



Clasificación automática de categorías constitucionales del proceso constitucional chileno de 2016, usando IA generativa

Álvaro Sandoval Pizarro

Biblioteca del Congreso Nacional
Victoria s/n, Edificio del Congreso Nacional, Valparaíso

`alsandoval@bcn.cl`

Profesor guía: Carlos Valle Vidal, PhD

`carlos.valle@pucv.cl`

Resumen:

La etapa participativa del Proceso Constituyente Chileno de 2016 (segundo gobierno de M. Bachelet), culminó de manera exitosa. Si bien la metodología original facilitaba su sistematización, dejó un desafío: la categoría "otros" concentró el mayor porcentaje de las respuestas. En 2020, la Biblioteca del Congreso Nacional aplicó una combinación de clasificación manual y Machine Learning (ML), logrando reclasificar más del 93% de dichas respuestas e incorporando nuevas categorías emergentes.

Esta investigación evalúa si la Inteligencia Artificial Generativa (IAG) puede clasificar masivamente estas respuestas con resultados comparables o superiores al ML clásico. Se desarrolló una aplicación que clasificó más de 36.000 respuestas "otros" en las meta-categorías de valores, derechos, deberes e instituciones.

Los resultados muestran una coincidencia exacta del 58,1% entre ML e IAG en la clasificación del subconjunto "otros", y una cobertura superior al 99% por parte de la IAG sobre dicho subconjunto. Ambos métodos alcanzan coberturas comparables sobre el corpus total, con diferencias menores atribuibles a la depuración previa del corpus ML. El análisis comparativo indica que ambos métodos preservan las prioridades centrales del debate ciudadano, aunque la IAG incrementa moderadamente la diversidad distributiva y amplía la visibilización de categorías ciudadanas emergentes. La validación experta no evidenció un deterioro sistemático del desempeño de la IAG; la baja concordancia inter-evaluadora (Kappa de Cohen promedio $\approx 0,09$; Kappa de Fleiss = 0,059) refleja la naturaleza interpretativa de la clasificación temática constitucional, más que una limitación del modelo.

Se concluye que la IAG constituye una herramienta válida y escalable para la sistematización de participación ciudadana en procesos deliberativos complejos, confirmando así la relevancia y validez de los nuevos conceptos aportados por la ciudadanía al proceso de redacción constitucional.

Palabras Clave: proceso constituyente, constitución política, participación ciudadana, clasificación de texto, inteligencia artificial generativa.



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☐ Memoria o trabajo de título ☒ Tesis de Postgrado

Título del trabajo: Clasificación automática de categorías constitucionales del proceso constitucional chileno de 2016, usando IA generativa

Nombre del candidato(a): Alvaro Sandoval Pizarro

Carrera / Grado: Magister en Tecnologías de Información

Campus: San Joaquín **Departamento:** Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Carlos Valle Vidal, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente

DEJO CONSTANCIA que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 15 de junio de 2026

Firma: 

Estudiante o Candidato(a):

Fecha: 15 de junio de 2026

Firma: 

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.



1 Introducción

1.1 Objetivos

1.1.1 Objetivo general

Validar que el uso de Inteligencia Artificial Generativa para clasificar conceptos constitucionales en el contexto de la participación ciudadana en un proceso constituyente permite obtener resultados similares o mejores que utilizando Machine Learning clásico.

1.1.2 Objetivos específicos

- Desarrollar una aplicación basada en IAG capaz de clasificar automáticamente respuestas "otros" del proceso constituyente 2016
- Comparar cuantitativamente los resultados de clasificación IAG vs ML en términos de cobertura y precisión
- Evaluar cualitativamente la calidad semántica de las clasificaciones generadas por IAG
- Determinar la relevancia de los conceptos constitucionales emergentes aportados por la ciudadanía

1.2 Contexto, motivación y problemática

La Constitución Política de la República de Chile, promulgada en 1980, ha experimentado numerosas modificaciones desde su entrada en vigor [1]. Sin embargo, a pesar de las reformas, su origen, ocurrido durante un período no democrático, ha mantenido abierto el debate respecto de su legitimidad y adecuación a las transformaciones políticas, sociales y culturales del país [2].

Durante el segundo mandato de la presidenta Michelle Bachelet (2014–2018), y en el marco de su programa de gobierno [3], se impulsó el “Proceso Constituyente Abierto a la Ciudadanía” con el objetivo de elaborar una nueva carta fundamental a partir de un amplio proceso participativo. Dicho proceso culminó con la presentación de un proyecto de ley de Reforma Constitucional, ingresado al Congreso Nacional el 6 de marzo de 2018 [4], pocos días antes del término del mandato presidencial. No obstante, la tramitación del texto no avanzó durante el gobierno siguiente, encabezado por el presidente Sebastián Piñera.

En octubre de 2019, el país vivió el denominado “estallido social” [5], una masiva movilización ciudadana que expresó múltiples demandas sociales y políticas. Este acontecimiento dio origen a un nuevo proceso constituyente, esta vez impulsado desde la ciudadanía. Como resultado, el 15 de noviembre de 2019 se firmó el “Acuerdo por la Paz Social y la Nueva Constitución” [6], que definió el mecanismo para redactar una nueva constitución mediante una Convención Constitucional cuyos integrantes serían elegidos por voto popular.

Previo a este segundo proceso, durante 2019 y 2020, se realizaron cabildos autoconvocados y diálogos ciudadanos organizados por el gobierno, con el propósito de recoger las visiones y propuestas de la población. En abril de 2021, se publicó un documento con la sistematización de 1.233 cabildos autoconvocados, elaborado por un consorcio de 12 universidades y la organización Unidad Social [7].

En paralelo, durante 2020 y 2021, la Biblioteca del Congreso Nacional de Chile (BCN) desarrolló y publicó diversos documentos e iniciativas de apoyo al proceso constituyente [8], [9]. Entre ellas, destaca la visualización de los resultados de los cabildos del proceso constituyente de 2016 [10], una herramienta digital entregada como insumo al trabajo de la Convención Constitucional [11], [12].

1.3 Definición del problema

La etapa participativa del “Proceso Constituyente Abierto a la Ciudadanía”, realizado entre el 23 de abril y el 6 de agosto de 2016, tuvo 3 niveles de participación ciudadana:

- Nivel local
 - Consulta individual
 - Encuentros locales
- Nivel provincial: cabildos provinciales
- Nivel regional: cabildos regionales

Las personas participantes podían expresar sus preferencias a las preguntas orientadoras de la metodología de la etapa participativa: qué valores y principios, qué derechos, qué deberes y responsabilidades y qué instituciones del Estado consideran más importantes y, por lo tanto, deberían ser incluidas en la nueva constitución [13].

Para dar respuesta a tales preguntas, se incluyeron listados de referencia de conceptos constitucionales, los que orientaron el debate. Se utilizaron los mismos listados de referencia para cada nivel de participación. Para la elaboración de tales listados, se utilizó el análisis comparado de 16 constituciones del mundo (Alemania, Argentina, Brasil, Canadá, Chile, Colombia, España, Estados Unidos, Finlandia, Francia, Holanda, Italia, México, Suecia, Suiza, Uruguay), que realizó la división de Estudios de la SEGPRES. Se incluyó también la opción de generar nuevos aportes, en el espacio denominado “otros”, según detalla la guía metodológica del proceso [13]. En la fase local, se podía escoger un máximo de 7 conceptos constitucionales por cada pregunta, y un máximo de 14 en las fases provincial y regional.

