complementarios mediante OCR cuando sea necesario. Tercero, un agente de análisis con herramientas que examina cada unidad del proceso de forma autónoma, decidiendo en tiempo de ejecución qué operaciones realizar en función del contenido que encuentra, con una fase opcional de síntesis transversal. Y cuarto, un agente de personalización que cruza los resultados del análisis con los datos individuales de cada participante para producir los documentos finales de rendición de cuentas.

## Una arquitectura híbrida; orquestador, pipelines deterministas y agentes con herramientas

Los procesos participativos adoptan dinámicas muy diversas: algunos se organizan por sesiones cronológicas, otros por mesas temáticas, otros por grupos de trabajo, y algunos transcurren como un único flujo continuo de deliberación.

Un *pipeline* con una secuencia fija de pasos internos (como un grafo acíclico predefinido donde cada nodo ejecuta una tarea concreta en orden) resultaría demasiado rígido, ya que los nodos y el orden de ejecución quedarían vinculados a una dinámica participativa concreta. En su lugar, se propone una arquitectura en la que un orquestador central coordina el flujo general de forma determinista, mientras que los módulos de análisis y personalización se implementan como agentes autónomos con acceso a un conjunto de herramientas (*tools*). Cada agente recibe un objetivo y un contexto, y decide autónomamente qué herramientas invocar, en qué orden y cuántas veces, en función de lo que encuentra en los datos.

El orquestador se encarga de la lógica predecible (qué unidades analizar, cuándo sintetizar, para qué participantes personalizar); los agentes se encargan de la lógica adaptativa (cómo analizar cada unidad concreta, qué aspectos profundizar, qué omitir).

#### **a. Agente de análisis**

El agente de análisis opera sobre cada "unidad de análisis", un concepto flexible que se adapta a la estructura del proceso: puede ser una sesión presencial, un hilo temático en una plataforma digital, una mesa de trabajo, un capítulo de un juego prospectivo, o el proceso completo si su volumen lo permite.

El agente dispone de un repertorio de herramientas como: analizar la estructura de la discusión, extraer citas textuales relevantes, identificar consensos y disensos, detectar propuestas concretas, evaluar la intensidad del debate, o clasificar intervenciones por tipo (argumentativa, propositiva, procedimental). En lugar de ejecutar todas estas operaciones en secuencia fija, el agente decide cuáles son pertinentes para cada unidad. Si una sesión contiene un debate polarizado, podrá invocar repetidamente la herramienta de consensos y disensos; si otra es una presentación informativa con poca discusión, podrá centrarse en la extracción de compromisos y omitir el análisis de polarización.

Cada ejecución del agente produce un informe estructurado (definido mediante esquemas como Pydantic) que garantiza un formato unificado y procesable independientemente de qué herramientas se hayan invocado. Cuando el proceso se compone de múltiples unidades, el orquestador lanza un paso de síntesis (que también puede implementarse como agente con herramientas) para integrar los informes parciales y construir una visión transversal: patrones recurrentes, evolución del discurso, dinámicas de participación, recomendaciones y áreas para exploración futura. En procesos con una única unidad, el orquestador omite este paso.

### **b. Agente de personalización**

El agente de personalización es la pieza diferencial del sistema y opera con independencia de cómo se haya segmentado el análisis. El orquestador lo invoca una vez por cada actor participante, proporcionándole el informe de síntesis (o un informe personalizado, según el caso) junto con los datos individuales: sus intervenciones concretas (extraídas del módulo de ingesta y vinculadas a su perfil mediante identificadores de usuario), las unidades en las que participó, y las propuestas que formuló o apoyó. Este agente dispone a su vez de herramientas específicas: buscar las intervenciones de un participante en el corpus, localizar cómo sus aportaciones se reflejan en los datos del proceso, comparar su posición con las posiciones mayoritarias, e identificar los próximos pasos relacionados con los temas en los que participó.

A partir de estas consultas, genera un documento de rendición de cuentas personalizado que responde a preguntas como: ¿qué propuestas concretas hizo esta persona o colectivo y cuál fue su recorrido en el proceso?, ¿qué decisiones o compromisos se derivaron de las temáticas en las que participó?, ¿dónde coincidió con posiciones mayoritarias y dónde sus aportaciones representaron perspectivas minoritarias?

Esta separación entre orquestación determinista y ejecución adaptativa mediante *tool calling* presenta una ventaja fundamental frente a los *pipelines* íntegramente lineales: el mismo sistema puede operar sobre un proceso organizado por sesiones cronológicas, por temáticas transversales, por grupos de trabajo o como flujo continuo, sin modificar ni la lógica del orquestador ni las herramientas de los agentes. Lo que cambia entre procesos es únicamente la configuración de las unidades de análisis y los *prompts* de los agentes, no la arquitectura del sistema.

## Extracción de datos en procesos digitales e híbridos

Un aspecto clave del sistema es su capacidad para extraer datos directamente de las plataformas donde se desarrolla la participación. En procesos total o parcialmente digitales, los mensajes, reacciones e interacciones de los participantes quedan registrados en plataformas conversacionales (Decidim, Discord, foros web u otras plataformas de participación dedicadas) y pueden extraerse de forma sistemática mediante las APIs de dichos sistemas. Por ejemplo, es posible extraer automáticamente todos los mensajes de una plataforma como Discord junto con los perfiles de participantes registrados en herramientas de gestión de la información como Notion, vinculándolos mediante identificadores de usuario para construir un conjunto de datos completo que incluya quién dijo qué, en qué contexto y en respuesta a qué tema.

Para procesos presenciales o híbridos donde no exista registro digital nativo, el módulo de ingesta incorpora capacidades de procesamiento de audio (*speech-to-text* con diarización de hablantes para identificar quién interviene en cada momento) y de documentos (OCR para relatorías manuscritas o escaneadas, extracción de texto de documentos en diversos formatos). La combinación de ambas vías permite que el sistema opere tanto en procesos nativamente digitales, donde la trazabilidad individual es inmediata, como en procesos presenciales, donde se requiere un paso adicional de transcripción y atribución.

## Generación automatizada de relatorías

Cuando el proceso no cuente con relatorías elaboradas manualmente, el sistema puede generar relatorías sintéticas a partir de los datos brutos. Este módulo toma como entrada las transcripciones de audio procesadas o los mensajes extraídos de plataformas digitales y produce narrativas estructuradas que resumen el desarrollo de cada sesión o temática: contexto del debate, principales líneas argumentales, propuestas formuladas y dinámica de la participación.

Estas relatorías generadas automáticamente sirven tanto como insumo para el pipeline de análisis como para la documentación del proceso, y pueden someterse a revisión humana antes de su uso si se considera necesario.