Los datos de la consulta individual se registraron directamente en la plataforma web disponible para tal efecto y las actas de cabildos fueron digitalizadas y subidas a la misma plataforma. Todos los datos fueron anonimizados con posterioridad y quedaron disponibles como datos abiertos [14].

Al analizar tal conjunto de datos, se detectó que la categoría “otros” tuvo la mayor frecuencia en las cuatro preguntas, en todos los cabildos (local, provincial y regional). Este hallazgo orientó el trabajo de la BCN, de clasificación semiautomática, realizado para comprender los aportes ciudadanos, logrando reclasificar el 96% de todas las respuestas “otros” de cabildos [10].

1.4 Breve descripción de la solución

El enfoque de la presente tesina fue no utilizar a un experto humano para clasificar los textos “otros”, sino que construir una aplicación que automatice la clasificación utilizando como taxonomía el conjunto de conceptos constitucionales predefinidos en la metodología de la etapa participativa, enriquecida con los nuevos conceptos constitucionales surgidos de la clasificación manual.

La aplicación desarrollada utiliza la API de OpenAI y el modelo gpt-4.1-mini y permite categorizar las respuestas de texto libre, junto al fundamento, también texto libre, si existe. La aplicación realiza procesamiento paralelo para optimizar los tiempos, así como trazabilidad de cada clasificación.

Además, se incorporaron mecanismos de registro de errores, visualización de métricas de rendimiento y la capacidad de reanudar el trabajo sin perder progreso. Esta solución está diseñada para ser escalable, reproducible y adaptable a distintos conjuntos de categorías temáticas dentro del proceso constituyente.

1.5 Hipótesis y metodología aplicada

La hipótesis de esta investigación sostiene que la Inteligencia Artificial Generativa (IAG) puede clasificar automáticamente las respuestas “otros” del proceso constituyente 2016 con cobertura y calidad semántica

comparable o superior al enfoque basado en Machine Learning (ML) clásico, preservando las prioridades centrales del debate ciudadano identificadas en el corpus original.

El corpus total considerado corresponde a 264.383 respuestas ciudadanas recopiladas en distintos tipos de cabildos. De ellas, 36.322 fueron clasificadas como "otros", es decir, no utilizaron ninguna de las categorías del listado oficial de conceptos constitucionales definido por la metodología participativa. Estas respuestas incluyeron campos de texto libre para describir el contenido y su fundamento argumentativo.

Para validar la hipótesis se desarrolló una aplicación basada en la API de OpenAI (modelo gpt-4.1-mini) que implementa clasificación automática con procesamiento paralelo, trazabilidad y recuperación ante interrupciones. La taxonomía utilizada combina las categorías originales de la metodología participativa con los conceptos emergentes identificados en el análisis manual previo realizado por la Biblioteca del Congreso Nacional.

La validación se estructuró en tres dimensiones complementarias:

- (a) Cobertura: proporción de respuestas "otros" efectivamente clasificadas por cada método.
- (b) Consistencia temática: análisis de coincidencia exacta entre clasificaciones IAG y ML, preservación del núcleo temático dominante (categorías más frecuentes), y métricas de diversidad distributiva (índices de Shannon y Simpson).
- (c) Calidad semántica: validación cualitativa mediante evaluación experta independiente. Se seleccionó una muestra aleatoria de 200 casos donde ML e IAG difirieron en su clasificación, priorizando aquellos con mayor extensión de texto en el fundamento. Tres expertos de distintas disciplinas (un sociólogo con experiencia en el proceso de clasificación previo, un abogado constitucionalista y una abogada) evaluaron cada caso según una escala ordinal que permitía indicar cuál método ofrecía mejor clasificación, si ambas eran equivalentes, incorrectas o no evaluables.

1.6 Breve descripción de la estructura del informe

El capítulo N°1 corresponde a la introducción.

El capítulo N°2 contiene la definición del marco teórico y el estado del arte.

El capítulo N°3 describe cómo se abordó y desarrolló del proyecto, que considera una descripción del corpus utilizado, hallazgos y decisiones surgidas a partir de tales hallazgos, definición de validación de la propuesta y análisis de resultados.

Se finaliza con el capítulo N°4 con las conclusiones del trabajo.

El cuerpo del informe se cierra con la lista de referencias.

2 Marco teórico y estado del arte

2.1 Marco teórico

- ¿Qué es Constitución?

García de Enterría [15] define la Constitución como la norma fundamental del Estado, encargada de organizar los poderes públicos, establecer los derechos de las personas y garantizar la supremacía del orden constitucional.

El Diccionario Panhispánico del español jurídico establece que la Constitución [16] es la ley suprema de un Estado, que regula la organización de los poderes públicos y establece las garantías de los derechos; ostenta una

posición de jerarquía sobre todas las demás normas que integran el ordenamiento jurídico y solo puede ser reformada siguiendo los procedimientos en ella establecidos.

- ¿Qué es un proceso de reforma constitucional?

La Constitución Política de la República de Chile, en su texto refundido vigente, disponible en BCN.cl [17], establece un procedimiento específico para introducir reformas a su contenido. Este procedimiento se encuentra regulado en el Capítulo XV, artículos 127 al 129, bajo el subtítulo “Reforma de la Constitución”.

- ¿Qué es un proceso constituyente?

Un estudio comparado de PNUD [18] define el proceso constituyente, como “el conjunto de actos que conducen a la creación de una nueva constitución”, de acuerdo con la definición de Pisarello [19], que lo describe como “todo aquel conjunto de hechos, actos e instituciones que intervienen y dan origen a un nuevo orden constitucional.”.

Para el caso chileno, al momento del Proceso Constituyente Abierto a la Ciudadanía de 2016, en la Constitución solo estaba disponible el mecanismo de reforma constitucional [20].

- La participación ciudadana en el Proceso Constituyente

El Proceso Constituyente realizado en Chile el 2016 es definido como “el mecanismo de cambio constitucional que incluye la discusión pública con los ciudadanos y ciudadanas sobre los asuntos constitucionales, con el objetivo de permitir el diálogo y la convergencia para la construcción de acuerdos entre las personas sobre la Constitución. De la misma manera, se registran los acuerdos parciales o desacuerdos entre ellos” [13].

2.2 Estado del arte

- Los resultados de la etapa participativa

La sistematización de los resultados de la etapa participativa fue insumo utilizado para la redacción de la nueva constitución.

El informe final de síntesis de resultados cuantitativos de la etapa participativa da cuenta de la cantidad oficial de participantes, 218.689 personas, y muestra también un resumen de las 14 menciones más significativas, sin compararlas, para las 4 preguntas y para todos los niveles de participación [21]. No incluye análisis de respuestas “otros”.

Figueroa y Jordán, quienes formaron parte del equipo de gobierno que diseñó la metodología de convergencia deliberativa, y fueron nombrados por la presidenta Bachelet para fiscalizar el fiel cumplimiento de ésta, destacan y comparan las 7 primeras menciones en las 4 instancias, para las 4 preguntas conductoras del proceso [22]. Se puede apreciar que tales conceptos constitucionales mencionados se mantienen de manera similar entre las 4 instancias de participación ciudadana. Más adelante, en análisis propio, se muestra que tales resultados tampoco incluyen las respuestas “otros”.

Figueroa y Jordán también destacan el gran nivel participativo del proceso chileno, “uno de los mayores niveles de participación ciudadana a nivel mundial”. Concluyen que el proceso constituyente quedó suspendido, y en la práctica, no finalizó, debido a la “ausencia de acuerdo político amplio por el cambio constitucional” [22].