## Riesgos y consideraciones de diseño

La personalización automatizada de la rendición de cuentas introduce consideraciones específicas. En primer lugar, la calidad de la atribución individual depende directamente de la calidad de los datos de entrada: en procesos digitales, la trazabilidad es alta porque cada mensaje está vinculado a un usuario identificado; en procesos presenciales, la diarización de audio tiene tasas de error que pueden afectar a la correcta atribución de intervenciones. Se recomienda incorporar un paso con validación humana (*human-in-the-loop*) especialmente en los materiales personalizados, donde un error de atribución podría generar desconfianza en lugar de reforzarla.

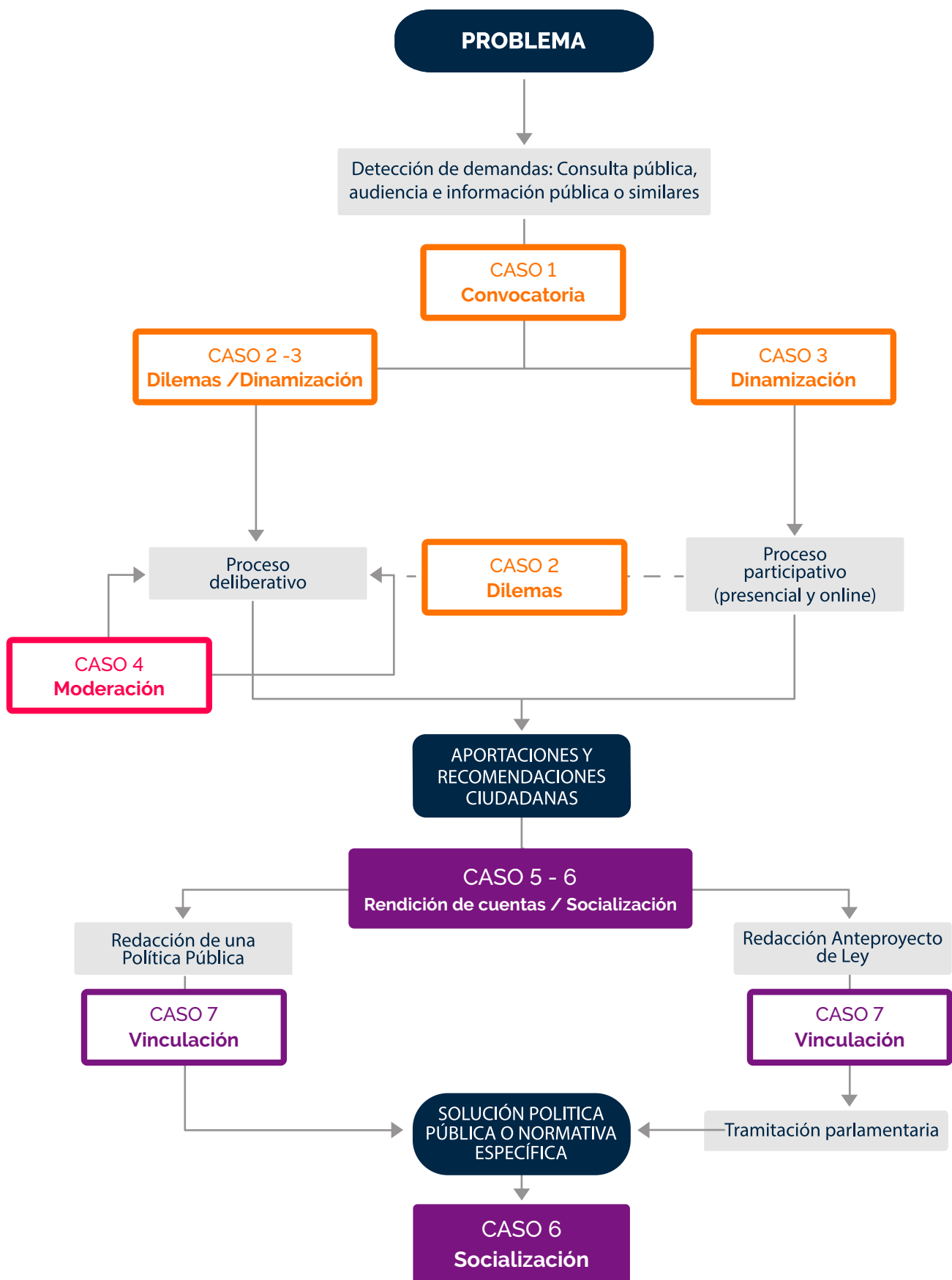
En segundo lugar, el sistema debe manejar con cuidado el equilibrio entre personalización y privacidad: los materiales de rendición de cuentas dirigidos a un participante concreto no deben revelar información atribuible a otros participantes más allá de lo que ya fuera público durante el propio proceso participativo. El diseño del agente de personalización debe incluir reglas explícitas sobre qué información es compartible y cuál debe anonimizarse o agregarse.

Por último, la escalabilidad del agente de personalización es un factor crítico: en procesos con decenas o cientos de participantes, el sistema debe ser capaz de generar materiales individualizados de forma eficiente. La arquitectura modular facilita esta escalabilidad, ya que la fase de personalización puede ejecutarse en paralelo para distintos participantes, reutilizando los informes de análisis y síntesis generados una única vez en la fase anterior.

6

Producción de  
materiales para la  
socialización de  
resultados

# Diagrama



# 6.

## Producción de materiales para la socialización de resultados

### Descripción del caso de uso

Una de las asignaturas pendientes habituales de los procesos deliberativos es la comunicación de los resultados una vez el proceso ha concluido, ya sea para socializar los resultados ante el público general o para rendir cuentas ante las personas participantes.

Aunque ambas cuestiones son importantes, no plantean los mismos retos ni requieren el mismo grado de innovación. Mientras que **la socialización de los resultados tiene más que ver con la comunicación institucional tradicional (aunque podría beneficiarse de innovaciones provenientes del campo del lenguaje fácil o de la generación de diversos contenidos de comunicación para públicos diferenciados)**, la rendición de cuentas plantea el particular reto de personalizar lo máximo posible la respuesta al perfil y aportaciones de cada persona o colectivo que haya participado.

En cualquier caso, la automatización de estos procesos permitiría a las administraciones efficientar la gestión de los procesos participativos y solventar una de sus tradicionales debilidades que es además fuente de desafección y escepticismo de la ciudadanía respecto de la participación.

## Actores involucrados

En la elaboración de materiales y su revisión previa publicación deberían involucrarse:

- **Personal técnico del Ministerio a cargo del proceso participativo:** son quienes mejor conocen el proceso realizado y el contenido de sus resultados, y además quienes deberán determinar los requisitos mínimos a cumplir en materia de identidad y lenguaje oficial para las comunicaciones. Se recomienda especialmente contar con personal vinculado con el área de comunicación, que pueda determinar las posibles audiencias a las que se dirigirán los materiales.
- **Personal técnico con conocimientos en lenguaje fácil:** ya sea que se realicen iteraciones con entidades o actores expertos en lenguaje fácil o que alguien que conozca sobre estos temas participe en el proceso de revisión de los contenidos generados.
- **Personal técnico a cargo de dinamizar y/o gestionar el proceso participativo:** resulta clave mantener un diálogo con las personas que hayan estado involucradas directamente en el proceso participativo, y que hayan mantenido diálogo o contacto con las personas participantes, ya que pueden realizar aportaciones sobre el tono a utilizar o los elementos más destacados del debate y contrastar esas ideas con los resultados generados por la IA.

## Causas y consecuencias

La comunicación y socialización de resultados es una de las grandes tareas pendientes de parte de las administraciones públicas, y representa un elemento clave dentro del proceso de participación ciudadana, ya que brinda confianza en el proceso, facilita la evaluación de su impacto y resultados

y permite mantener un vínculo estrecho con la ciudadanía más allá de los participantes directos del proceso (OCDE, 2022).

Este proceso de socialización requiere necesariamente de contar con materiales que sean capaces de cumplir una serie de requisitos, desde la fidelidad de los resultados respecto al proceso llevado a cabo, hasta el respeto de los criterios vinculados con la comunicación clara y la accesibilidad para todas las personas según establece la normativa vigente, en concreto la Ley 6/2022 que establece la necesidad de tomar medidas para garantizar la accesibilidad universal en todas las relaciones con la administración pública, como así también para facilitar la participación en la vida pública y en procesos electorales por parte de todas las personas.

No cumplir con estos requisitos de partida, que requieren de un proceso de reflexión y definición de los elementos clave a tener en cuenta antes de la elaboración de cualquier tipo de material comunicativo, podría implicar que ni siquiera utilizando tecnología de vanguardia se consiga mejorar el vínculo entre ciudadanía y administraciones. Teniendo esto en cuenta, se recomienda tener en cuenta los siguientes elementos para garantizar buenos resultados:

- **Cumplir con lo estipulado** en el marco de la Ley 6/2022, asegurando la accesibilidad universal en el ámbito de las telecomunicaciones, las relaciones con las administraciones públicas y la participación en la vida pública (BOE, 2022)
- **Definir criterios y protocolos de comunicación** clara que deban ser respetados por cualquier tipo de comunicación destinada a socializar resultados de un proceso participativo siguiendo las recomendaciones de entidades expertas en la materia (*Prodigioso Volcán*, 2025).
- **Realizar iteraciones** con actores provenientes de diversos contextos y colectivos vulnerables, que puedan testear las primeras versiones de los contenidos elaborados de manera automatizada, para incorporar mejoras y ajustar el protocolo.

- **Definir las posibles audiencias** que recibirán el contenido elaborado, los canales oficiales a través de los cuales se difundirá la información y el formato de comunicación elegido. Todo esto permitirá construir soluciones basadas en IA que respeten aquellos requisitos formales a cumplir al tratarse de comunicación oficial a la vez que facilitar su lectura, comprensión y apropiación por parte de actores variados.

## Importancia de su resolución

Incorporar una tecnología que facilite la preparación de los materiales para la socialización de los resultados, de manera clara, sencilla y accesible a las personas con discapacidad, no sólo ahorraría tiempo y recursos sino que significaría un acto de transparencia y responsabilidad pública, institucional y política.

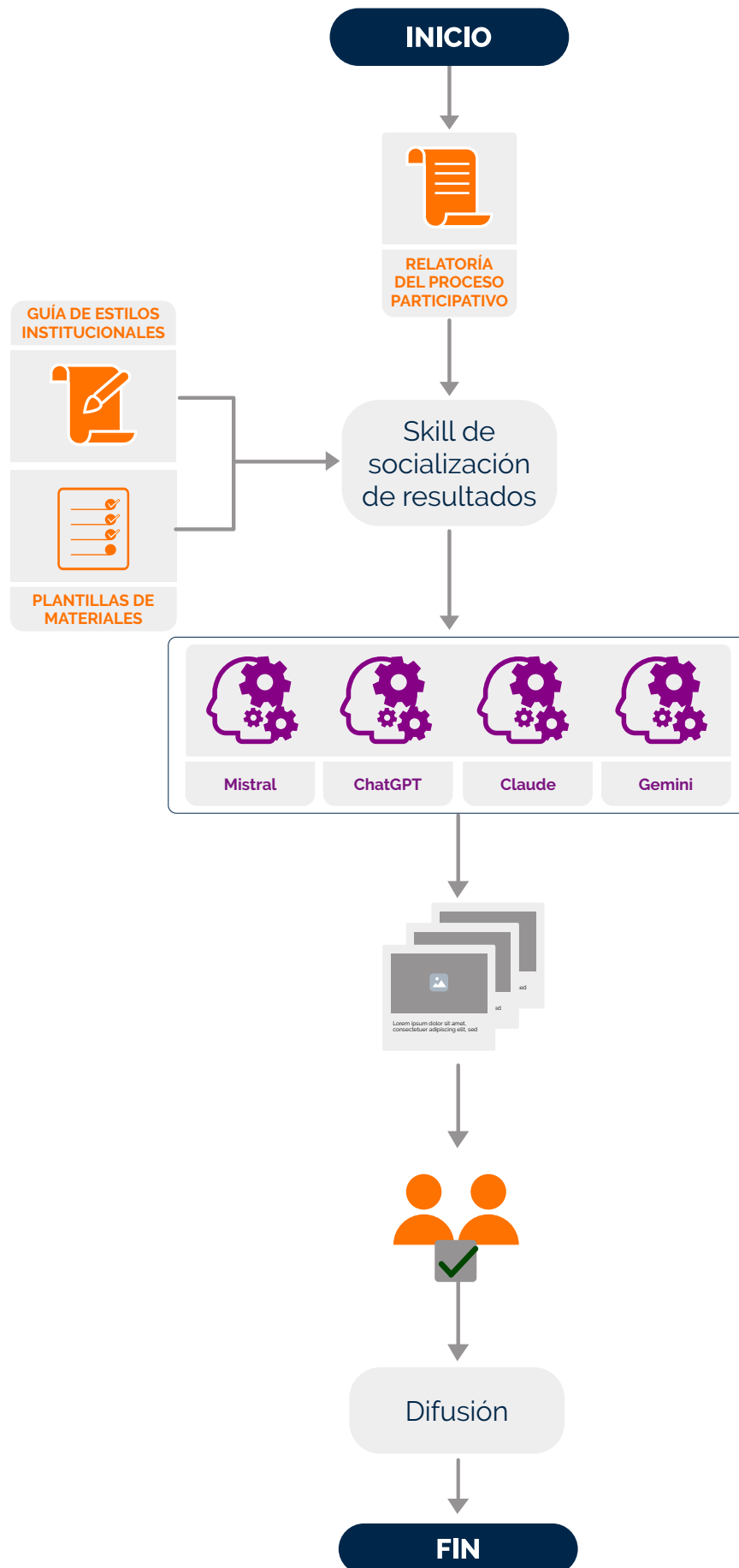
Que los diversos actores sociales, de los más amplios sectores, y la ciudadanía en general, incluidos los colectivos vulnerables y/o minoritarios, conozcan los frutos de un proceso participativo/deliberativo es un acto de promoción de la participación en sí mismo, ya que la exposición de los productos, resultados e impactos del proceso en los problemas o dilemas públicos abordados, contribuye a fortalecer la confianza en él y alienta a que más personas e instituciones se involucren en próximas instancias.