2.3 Trabajos relacionados

La investigación de 2017 sobre minería de argumentos en el Proceso Constituyente chileno de 2016 [23] presenta y analiza el corpus de más de 200.000 argumentos políticos en español, obtenidos a partir de los resultados de la instancia local de la etapa participativa del proceso constituyente abierto a la ciudadanía. El estudio destaca que el corpus es uno de los más grandes en idioma español. Concluye que es posible procesar de manera automática grandes volúmenes de datos generados en procesos participativos mediante técnicas de aprendizaje. Si bien en la publicación no se menciona de manera explícita el análisis de los aportes ciudadanos, al analizar el corpus por ellos publicado, es evidente que sí lo realizaron. Es destacable también que los resultados de la investigación fueron utilizados en la elaboración del Informe Final sobre el proceso de participación y diálogos constitucionales [24].

El trabajo realizado en 2023 por un equipo de investigadores de la Universidad de Texas y el CCP (Comparative Constitutions Project) [25] propone una metodología deductiva para analizar datos masivos y no estructurados de consultas públicas constitucionales, aplicada en el caso chileno de 2016. Considera como corpus los resultados de los cabildos de todas las fases, es decir, excluyendo la consulta individual, lo que representa más de 265.000 respuestas. Utilizaron como referencia la ontología de 334 conceptos constitucionales del CCP, y aplicaron técnicas de procesamiento de lenguaje natural, para medir la similitud semántica entre las respuestas ciudadanas y los tópicos de la ontología, demostrando así que esta metodología es escalable y que las consultas ciudadanas son herramientas poderosas para identificar los temas de mayor preocupación para la ciudadanía y refuerza la legitimidad de las consultas públicas como insumo para el diseño constitucional moderno.

La clasificación automática de textos legislativos y de política pública mediante IAG ha sido objeto de creciente investigación en los últimos años. Chalkidis et al. (2021) [26] introdujeron MultiEURLEX, un conjunto de datos de referencia público para la clasificación temática automática de documentos legales mediante modelos de lenguaje preentrenados, utilizando una taxonomía predefinida como esquema de clasificación. El dataset comprende 65.000 leyes de la Unión Europea, clasificadas según la taxonomía EuroVoc, y abarca 23 idiomas. El trabajo provee un conjunto de datos de referencia para evaluar modelos de lenguaje preentrenados en la tarea de clasificación automática de textos legales.

Gunes y Florczak (2023, 2025) [27] utilizaron los modelos GPT-3.5 y GPT-4 de OpenAI para clasificar proyectos de ley y audiencias del Congreso de Estados Unidos en 21 categorías temáticas del Comparative Agendas Project. En la versión extendida del trabajo, publicada en el Journal of Computational Social Science, ampliaron la comparación a seis modelos de OpenAI y Google, obteniendo resultados que varían considerablemente según el escenario y el nivel de intervención humana requerido. Los autores concluyen que la dependencia exclusiva de LLMs de instrucción resulta insuficiente, pero que la combinación de modelos con supervisión humana moderada puede rivalizar con algoritmos entrenados específicamente para la tarea.

Estos antecedentes son relevantes para la presente investigación en cuanto demuestran que los LLMs, incluidos los modelos de la familia GPT, son capaces de clasificar textos con contenido jurídico y político en esquemas taxonómicos predefinidos, con resultados comparables a enfoques supervisados clásicos, en especial cuando se dispone de una ontología de referencia bien estructurada.

3 Desarrollo

3.1 Descripción del corpus

Los resultados de la etapa participativa del proceso constituyente de 2016 están publicados en formato abierto [14]. El conjunto de datos está compuesto por:

- 4 archivos con los resultados de las 4 instancias de participación: individual, local, provincial y regional

- 3 archivos con los perfiles de los participantes de las 3 instancias de cabildos. El archivo con los resultados individuales incluye también el perfil de cada persona que respondió

Los resultados y perfiles están anonimizados y salvo el caso de las respuestas individuales, resultados y perfiles no están relacionados, de esta manera, no es posible responder preguntas como ¿cuáles son las preferencias para determinado grupo etario o género. Los resultados de las 4 instancias están asociados a una comuna, por lo que todos los datos se pueden consultar por ubicación, incluso aquellos encuentros locales realizados en el extranjero.

En la siguiente tabla se muestra un resumen de las 4 instancias:

Instancia	#participantes	#personas/cabildo/mesa	#cabildos	#mesas
Individual	90.804	No aplica	No aplica	No aplica
Local	106.412	13,1	8.113	No aplica
Provincial	12.852	19,6	71	656
Regional	8.621	17,6	15	491
Total	218.689			9.260

Tabla 1. Resumen de participación en las distintas instancias.

Existen diferencias significativas entre la instancia individual y el resto de las instancias. Las más relevantes son:

- La consulta individual se realizó a través de una encuesta en línea, y debido a su naturaleza personal, no existe debate ni acuerdo, aunque sí hay argumentación, como en las 4 instancias
- Las instancias local, provincial y regional se organizaron como cabildos, hubo debate y niveles de acuerdo por cada respuesta, las que fueron registradas en actas, las que fueron digitalizadas y subidas a la plataforma digital disponible para el proceso
- Las instancias provincial y regional se realizaron en establecimientos de educación a lo largo del país, con excepción del cabildo regional de la región metropolitana, que se realizó en el Centro Cultural Estación Mapocho [28]. Cada mesa representa un cabildo o grupo de personas que debatieron y registraron sus respuestas

Los niveles de acuerdo definido por la metodología eran 3:

- Acuerdo total, para mayoría absoluta
- Acuerdo parcial, para mayoría simple
- Desacuerdo, cuando no hay mayoría

Respecto de las preguntas, hay 2 diferencias significativas:

- Las instancias individual y local tenían una quinta pregunta: "Este espacio está destinado para realizar un aporte a la memoria histórica de este proceso. Podrá ser completado con ideas, propuestas y visiones sobre nuestro país para las futuras generaciones". Era un texto libre, la gran mayoría lo respondió, lo que representa un aporte ciudadano muy relevante, aunque no estructurado
- Las instancias provincial y regional utilizaron como insumo los resultados de la etapa anterior, por tal razón en la metodología estas instancias podían escoger entre 14 respuestas por cada una de las 4 preguntas, tomando como base las 7 más votadas de la instancia anterior. En cambio, las instancias individual y local solo podían escoger un máximo de 7 por cada pregunta

La instancia local, donde se realizaron múltiples cabildos los que se denominaron Encuentros Local Autoconvocados (ELA), tuvo 2 características adicionales:

- Se podía realizar en el extranjero. Se registraron 148 ELA fuera de Chile
- Podía ser registrado como un encuentro de pueblos originarios. Se realizaron 142 ELA de organizaciones de pueblos originarios

3.2 Definición del corpus de estudio

3.2.1 Hallazgo de sesgo en instancia individual

El conjunto de datos de la instancia individual adolece de un elemento clave: no tuvo debate, porque en esencia es una encuesta, no un cabildo.

Lo anterior, podría explicar por qué al analizar las respuestas de la instancia individual, se aprecia de manera muy directa que más del 35% de las respuestas a la quinta pregunta son idénticas. Si consideramos que la metodología no permitía la duplicación de registro individual, tal comportamiento no puede atribuirse a un error o a la casualidad, sino más bien a coordinación entre individuos.