Asimismo, la socialización de resultados implica medirlos de alguna manera y por lo tanto identificar oportunidades de mejora, aprender y realizar los cambios necesarios en experiencias futuras.



Solución  
propuesta

# Diagrama



# Socialización de resultados participativos mediante asistentes de IA y el estándar abierto de *skills*

## Introducción

Partiendo del escenario en el que se dispone de una relatoría del proceso participativo (ya sea elaborada manualmente o mediante herramientas de IA), se propone un enfoque progresivo que prioriza la rapidez de implantación y el bajo coste inicial, con capacidad de evolucionar hacia soluciones más avanzadas y a medida que las necesidades lo justifiquen.

## Estructura de la solución propuesta

### 1. Fase Enfoque low-code con *skills* y *plugins* para asistentes de IA (corto plazo)

La forma más rápida y accesible de comenzar es aprovechar el estándar abierto de *skills* que ha sido adoptado por los principales proveedores de IA. Una *skill* es un paquete autocontenido —típicamente un archivo de manifiesto con instrucciones, plantillas y herramientas— que define cómo un asistente de IA debe abordar una tarea específica.

Al tratarse de un estándar abierto y no de una funcionalidad propietaria, las *skills* son portables entre proveedores: la misma *skill* puede ejecutarse en Claude (conversaciones estándar, Cowork, Claude Code y API), en ChatGPT (a través de Codex y Responses API), en Mistral (con Mistral Vibe) o en Gemini

(mediante Gemini CLI). Esta interoperabilidad elimina la dependencia de un único fabricante y permite a la administración pública evaluar o cambiar de proveedor sin perder el trabajo invertido en la definición de sus flujos.

En la práctica, se configuraría un agente o *skill* de "socialización de resultados participativos" en la plataforma elegida, que encapsularía: las plantillas de cada tipo de material (resumen ejecutivo, comunicado en lenguaje fácil, nota de prensa, contenidos para redes sociales segmentados por público, cartas de devolución para participantes), la guía de estilo institucional, los glosarios relevantes y los criterios de calidad.

Un técnico de participación simplemente cargaría la relatoría del proceso y el asistente se encargaría de generar el paquete completo de materiales siguiendo las pautas predefinidas. El técnico revisa, ajusta y aprueba cada material antes de su difusión. Si la configuración necesita mejoras, el propio equipo puede iterar sobre ella sin depender de desarrolladores externos.

Las ventajas de este enfoque son significativas: no requiere infraestructura propia ni desarrollo a medida, el coste es el de las suscripciones a la plataforma elegida, se puede poner en marcha en días y el personal técnico puede iterar con las *skills* sin depender de equipos de desarrollo. Al basarse en un estándar abierto, la misma *skill* puede reutilizarse en distintos departamentos, compartirse como buena práctica entre instituciones o incluso publicarse como recurso abierto para otras administraciones.

Dado que las *skills* son archivos de texto plano, pueden versionarse con sistemas de control de cambios (como Git), auditarse y documentarse de forma transparente, lo que responde a las exigencias de trazabilidad y rendición de cuentas de la administración pública.

## 2. Fase – Automatización con *workflows low-code* (medio plazo)

Si la Fase 1 demuestra su utilidad y el volumen de procesos lo justifica, el siguiente paso natural es automatizar el flujo completo mediante plataformas de orquestación *low-code* como n8n, *Activepieces* u otras plataformas de orquestación *open-source* y auto-alojables.

Estas herramientas permiten construir *pipelines* que, al recibir una relatoría, ejecutan automáticamente la cadena de generación: extraen las secciones clave del documento, invocan al modelo de lenguaje con los *prompts* ya validados en la Fase 1, generan el paquete completo de materiales en distintos formatos y los depositan en una carpeta de revisión para el técnico. Se mantiene el *human-in-the-loop*: nada se publica sin aprobación humana, pero el trabajo manual se reduce drásticamente.

### 3. Fase – *Pipeline* integral de procesamiento y generación (largo plazo)

Las Fases 1 y 2 asumen que ya se dispone de una relatoría elaborada del proceso participativo. Sin embargo, la realidad de muchos procesos es que la documentación disponible al concluir es un conjunto heterogéneo de materiales sin procesar: actas manuscritas, grabaciones de audio de sesiones presenciales, fotografías de paneles o pósit, formularios escaneados, hilos de debate en plataformas digitales y datos de votaciones o encuestas.

El salto cualitativo de la Fase 3 consiste en construir un *pipeline* integral que cubra todo el camino desde estos insumos brutos hasta los materiales de comunicación finales, incluyendo la generación automatizada de la propia relatoría como paso intermedio.

Este *pipeline* se estructura en tres bloques encadenados. El primero es la ingesta y procesamiento de documentos: OCR para documentos escaneados y fotografías de materiales físicos, transcripción automática de grabaciones de audio y vídeo (con diarización para identificar hablantes), extracción estructurada de datos de formularios y encuestas, y normalización de aportaciones procedentes de plataformas digitales de participación.

El segundo bloque es la síntesis y generación de la relatoría: un agente de IA que, a partir de todos los insumos procesados, elabora un borrador de relatoría estructurada —identificando temas principales, puntos de consenso y disenso, propuestas concretas y datos cuantitativos de participación— que un técnico revisa y valida.

El tercer bloque es la generación de materiales de comunicación, donde se reutilizan directamente las *skills* definidas en las Fases 1 y 2: la relatoría validada alimenta las mismas plantillas y criterios de calidad ya probados para producir el paquete completo de materiales de socialización.

La construcción de un *pipeline* de esta complejidad requiere *frameworks* de ingeniería de agentes que permitan orquestar múltiples pasos de IA, gestionar el flujo de datos entre ellos y mantener el control humano en los puntos críticos. Existen varias opciones *open-source* maduras para este propósito: LangGraph, que ofrece una arquitectura basada en grafos dirigidos especialmente adecuada para *workflows* multi-paso con bifurcaciones y validaciones intermedias; Haystack, que está específicamente diseñado para *pipelines* de procesamiento de documentos (OCR, extracción, RAG) y es una opción natural para el bloque de ingesta; o combinaciones de ambos según las necesidades de cada bloque del *pipeline*. Estos *frameworks* son compatibles con modelos *open-source* desplegados en infraestructura propia (como Mistral, Granite o Gemma), lo que permite cumplir con requisitos de soberanía de datos cuando sea necesario.

Es importante subrayar que esta fase no parte de cero. Las *skills* de generación de materiales desarrolladas y validadas en las fases anteriores se integran como componentes reutilizables del último bloque del *pipeline*, lo que reduce significativamente el esfuerzo de desarrollo y garantiza la coherencia con los estándares de calidad ya establecidos. Del mismo modo, los *prompts* optimizados, las guías de estilo y los criterios de validación acumulados se incorporan directamente. La Fase 3 añade complejidad en la ingesta y síntesis, pero no reinventa la capa de generación de salidas.

Esta fase se justifica cuando confluyen tres condiciones: el volumen de procesos participativos es suficiente para amortizar la inversión en desarrollo, la administración necesita reducir el cuello de botella que supone la elaboración manual de relatorías (que a menudo es la razón principal por la que la comunicación de resultados se retrasa o no llega a producirse), y existen requisitos de soberanía de datos o integración con sistemas internos que impiden depender de plataformas comerciales externas. Sin estas condiciones, las Fases 1 y 2 cubren sobradamente las necesidades.

## ¿Por qué este enfoque progresivo y no una solución directa?

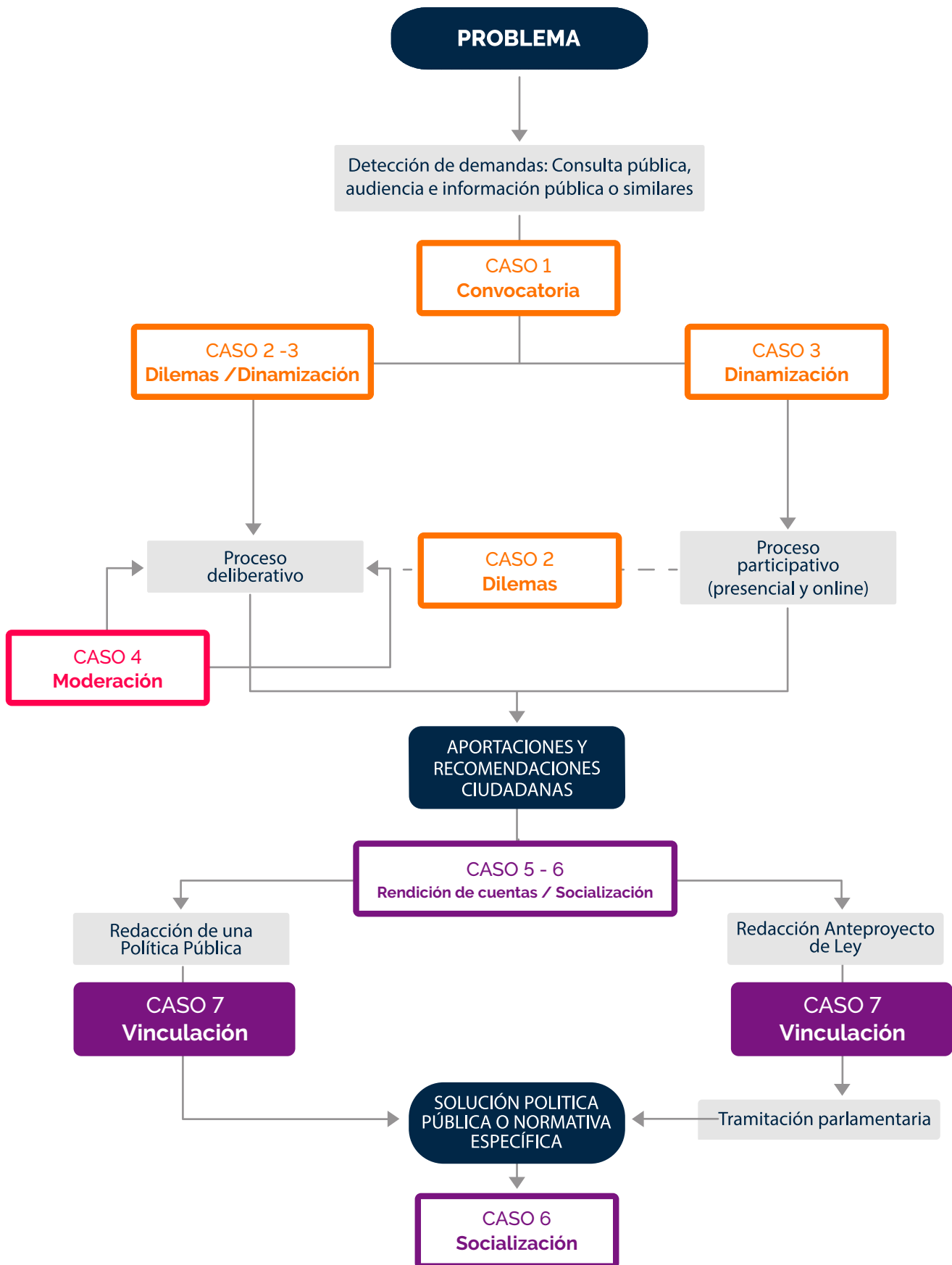
El campo de los modelos de lenguaje evoluciona con una rapidez sin precedentes. Invertir hoy en una solución a medida implica el riesgo de que quede obsoleta en meses, mientras que un enfoque basado en el estándar abierto de *skills* se beneficia automáticamente de cada mejora del modelo subyacente y protege la inversión frente a cambios en el mercado de proveedores: si un proveedor deja de ser competitivo, las *skills* migran a otro sin reescritura.

La Fase 1 permite validar la utilidad real de la IA en este contexto con una inversión mínima: si los materiales generados no alcanzan la calidad esperada o si el equipo técnico no adopta la herramienta, se habrá perdido muy poco. Y si funciona, las *skills*, las plantillas y el conocimiento acumulado migran directamente a las fases siguientes; de hecho, las *skills* definidas en la fase 1 pueden integrarse tal cual en los *workflows* automatizados de la fase 2. Este enfoque es el más prudente para una administración pública que debe justificar cada inversión tecnológica.



Vinculación  
entre insumos  
participativos  
y producción  
normativa

# Diagrama



# 7.

## Vinculación entre insumos participativos y producción normativa

### Descripción del caso de uso

En ocasiones, el resultado final de un proceso participativo es una propuesta de texto normativo (i.e. un Proyecto de Ley) que es sometido por el Ejecutivo a tramitación parlamentaria para su aprobación. Se considera oportuno que los legisladores, los medios de comunicación y el público en general, puedan conocer cuáles son los insumos que han sido considerados por el Ejecutivo para la elaboración de cada una de las disposiciones del texto legislativo, para mejorar su comprensión acerca de las razones y argumentos que lo sostienen y para mostrar su pertinencia en términos de legitimidad ciudadana y científica/académica, así como su solvencia técnica, jurídica o económica.