Este es el texto incluido en 26.208 de las respuestas a la quinta pregunta, de un total de 74.652 respuestas válidas:

“Solicito mejorar la Constitución política de Chile en los siguientes 2 artículos que incorporan 4 propuestas claras

1) Agregar el siguiente inciso final al N°24 del artículo 19:

Las personas que coticen o hayan cotizado en cualquier institución del sistema previsional, serán titulares del derecho de dominio sobre sus cotizaciones y ahorros previsionales.

2) Agregar al N°18 del artículo 19, como incisos cuarto, quinto y sexto, los siguientes incisos:

Las prestaciones básicas uniformes se financiarán a través de un Préstamo Estatal de Reconocimiento, a través del cual, el jubilado podrá usufructuar exclusivamente de su rentabilidad, pero a su fallecimiento se restituirá automáticamente el Préstamo hacia el Estado.

Las prestaciones básicas uniformes tendrán un valor no inferior al 70% del sueldo mínimo vigente en el país.

Los afiliados al sistema de pensiones tienen el derecho a acceder al 100% del detalle de su portafolio de inversiones, para poder auditar los valores cuota de los fondos previsionales.”

3.2.2 Comparación de frecuencias de “otros”

Al comparar las respuestas “otros” entre las 4 instancias, se puede apreciar similitudes entre los 3 cabildos, pero un comportamiento muy diferente con la instancia individual. Las respuestas “otros” para la instancia individual, está en el lugar 30° de mayor frecuencia, entre 113 opciones.

Sin embargo, para la instancia local, se encuentra en primer lugar, siendo la respuesta de mayor frecuencia. Lo mismo ocurre con las instancias provincial y regional, donde incluso hay mayor distancia entre la primera frecuencia y la segunda, como se aprecia en la tabla comparativa y los gráficos a continuación:

Instancia	Total respuestas	Respuestas “otros”	Porcentaje “otros”	Ranking
Individual	2.073.163	23.903	1,15%	30° de 113
Local	210.433	23.491	11,16%	1° de 113
Provincial	29.946	6.657	22,23%	1° de 113
Regional	23.708	5.878	24,79%	1° de 108

Tabla 2. Resumen de las 4 instancias para respuestas “otros”.

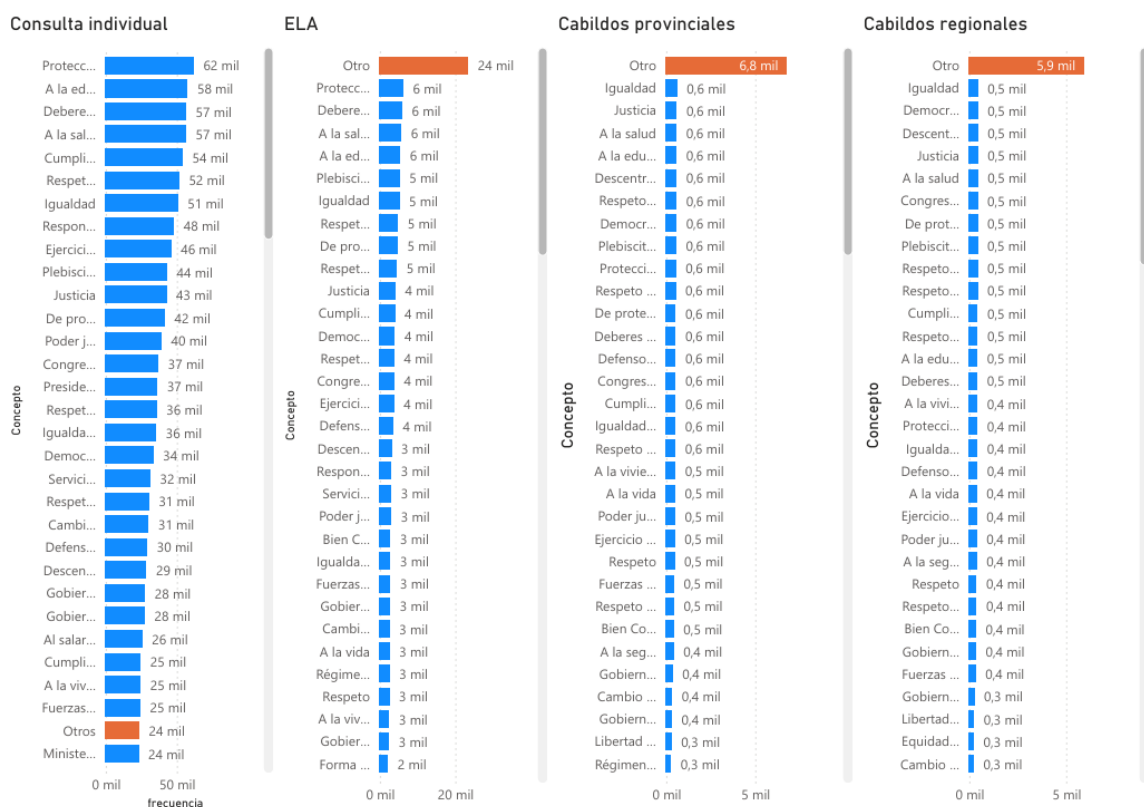


Figura 1. Gráfico de frecuencias en las 4 instancias.

Las respuestas “otros” representan nuevas categorías, no incluidas en el instrumento. Es destacable que, para el caso individual, las personas en su gran mayoría no consideraron necesario aportar con nuevas categorías, sin embargo, en todas las instancias donde existió debate, las respuestas “otros” fue mayoría, dando cuenta de la necesidad de incluir categorías que la metodología no consideró.

Debido a lo anterior, para este trabajo de investigación solo consideré las instancias local, provincial y regional, buscando categorizar las respuestas otros.

3.2.3 Preparación diferencial de los datos según método

Este trabajo aplica un modelo de Inteligencia Artificial Generativa (IAG) para clasificar las respuestas "otros" del corpus, y compara los resultados con los obtenidos mediante un proceso de Machine Learning clásico (ML) realizado en 2020. Ambos métodos operan sobre el mismo dataset de origen, las 264.383 respuestas publicadas por el gobierno en 2017 como datos abiertos, y aplican la clasificación exclusivamente sobre el subconjunto de respuestas originalmente etiquetadas como "otros". Sin embargo, difieren en el grado de preparación previa del corpus y en la estructura de sus asignaciones, lo que condiciona el universo efectivo de análisis en cada caso.

El proceso con ML utilizó un corpus depurado mediante un trabajo manual exhaustivo realizado en 2020. La depuración excluyó 267 respuestas correspondientes a texto basura, vacío, semánticamente irrelevante o asociado a referencias procedimentales del cabildo sin contenido sustantivo (por ejemplo: "no hubo acuerdo", "concepto no trabajado"). Adicionalmente, se eliminaron 23.910 fundamentos que contenían basura o información semánticamente irrelevante, conservando en estos casos la respuesta asociada. Se realizó también

una fusión de registros duplicados, en aquellos casos en que una misma respuesta contaba con más de un fundamento registrado en el acta del cabildo. El resultado fue un corpus depurado de 264.087 respuestas, de las cuales 36.026 correspondían a la categoría "otros". Este proceso fue realizado íntegramente por el autor.

El conjunto de entrenamiento fue construido por un sociólogo con experiencia en análisis político, quien recibió la instrucción de etiquetar manualmente aproximadamente el 20% del corpus depurado. El corpus fue dividido en cuatro subconjuntos según la meta-categoría temática (Deberes y Responsabilidades, Derechos, Instituciones, y Valores y Principios), trabajando cada subconjunto de manera independiente. El texto de entrada para cada registro fue la concatenación de la respuesta "otro" y su fundamento asociado. El preprocesamiento incluyó eliminación de stopwords en castellano y stemming, seguido de vectorización TF-IDF. La clasificación automática del corpus restante se realizó mediante similitud de coseno entre el vector de cada respuesta y los vectores del conjunto de entrenamiento, asignando la categoría del registro más similar. Los registros cuya similitud máxima no alcanzó el umbral de 0,5 fueron dejados sin clasificar.

En la práctica, la cobertura manual superó la instrucción original, alcanzando el 37,2% del corpus, con variaciones significativas entre meta-categorías: mientras Deberes y Responsabilidades y Derechos se mantuvieron cercanas al 20% original (17,4% y 17,2% respectivamente), Instituciones alcanzó el 35,8% y Valores y Principios llegó al 90,2%. El experto tuvo además la facultad de crear nuevas categorías a partir de las respuestas encontradas, aplicando como criterio de validación que cada categoría candidata contara con al menos 50 respuestas distintas asociadas. Este proceso fue iterativo: las categorías candidatas fueron confirmadas o descartadas a medida que avanzaba el etiquetado. El resultado fue que el 54,1% de las asignaciones totales correspondió a categorías nuevas o emergentes, no contempladas en la taxonomía original, lo que refleja la riqueza semántica del corpus y la capacidad del proceso para capturar conceptos constitucionales no anticipados.

Una característica relevante del método ML es que puede asignar más de una categoría por respuesta, heredando el criterio del experto humano que construyó el conjunto de entrenamiento, quien asignó hasta seis categorías en los casos más complejos. Las 36.026 respuestas "otros" del corpus ML generaron 35.177 asignaciones de categoría, resultado que se integra al universo total de 263.238 asignaciones utilizadas en el análisis comparativo.

El proceso con IAG, en cambio, utilizó el corpus en bruto sin intervención previa, preservando las 264.383 respuestas originales, de las cuales 36.322 correspondían a la categoría "otros". El prompt instruyó al modelo a asignar una única categoría por respuesta, y a evaluar si el fundamento aportaba o no información útil para la clasificación, registrando este juicio en el campo fundamento aporta del output. Este enfoque es funcionalmente equivalente a la depuración manual de fundamentos realizada para ML, pero implementado de manera automatizada en tiempo de clasificación. Las 36.322 respuestas generaron 36.080 asignaciones, que se integran al universo total de 264.149 asignaciones del análisis comparativo.

En ambos casos, el universo de análisis comparativo corresponde a asignaciones de categoría, excluyendo los registros sin clasificar. La diferencia entre ambos universos (263.238 en ML versus 264.149 en IAG) refleja el efecto combinado de la depuración previa del corpus ML y la mayor cobertura del proceso IAG sobre el corpus completo.

Dimensión	ML	IAG
Dataset de origen	Dataset gobierno 2017	Dataset gobierno 2017
Preparación del corpus	Depuración manual (2020)	Sin depuración previa
Conjunto de entrenamiento	~37,2% etiquetado manual	No aplica
Preprocesamiento	Stopwords + stemming + TF-IDF	Prompt con taxonomía explícita
Método de clasificación	Similitud de coseno (umbral $\geq 0,5$)	Modelo generativo (gpt-4.1-mini)
Criterio nuevas categorías	Mínimo 50 respuestas distintas	No aplica
Respuestas totales	264.087	264.383
Respuestas "otros" clasificadas	36.026	36.322

Categorías por respuesta	1 a 6 (hereda criterio experto)	1 (por diseño del prompt)
Asignaciones totales ("otros")	35.177	36.080
Universo análisis comparativo	263.238 asignaciones	264.149 asignaciones
Requests API	N/A	106.382 (2 ejecuciones)
Tokens procesados	N/A	77.345.505 (2 ejecuciones)
Costo directo	Sin costo; clasificación experta previa	USD 62 (2 ejecuciones)

Tabla 3. Comparación de parámetros del corpus y proceso de clasificación entre ML e IAG.

- Detalle de ejecuciones IAG

El proceso IAG requirió dos ejecuciones completas sobre las 36.322 respuestas "otros". La primera (junio 2025) fue descartada por errores en la configuración del prompt que comprometían la trazabilidad de los resultados, y utilizó inicialmente el modelo gpt-4o y luego el gpt-4o-mini. La segunda (diciembre 2025), con el prompt consolidado, utilizó el modelo gpt-4.1-mini (disponible en esa fecha) y produjo los datos definitivos. La diferencia de costo entre ejecuciones (USD 44 vs USD 18) refleja tanto la selección del modelo como el proceso iterativo de ajuste del prompt de la primera ejecución.

3.3 Especificación de la solución

La Figura 3 ilustra el flujo completo del proceso de clasificación desarrollado en este capítulo, mostrando en paralelo las etapas de cada método desde el corpus de origen hasta el universo de análisis comparativo. Las subsecciones siguientes detallan cada componente del proceso IAG.

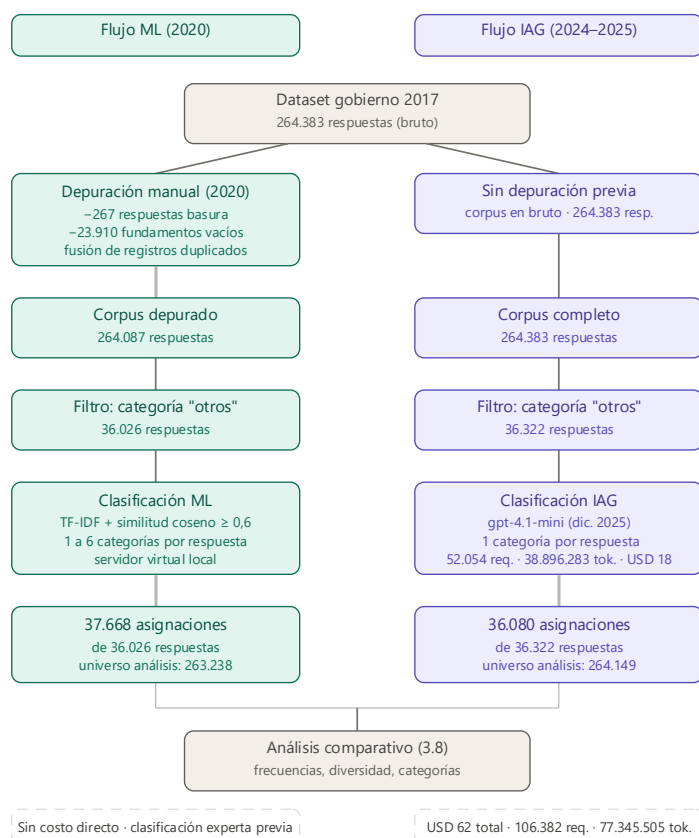


Figura 3. Flujo de proceso de clasificación ML vs IAG.

3.3.1 Descripción general de la solución

En el contexto de la presente tesis, se desarrolló un script en Python que permite la clasificación automática de respuestas abiertas del tipo “otros”, recogidas durante la etapa participativa del proceso constituyente impulsado por el gobierno de Michelle Bachelet.