Para ello, se plantea la elaboración de una **herramienta/artefacto que permita una lectura del texto legislativo enriquecida con dichas capas de información**. En el diseño de esta herramienta/artefacto habrá que valorar la posibilidad de ofrecer tanto un resumen cuantitativo de los insumos vinculados con cada disposición del texto normativo (por ejemplo X propuestas del proceso de consulta ciudadana e Y referencias bibliográficas) como la posibilidad de consultar el detalle del contenido de dichos insumos.

En este caso de uso particular, vinculado al proceso de elaboración del Proyecto de Ley de Juventud, estos insumos consisten por un lado en las aportaciones de entidades y ciudadanía durante el proceso participativo previo (consulta ciudadana y talleres de diálogo organizados por el Ministerio de Juventud e Infancia), los insumos aportados por diferentes grupos de interés, y las aportaciones de la comunidad científica, recogidas a modo de recomendaciones de política pública en los capítulos del Informe de Juventud en España.

### Actores involucrados:

En la detección, recopilación y selección de la información pertinente:

- **Personal del Ministerio** involucrado en los procesos de consulta previos a la elaboración del Proyecto de Ley. Aquí se recomienda la existencia de un punto focal o una persona responsable de coordinación del proceso de recopilación de información, en el supuesto de que haya diversos procesos de consulta.
- **Asesores científicos del Ministerio** correspondiente, cuyo rol principalmente consiste en brindar asistencia técnica al gabinete del Ministro/a en procesos de elaboración de política pública o de elaboración de normativa específica.
- **Sociedad civil organizada**, cuya labor de incidencia no solamente representa una fuente de información acerca de las demandas ciudadanas que pueden aportar a un proyecto normativo, y al mismo tiempo son un actor clave que puede formar parte de los procesos de iteración vinculados a la mejora de la herramienta planteada en este caso de uso, verificando su funcionamiento y valorando si el análisis y las interconexiones entre las aportaciones y el texto normativo planteado son adecuadas.

## Causas y consecuencias:

Este caso de uso surge a partir de una realidad de los procesos normativos actuales, y es la falta de transparencia y trazabilidad entre los insumos recogidos en un proceso de elaboración normativo y el texto del proyecto de ley, así como la desconexión entre dicha fase previa y la fase de tramitación parlamentaria de la propuesta normativa.

El proceso de consulta y audiencia pública se encuentran regulados por la Ley 39/2015 del Procedimiento Administrativo Común de las Administraciones Públicas, la Ley 50/1997 del Gobierno y la Orden PRE/1590/2016 donde se dictan instrucciones para habilitar la participación pública en el proceso de elaboración normativa a través de los portales web de los departamentos ministeriales. El objetivo es recabar la opinión de los sujetos y organizaciones más representativas "potencialmente afectados por la futura norma" y se puede consultar sobre la problemática a solucionar, la necesidad de la norma, los objetivos específicos y las soluciones planteadas. Sin embargo, no existe una normativa específica que regule la publicación de la información recabada en estos procesos de consulta, no más allá de lo expresado en la Ley 19/2013 de Transparencia, acceso a la información pública y buen gobierno, la cual no cuenta con disposiciones específicas respecto a esta fase del ciclo de política pública, aunque en su artículo 5 define que los sujetos determinados por la ley deberán publicar de manera "periódica y actualizada la información cuyo conocimiento sea relevante para garantizar la transparencia de su actividad relacionada con el funcionamiento y el control de la actuación pública". Sin embargo, esta ambigüedad ha generado una brecha respecto a este tipo de información que se recoge en la fase de diseño de nuevas normativas, y que en términos generales, no es publicada por los organismos responsables.

Por otra parte, los Ministerios también recurren habitualmente a otras fuentes de información a la hora de elaborar un texto normativo. Estas fuentes pueden provenir de producción científica y académica (producida ad hoc o no); otros espacios de diálogo con los actores interesados, etc.

Se trata de información que permite comprender mejor el contexto y las necesidades que pueden dar origen y orientar las medidas promovidas por un texto normativo, y cuyo conocimiento puede ser útil en la apropiación de ese texto por parte de la ciudadanía y en el debate parlamentario posterior.

Una vez realizada la fase de consultas previas, redactada la propuesta normativa y sometida a su debate en el ámbito parlamentario, no existe ninguna herramienta que permita trazar de forma clara la relación entre los insumos considerados y las disposiciones finalmente incluidas en los textos normativos aprobados en Consejo de Ministros y enviados al Congreso de los Diputados para su tramitación.

### **Esto genera una serie de consecuencias:**

- **En primer lugar**, genera un problema de transparencia con la ciudadanía y colectivos potencialmente afectados por la normativa sometida a tramitación. No es posible siendo un colectivo o un individuo posiblemente afectado por una normativa conocer en qué medida las aportaciones enviadas han influido en el texto enviado a la sede parlamentaria, generando grandes dificultades en la trazabilidad y la rendición de cuentas del proceso participativo. Aunque la rendición de cuentas a las personas y colectivos participantes en los procesos de consulta puede tener otros canales (que son objeto de un caso de uso distinto en el marco de esta publicación) resulta esencial que estos puedan consultar de una forma clara y accesible qué insumos han sido tenidos en cuenta en la redacción final del Proyecto de Ley.
- **En segundo lugar**, se genera el mismo efecto con los diputados y diputadas a cargo de la consideración, deliberación y potencial aprobación de la normativa, para que puedan conocer en qué medida el articulado de la normativa cuenta con legitimidad pública y social, así como su respaldo técnico o científico.

- **En tercer lugar**, esta falta de transparencia y trazabilidad dificulta el consenso en el proceso legislativo. Si bien el consenso parlamentario depende de un complejo conjunto de factores (que van desde la temática que aborde la Ley en concreto, su urgencia, los colectivos potencialmente afectados, el momento exacto en el que se presente o la situación política en su conjunto), sí parece cierto que una mayor trazabilidad podría influir positivamente en la calidad del debate y en el proceso de enmiendas por parte de los grupos parlamentarios, que contarían con mayor y mejor información acerca de los insumos que han motivado cada uno de los artículos de la norma.