El objetivo del script es asignar a cada respuesta una categoría constitucional significativa, de manera automatizada, utilizando para ello modelos generativos de lenguaje provistos por la API de OpenAI, GPT-4.1-mini. La respuesta está conformada por el nombre de la categoría y el fundamento.

3.3.2 Taxonomía utilizada

Un aspecto fundamental del script fue el uso de una taxonomía temática de conceptos constitucionales, estructurada en cuatro conjuntos independientes, según las preguntas del cuestionario ciudadano:

- Valores y principios
- Derechos
- Deberes y responsabilidades
- Instituciones del Estado

La base inicial de estas taxonomías provino del conjunto de conceptos incluidos en la metodología oficial del proceso participativo de 2016, definidos por el gobierno de Michelle Bachelet para estructurar la deliberación ciudadana.

Posteriormente, estas taxonomías fueron enriquecidas con nuevos conceptos identificados por la Biblioteca del Congreso Nacional (BCN) durante el trabajo realizado en 2020, mediante una combinación de clasificación manual y técnicas de aprendizaje de máquina (ML) sobre el mismo corpus de respuestas. Esto permitió extender el alcance del clasificador a categorías emergentes que no estaban previstas inicialmente, aumentando su cobertura y profundidad analítica.

A continuación, un resumen de los conceptos constitucionales:

Categoría	# conceptos originales	# conceptos nuevos	Ejemplos de conceptos originales	Ejemplos de conceptos nuevos
Valores y principios	37	29	Democracia Igualdad Descentralización	Respeto derechos humanos Cohesión social (equidad) Libertad
Derechos	44	25	A la salud A la educación A la vivienda digna	Al agua (recursos hídricos) De propiedad sobre los recursos naturales A una muerte digna
Deberes y responsabilidades	12	32	Respeto de derechos de otros De protección y conservación de patrimonio histórico y cultural	De participación ciudadana (voto obligatorio) De probidad y transparencia en el Estado De actuar por el bien común
Instituciones	21	61	Plebiscitos, referendos y consultas Congreso o parlamento (estructura y funciones) Defensor del Pueblo/Ciudadano	De la sociedad civil Características del Estado Banco Central
Total	114	147		

Tabla 4. Resumen de las 4 categorías de conceptos constitucionales.

3.3.3 Descripción de la funcionalidad del script

3.3.3.1 Entrada de datos

El script recibe como entrada un archivo JSON estructurado con los siguientes campos:

- `id_cabildo`: identificador del cabildo correspondiente
- `otro`: texto libre que describe la respuesta
- `fundamento`: texto libre de justificación
- `tipo_eleccion`: tipo de proceso participativo
- `tipo_cabildo`: tipo de evento
- `pregunta`: uno de los 4 ejes temáticos mencionados

3.3.3.2 Construcción del prompt

El script genera automáticamente un prompt para el modelo de IAG con la siguiente información:

- El texto libre de la respuesta (`otro`)
- Su fundamento, si existe (`fundamento`)
- La lista de categorías disponibles en la taxonomía específica de la pregunta
- La definición de la pregunta obtenida a partir del documento metodológico del proceso constituyente

El modelo recibe además un system prompt que le indica actuar como clasificador experto y responder siempre en formato JSON con tres claves: `id_categoria`, `categoria` y `fundamento_aporta`.

La definición completa del system prompt y user prompt utilizados se encuentra disponible en el repositorio del proyecto: <https://github.com/alvaroisp/mti-cabildos2016>

3.3.3.3 Clasificación con IAG

La clasificación se realiza mediante la API de OpenAI, utilizando el método `client.chat.completions.create(...)` y el modelo `gpt-4.1-mini`. El script maneja automáticamente errores de tipo `RateLimitError`, interrupciones del sistema, y realiza reintentos exponenciales si la respuesta falla.

3.3.3.4 Paralelización

Para mejorar el rendimiento, el procesamiento se realiza en paralelo utilizando `ThreadPoolExecutor`, permitiendo escalar la clasificación a decenas de miles de respuestas en tiempos razonables. Cada hilo mantiene su propio log de trazabilidad.

3.3.3.5 Trazabilidad, métricas y salida del script

El script genera además artefactos de trazabilidad y métricas por cada ejecución: registros de respuestas y de errores del modelo y un archivo de avance temporal (`progreso_temp.json`) que permite reanudar el proceso ante interrupciones; métricas de desempeño por eje temático (respuestas procesadas, porcentaje de errores, tiempo promedio por entrada e histogramas de tiempos); y, como salida principal, el archivo `resultados_categorizados.json`, que para cada respuesta almacena la clasificación automática (`id_categoria`, `categoria`), la evaluación del fundamento (`fundamento_aporta`), los metadatos originales, la pregunta asociada y los tiempos de ejecución.

3.4 Resultado de la ejecución del script

La ejecución consideró un archivo JSON con 36.322 entradas asociadas a las respuestas “otros” de los ELA, cabildos provinciales y regionales, 9,9 horas de proceso, un promedio de 0,98 entradas/segundo, sin errores de ejecución.

Se clasificaron exitosamente 36.079 respuestas en 1 categoría, 205 no fueron clasificados y 6 fueron clasificados en más de 1 categoría. También se detectaron 57 registros aparentemente duplicados, que al deduplicar quedaron convertidos en 25 registros: 11 de ellos con datos correctos y 14 con datos no clasificables.

En resumen, el proceso de clasificación utilizando el modelo gpt-4.1-mini clasificó el 99,36% de las respuestas, y no clasificó el 0,56% de los datos.

Al comparar con el proceso anterior utilizando Machine Learning, se clasificaron 35.177 respuestas, y no se clasificaron 2.491 respuestas, lo que equivale al 93,39% y 6,61% respectivamente.

La diferencia entre totales de un proceso y otro se explica porque con Machine Learning se excluyeron manualmente respuestas no clasificables, en cambio no se excluyeron registros para IAG. Aun así, el rendimiento del proceso de clasificación con IAG fue 1 orden de magnitud mejor que el de Machine Learning, al comparar registros no clasificados: 205 versus 2.491.

3.5 Resultados de la clasificación

3.5.1 Evaluación general de resultados

El 58,1% de las clasificaciones de todo el corpus de respuestas fueron idénticas entre AIG y ML, lo que representa 20.961 respuestas con 1 categoría.

El proceso de clasificación con IAG tuvo mayor cobertura: Con IAG se clasificaron 2.473 respuestas que el proceso con ML no clasificó, lo que representa un 6,85% de todas las respuestas con 1 categoría. Por otro lado, no existen respuestas que fueron clasificadas con ML pero que no fueron clasificadas por IAG.

El 39,3% de las clasificaciones con 1 categoría del corpus de respuestas fueron distintas entre IAG y ML, lo que equivale a 14.190 respuestas.

3.6 Control de calidad y depuración del dataset

Con el objeto de asegurar la validez interna del análisis comparativo entre los distintos escenarios de clasificación, se implementó un proceso sistemático de control de calidad y depuración del dataset resultante de la clasificación con IAG. Este proceso se estructuró en tres etapas: (i) verificación de consistencia entre identificadores y categorías, (ii) detección de asignaciones de meta-categoría y (iii) recalibración del dataset para el análisis comparativo.