### Importancia de su resolución :

Entre los potenciales impactos positivos que se esperan del abordaje de este caso de uso se encuentran:

- **Brindar a los participantes de la consulta/proceso deliberativo una devolución adecuada de cómo sus aportaciones se han transformado en propuestas concretas de la normativa.** Esto no es óbice de que una primera rendición de cuentas habrá tenido que producirse al término del proceso participativo del que se trate, sin necesidad de esperar a la finalización de la elaboración del texto normativo, teniendo en cuenta el plazo que pueda transcurrir entre un momento y otro y el hecho que muchas de las aportaciones planteadas por los participantes han podido quedar descartadas de inicio sin ninguna posibilidad de quedar plasmadas en el Proyecto de Ley específico.
- **Ofrecer a los actores políticos vinculados al proceso legislativo insumos que favorezcan una comprensión integral y más amplia de la normativa** sometida a tramitación y de los argumentos y evidencias que la sustentan. De cara a una potencial rendición de cuentas por parte de los partidos políticos a la ciudadanía tras la tramitación

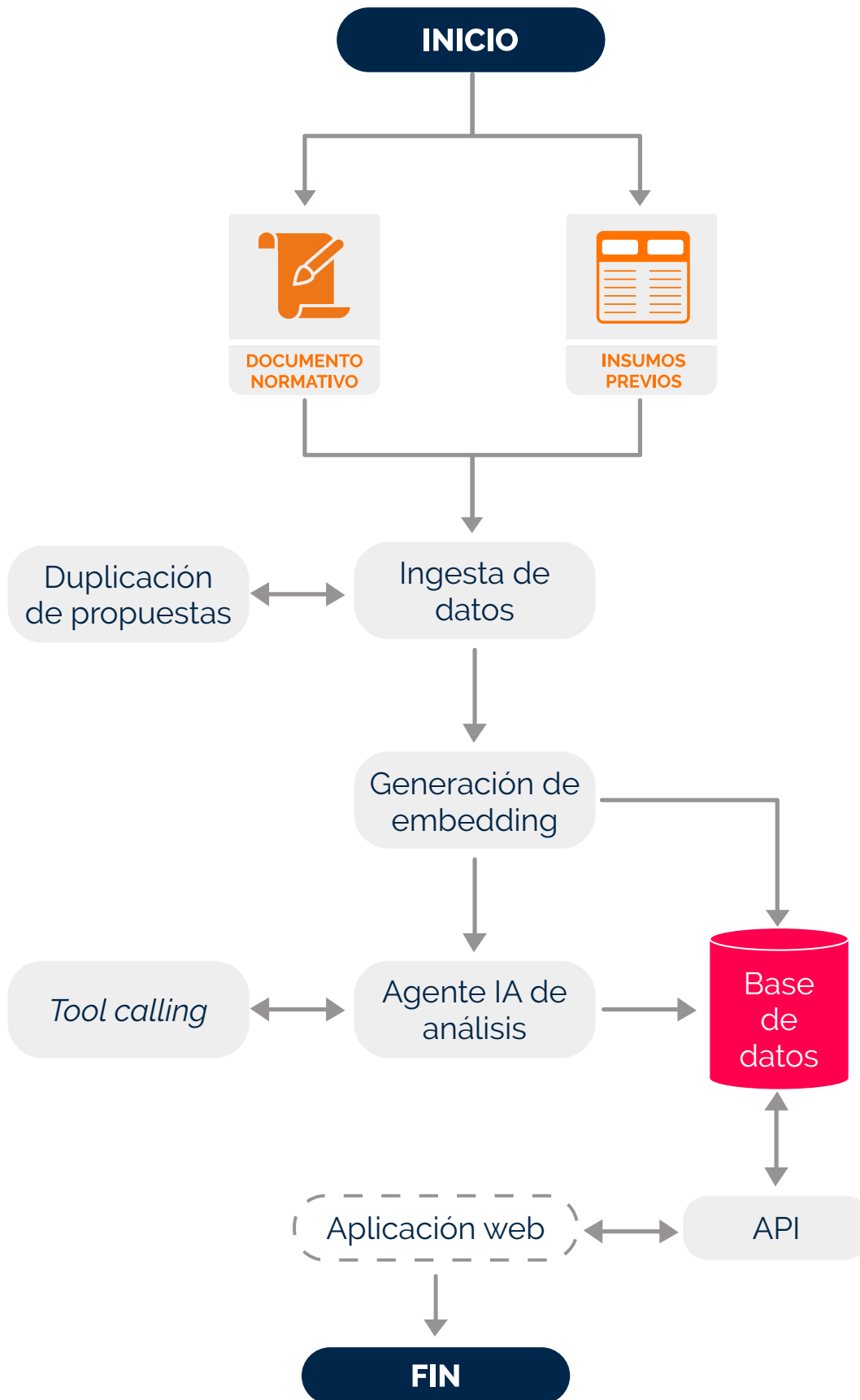
parlamentaria, esta herramienta también sería de valor, al informar a los partidos qué actores están detrás de las disposiciones que han apoyado o rechazado.

- **Aportar una herramienta de consulta con un valor añadido diferencial al que aportan las actuales Memorias de Análisis de Impacto Normativo (MAIN) la cual**, en su configuración y regulación actual tiene por objetivo justificar y dar a conocer la oportunidad y necesidad de una nueva norma, valorando su potencial impacto en diferentes ámbitos de la realidad. Las MAIN en la actualidad, aunque son de obligada elaboración, cuentan con sus propios puntos de mejora debido a su excesivo formalismo y carencias en materia de análisis ex ante de los potenciales impactos de una norma. Ante esta situación, la herramienta planteada en este caso de uso representa un valor añadido, ya que aporta información clave que en la actualidad no se encuentra contenida ni disponible para ninguno de los actores involucrados o potencialmente afectados por un proyecto normativo, brindando una capa extra de información acerca de la opinión pública y las aportaciones ciudadanas sobre una temática concreta.
- **El hecho de poder conocer los insumos y aportes que sustentan un texto normativo**, si se convierte en práctica habitual, habrá promovido la educación y la cultura cívica, y contribuido a fortalecer la confianza ciudadana en las instituciones.



Solución  
propuesta

# Diagrama



# Vinculación automática de insumos participativos con texto normativo mediante búsqueda semántica y agente de verificación

## Arquitectura general

La solución se estructura en tres componentes. En primer lugar, un *pipeline* de procesamiento que se encarga de ingestar los documentos, extraer y deduplicar las propuestas, generar los *embeddings* y ejecutar el agente de vinculación; es el componente que realiza todo el trabajo analítico pesado y se apoya en un agente orquestado con LangGraph. Segundo, una API (realizada con FastAPI) que expone los resultados almacenados y sirve como punto de acceso para el tercer componente: una aplicación web de exploración (basada en Nuxt) que permite navegar el texto normativo y consultar los insumos asociados a cada disposición.

## Ingesta de insumos participativos

El *pipeline* recibe dos tipos de entradas de insumos participativos que requieren procesadores distintos.

- Si se reciben como un único archivo .docx o .xlsx con una tabla resumen, el procesamiento es directo: se procesa la tabla y las entradas de cada fila se convierten en una propuesta con sus metadatos.
- Si en cambio se reciben como un conjunto de documentos PDF (uno por autor o entidad), el *pipeline* utiliza Docling para extraer el texto de cada documento y segmentarlo en propuestas individuales, ya que un mismo PDF puede contener múltiples aportaciones diferenciadas.

## Procesamiento del documento normativo

La segunda entrada es el documento normativo producido (un proyecto de ley, un plan estratégico o de acción u otro instrumento equivalente). El objetivo es intentar vincular cada propuesta con las disposiciones del texto normativo en las que ha sido incorporada, indicando el grado de incorporación y una explicación de por qué se considera que existe dicha vinculación.

El procesamiento del documento normativo también utiliza Docling, pero con un objetivo distinto: extraer su estructura semántica (títulos, capítulos, artículos, disposiciones) y generar fragmentos que respeten la jerarquía del texto legal. Un problema previsible en este paso es la distancia lingüística entre el lenguaje del texto normativo (técnico-jurídico) y el de las propuestas ciudadanas (más directo e informal). Para mitigar esta brecha, el *pipeline* genera una versión adaptada a lectura fácil de cada sección del documento normativo mediante un LLM. Esta versión se almacena junto con el texto original y es la que se utiliza como base para la búsqueda semántica, mejorando significativamente la calidad de la recuperación.

## Vinculación mediante RAG y verificación por LLM

Las secciones procesadas se almacenan en una base de datos PostgreSQL con la extensión pgvector, que permite realizar búsqueda vectorial sobre los *embeddings* de los textos. Para cada propuesta, el sistema recupera un conjunto amplio de secciones candidatas (las más cercanas semánticamente) y las pasa como contexto a un agente de verificación orquestado con LangGraph. El agente, mediante *tool calling* y potencialmente subagentes especializados, analiza cada par propuesta-sección para determinar si existe una vinculación real, con qué grado y por qué razón. Este diseño basado en agentes permite descomponer el análisis en pasos y añadir nuevas capacidades de verificación, respondiendo de forma dinámica sin rediseñar el

*pipeline*. El enfoque híbrido (recuperación semántica seguida de verificación mediante agente) equilibra eficiencia y calidad: la búsqueda vectorial reduce el espacio de comparación a un número manejable de candidatos, y el agente aporta la capacidad de razonamiento necesaria para confirmar o descartar la vinculación, incluso cuando el lenguaje de la propuesta y el del texto legal difieren considerablemente.

El paso de verificación utiliza salida estructurada. En lugar de generar texto libre, el LLM debe producir un objeto con campos predefinidos: grado de incorporación, explicación razonada, score de confianza y las posiciones exactas dentro de la sección donde se localiza el fragmento vinculado. Esta estructura cumple una doble función. Por un lado, garantiza resultados consistentes y directamente almacenables en base de datos sin necesidad de un procesamiento adicional. Por otro lado, actúa como mecanismo de anclaje contra la alucinación del modelo: exigir una explicación razonada obliga al LLM a articular la cadena de razonamiento que justifica la vinculación (un proceso que mejora la precisión de la respuesta respecto a una simple clasificación numérica), y exigir las posiciones exactas del texto dentro de la sección fuerza al modelo a señalar un fragmento concreto del documento normativo, lo que facilita tanto la experiencia de usuario en la interfaz como la verificación: si las posiciones apuntan a texto irrelevante, la vinculación se puede descartar.

## ¿Por qué RAG híbrido y no CAG?

CAG (*Context-Augmented Generation*) procesaría todas las secciones del documento para cada propuesta, eliminando la dependencia de la calidad de la búsqueda semántica. Es viable para documentos pequeños (50-100 secciones), pero el coste combinatorio (propuestas × secciones) lo vuelve poco práctico o incluso inviable cuando ambos conjuntos crecen. Además, al enviar todo el documento en cada consulta, se pierde foco y aumenta el riesgo de que el LLM distribuya mal la atención (degradación en forma de U en contextos largos).

Por otro lado, RAG híbrido (búsqueda semántica + verificación LLM) escala mejor y focaliza la atención del modelo del lenguaje en los candidatos más prometedores. La brecha lingüística se mitiga con la generación del texto en lenguaje fácil. Si para alguna propuesta concreta el RAG no encuentra candidatos con confianza suficiente, se podría incluso aplicar CAG como *fallback* (mecanismo de contingencia para casos excepcionales a la opción principal).

## Interfaz de exploración y revisión

Entre el *pipeline* de procesamiento y la aplicación web se sitúa una API (implementada con FastAPI) que actúa como punto de acceso a los resultados almacenados en la base de datos. La API expone los datos necesarios para que el *frontend* pueda renderizar el documento normativo con sus vinculaciones. Esta separación entre el *pipeline* (que procesa y genera las vinculaciones) y la API (que las sirve) permite ejecutar el procesamiento de forma independiente y repetible sin afectar al servicio en producción.

El *frontend* se implementa como una aplicación web (usando Nuxt) que ofrece como valor añadido una lectura enriquecida del texto normativo. La interfaz principal muestra el documento legal artículo por artículo; al seleccionar una sección, un panel lateral presenta los insumos vinculados a ella con sus metadatos (texto, autor, tipo de autor, referencia, temática, grado de incorporación y explicación). La aplicación permite navegar tanto desde el texto normativo hacia los insumos como desde un insumo concreto hacia las secciones en las que ha sido incorporado, ofreciendo la doble perspectiva que necesitan los distintos actores: legisladores que quieren conocer qué respalda cada artículo, y ciudadanía que quiere saber qué ocurrió con su aportación. Todas las vinculaciones establecidas por el sistema están sujetas a revisión humana antes de su publicación, lo que garantiza que la herramienta funcione como un asistente de trazabilidad y no como un sistema de decisión autónomo.

## Deduplicación de propuestas

En los procesos participativos es frecuente recibir propuestas significativamente similares presentadas por diferentes actores. Durante el proceso de ingesta, el sistema podría incorporar un mecanismo de deduplicación o agrupación basado en similitud semántica que relacione automáticamente las propuestas equivalentes. Esto permitiría evitar el ruido visual en los resultados, así como identificar propuestas que generan mucho interés o consenso.

MAYO  
2026

**D** DERECHOS DIGITALES **POLITICAL WATCH**



Financiado por  
la Unión Europea  
NextGenerationEU



GOBIERNO  
DE ESPAÑA

MINISTERIO  
PARA LA TRANSFORMACIÓN DIGITAL  
Y DE LA FUNCIÓN PÚBLICA

SECRETARÍA DE ESTADO  
DE DIGITALIZACIÓN  
E INTELIGENCIA ARTIFICIAL

red.es



Plan de  
Recuperación,  
Transformación  
y Resiliencia