3.6.1 Verificación de consistencia entre identificador y categoría textual

Se realizó una revisión exhaustiva de la coherencia entre el identificador numérico de categoría (id_categoria) y el nombre textual de la categoría asignada por el modelo IAG. Esta verificación permitió detectar un total de 69 inconsistencias distribuidas de la siguiente manera:

Taxonomía	Inconsistencias	% del total	Casos resueltos
Valores y principios	24	34,8%	23 + 1 s/c

Derechos	35	50,7%	35
Deberes y responsabilidades	7	10,1%	4 + 3 s/c
Instituciones	3	4,3%	3
Total	69	100,0%	65 + 4 s/c

Tabla 5. Distribución de inconsistencias detectadas por taxonomía. “s/c”: sin clasificar tras corrección manual

Dado que los análisis de frecuencia, concentración y diversidad se realizan sobre la base del identificador numérico, estas inconsistencias podían generar distorsiones en los cálculos agregados. Se procedió a corregirlas mediante un proceso mixto: automático cuando la categoría textual era directamente identificable en la taxonomía, y manual en los casos que requerían interpretación del fundamento. De los 69 casos, 65 fueron resueltos y 4 quedaron sin clasificar, restableciendo la integridad referencial del dataset

3.6.1.1 Tipos de inconsistencias

El análisis permite identificar cuatro patrones diferenciados (Tabla 9). El detalle completo con los 69 casos se incluye en el Anexo. Entre ellos, el tipo T3 (meta-categoría) admite dos variantes de resolución —recuperable por fundamento o irrecuperable por ausencia de categoría en la taxonomía— y el tipo T4 constituye el patrón más grave, en que el modelo generó de forma autónoma tanto el nombre como el identificador de la categoría, sin correspondencia con la taxonomía entregada.

Tipo	Nº	Descripción	Resolución	Resultado
T1 Error referencial puro	37	Texto correcto en taxonomía; ID incorrecto	Automática	37 resueltos
T2 Variante textual menor	4	Texto es variante mínima de categoría real; ID correcto	Automática	4 resueltos
T3 Meta-categoría	24	Modelo usó etiqueta de taxonomía como nombre de categoría	Manual por fundamento	22 resueltos; 1 s/c; 2 s/c (categoría inexistente)
T4 Categoría e ID inventados	2	Modelo generó nombre e ID inexistentes en taxonomía	Manual por fundamento	1 resuelto; 1 s/c (ID eliminado)
Total	69			65 resueltos; 4 s/c

Tabla 6. Distribución de inconsistencias detectadas por taxonomía. “s/c”: sin clasificar tras corrección manual

3.6.2 Recalibración del dataset y efectos sobre el análisis

Durante el control de calidad se detectaron además tres casos en que la IAG devolvió como categoría la etiqueta general de la macro-categoría (por ejemplo, “Valores y Principios” o “Deberes y Responsabilidades”), inexistente en la taxonomía específica: ante fundamentos amplios o ambiguos el modelo no logró discriminar una categoría puntual y optó por la denominación del dominio. Como estas meta-categorías no forman parte de la taxonomía formal ni son comparables con la clasificación previa por ML, se reclasificaron como “Sin clasificar” para evitar la incorporación de categorías espurias. Una vez corregidas las inconsistencias y reclasificados estos casos —y tras eliminar en etapas previas los registros sin contenido sustantivo (por ejemplo, “no existe” o “sin concepto”)— se recalcularon las frecuencias agregadas y los indicadores de diversidad distributiva. Esta depuración garantiza que los tres escenarios comparados (sin clasificación automática, con ML y con IAG) se analicen sobre un conjunto de datos coherente con la taxonomía, libre de inconsistencias formales, trazable y reproducible, y conceptualmente comparable entre métodos, de modo que los resultados reflejen con mayor precisión el impacto real de cada estrategia de clasificación sobre la representación temática del corpus.

3.7 Validación cualitativa mediante evaluación experta

Con el objetivo de complementar el análisis cuantitativo de concordancia entre Machine Learning (ML) e Inteligencia Artificial Generativa (IAG), se diseñó un proceso de validación cualitativa orientado a evaluar la calidad semántica de las clasificaciones propuestas por ambos enfoques, para aquellas respuestas con distinta clasificación.

3.7.1 Diseño metodológico de la evaluación

Del total de respuestas clasificadas como “otros”, el 58,1% fue asignado a categorías idénticas por ML e IAG. En consecuencia, la validación cualitativa se concentró en el subconjunto restante, correspondiente a los casos en que ambos métodos proponían categorías distintas.

A partir de dicho conjunto se consolidaron las respuestas únicas, obteniéndose un corpus superior a seis mil registros. De este universo se seleccionó una muestra aleatoria de 200 casos, priorizando aquellos con mayor extensión en el campo “fundamento”, con el objetivo de asegurar una mejor evaluación.

Para cada caso se presentaron ambas clasificaciones (ML e IAG). Los evaluadores debían calificar cada caso según la siguiente escala:

Código	Etiqueta	Criterio de aplicación
1	ML mejor	La categoría ML captura mejor el concepto ciudadano
2	IAG mejor	La categoría IAG es más precisa o específica
3	Equivalentes	Ambas clasificaciones son válidas
4	Ambas incorrectas	Ninguna refleja adecuadamente el concepto
9	No evaluable	Texto ambiguo o con insuficiente información para emitir juicio

Tabla 7. Escala de evaluación de clasificación ML vs IAG.

La evaluación consideró cuatro criterios: precisión semántica, especificidad apropiada, coherencia con el fundamento y cobertura del concepto principal.

La encuesta fue aplicada a tres especialistas con perfiles complementarios: un sociólogo con experiencia directa en el proceso de clasificación manual realizado en 2020 con ML, un abogado constitucionalista y una abogada experta en Derecho Internacional

3.7.2 Resultados de la evaluación

Los resultados muestran variabilidad en los patrones individuales de evaluación.

En la categoría “ML mejor”, los porcentajes fueron:

- Evaluador 1 (sociólogo): 30,0%
- Evaluador 2 (abogado constitucionalista): 13,5%
- Evaluador 3 (abogada): 37,0%

En la categoría “IAG mejor”, los porcentajes fueron:

- Evaluador 1: 24,0%
- Evaluador 2: 46,5%
- Evaluador 3: 47,0%

En la categoría “Equivalentes”:

- Evaluador 1: 40,0%

- Evaluador 2: 27,0%
- Evaluador 3: 0,5%

Estos resultados evidencian diferencias en los criterios de evaluación. Mientras uno de los evaluadores presenta una mayor proporción de casos considerados equivalentes, los otros dos muestran una mayor inclinación a valorar favorablemente la clasificación generada por IAG. Asimismo, se observa que uno de los evaluadores utiliza muy escasamente la categoría “equivalentes”, lo que sugiere un criterio evaluativo más dicotómico.

Par	Acuerdo observado	Acuerdo esperado	Kappa de Cohen
Eval 1-2	33.0% (66/200)	26.5%	0.089
Eval 2-3	31.0% (62/200)	28.2%	0.039
Eval 1-3	34.0% (68/200)	23.2%	0.140
Promedio	32.7%	26.0%	0.089
Kappa de Feiss	32,7%	28,5%	0,059

Tabla 8. Concordancia inter-evaluador.

El coeficiente Kappa de Cohen [29] promedio obtenido fue cercano a 0,09, y el Kappa de Fleiss de 0,059, ambos indicando nivel "leve" de acuerdo más allá del azar.

Adicionalmente, el Kappa de Fleiss, métrica específicamente diseñada para evaluar la concordancia entre tres o más evaluadores de manera simultánea, arrojó un valor de 0,059, consistente con los valores obtenidos por pares y confirmando el nivel de acuerdo leve entre los evaluadores.

Este resultado debe interpretarse considerando la naturaleza interpretativa de la tarea evaluada. La asignación de categorías constitucionales implica juicios conceptuales que pueden variar según la formación disciplinar, la experiencia previa y el marco analítico de cada experto. En consecuencia, la baja concordancia no puede interpretarse de manera directa como indicador de baja calidad del modelo, sino como reflejo de la complejidad semántica de la tarea.

No obstante, el acuerdo observado promedio (32,7%) supera consistentemente el acuerdo esperado por azar en los tres pares evaluados (rango: 23,2% a 28,2%, promedio: 26,0%). Esta diferencia de +6,7 puntos porcentuales indica que existe estructura sistemática en las evaluaciones pese a la variabilidad individual. El Kappa de Fleiss, que evalúa la concordancia de los tres evaluadores de manera simultánea, confirma este resultado con un valor de 0,059, igualmente en el rango "leve". El análisis por criterio de mayoría simple permite caracterizar con mayor precisión estos patrones de evaluación.

3.7.3 Interpretación metodológica

Aplicando un criterio de mayoría simple (coincidencia de al menos dos evaluadores) fue posible asignar una evaluación consolidada en la mayoría de los casos. En 62 de los 200 registros (31,0%) no se alcanzó consenso inicial.

A continuación, resultados usando mayoría simple (2 de 3 evaluadores):

Código evaluación	Casos con mayoría 2/3	% de total
1 ML mejor	41	20,5%
2 IAG mejor	70	35,0%
3 Equivalentes	19	9,5%
4 Ambas incorrectas	7	3,5%
9 No evaluable	1	0,5%
Desacuerdo total	62	31,0%

Tabla 9. Distribución cuando al menos 2/3 coinciden.

A partir de los resultados obtenidos, no se observa evidencia consistente de inferioridad sistemática de la IAG respecto del ML clásico. La evaluación experta no muestra un patrón estable en que ML sea considerado superior en la mayoría de los casos por parte de los evaluadores. Tampoco permite afirmar una superioridad concluyente de la IAG. Más bien, los resultados sugieren que ambos enfoques presentan desempeños comparables en términos de calidad semántica, con variaciones asociadas al criterio individual del evaluador. La calidad semántica se entiende como la consistencia conceptual entre el contenido del texto libre y su fundamento, y la categoría constitucional asignada dentro de la taxonomía correspondiente a la pregunta evaluada (valores, derechos, deberes o instituciones).

Este hallazgo refuerza la idea de que la clasificación temática de conceptos constitucionales no constituye una tarea puramente objetiva, sino que incorpora un componente interpretativo relevante. En consecuencia, la comparación entre ML e IAG no puede reducirse exclusivamente a métricas automáticas de concordancia, sino que requiere integrar evaluación cualitativa experta para una comprensión más completa del desempeño comparativo.

3.8 Análisis comparativo de frecuencias y estructura distributiva

El presente apartado examina el impacto de la clasificación automática sobre la estructura temática del corpus, comparando tres escenarios: (i) datos originales sin reclasificación automática, (ii) datos con clasificación mediante Machine Learning clásico (ML) y (iii) datos con clasificación mediante Inteligencia Artificial Generativa (IAG). El análisis considera el dataset depurado y excluye la categoría técnica “Sin clasificar”, a fin de evaluar únicamente la distribución sustantiva de categorías.

3.8.1 Impacto global sobre la categoría residual “otros”

Del total de 264.383 respuestas registradas en el proceso constituyente 2016, 36.322 fueron originalmente clasificadas como “otros”, es decir, no asociadas a ninguna categoría predefinida.

La aplicación de IAG permitió clasificar 36.080 de estas respuestas, alcanzando una cobertura de 99,33%. En el caso de ML, la redistribución fue igualmente sustantiva, integrando prácticamente la totalidad de estas respuestas al sistema categorial.

En ambos escenarios, la categoría residual pierde relevancia analítica como contenedor indeterminado y pasa a integrarse en la estructura temática general. Este resultado constituye el primer impacto metodológico de la clasificación automática: transformar información abierta no estructurada en información categorizable y cuantificable.

En el escenario sin proceso de clasificación automática, el 13,6% de las respuestas permanecía agrupado en la categoría residual “otros”. Tras aplicar ML e IAG, este porcentaje se redujo a valores entre el 7% y 8%, respectivamente, evidenciando que ambos métodos permiten clasificar una proporción significativa de respuestas previamente no estructuradas.

Escenario	% categoría “otros”	Reducción absoluta
Sin ML ni IAG	13,6%	-
Con ML	7,2%	-6,4 pp
Con IAG	8,2%	-5,4 pp

Tabla 10. Impacto de la clasificación automática sobre la categoría “otros”.

3.8.2 Diversidad distributiva global

Excluyendo la categoría “Sin clasificar”, los totales efectivos considerados fueron 263.238 respuestas (ML) y 264.149 (IAG). Sobre estos universos se calcularon indicadores de diversidad distributiva:

Indicador	ML	IAG
Índice de Shannon (H)	4,549	4,588
Número efectivo de categorías (exp(H))	94,5	98,3
Índice de Simpson (1-D)	0,9854	0,9859

Tabla 11. Indicadores de diversidad distributiva.

Ambos métodos generan una estructura altamente diversificada. No obstante, IAG presenta una diversidad levemente mayor. El número efectivo de categorías aumenta desde 94,5 en ML a 98,3 en IAG, lo que indica una ampliación moderada del espectro conceptual observable. La diferencia no es abrupta, pero es sistemática y consistente.

3.8.3 Concentración temática y regla del 80%

Se evaluó la concentración temática mediante el número de categorías necesarias para explicar el 80% de la frecuencia acumulada por taxonomía.

Se aprecia una alta concentración estructural en ambos métodos. En la clasificación con IAG, entre 9 (Deberes) y 22 (Derechos) categorías explican el 80% de las respuestas según la taxonomía, entre el 16% y el 32% del total disponible, y en ML el patrón es equivalente (entre 9 y 23 categorías, 26% a 36%).

Los resultados muestran que el núcleo temático dominante del debate ciudadano se mantiene estable bajo ambos enfoques. Las principales categorías que concentran la mayor parte de las respuestas coinciden ampliamente entre ML e IAG. En algunos casos varía el orden de frecuencia, pero no la relevancia sustantiva de las categorías más importantes.

Este hallazgo es relevante, pues indica que la clasificación con IAG no altera las prioridades centrales del corpus.

3.8.4 Activación de categorías ciudadanas emergentes

El análisis de categorías identificadas como aportes ciudadanos (no contempladas originalmente en la metodología base) revela diferencias relevantes: el ML activó 90 categorías ciudadanas con 19.023 respuestas (7,23% del total), mientras que la IAG activó 147 categorías con 21.423 respuestas (8,11%). La IAG presenta así mayor proporción de respuestas en categorías ciudadanas, mayor número de categorías activadas y mayor dispersión en las de baja frecuencia, lo que indica que amplía el espectro conceptual emergente y distribuye con mayor granularidad las respuestas originalmente clasificadas como “otros”.

3.8.5 Concentración de categorías originales y expansión de la cola larga

- El análisis de las categorías originales (aquellas incluidas en la metodología original), sin distinción por taxonomía, y excluyendo respuestas “Sin clasificar”, muestra que la concentración en las categorías más frecuentes es similar en ambos enfoques. Las diez categorías originales más relevantes mantienen una participación acumulada comparable entre ML e IAG (25,0% y 26,0%, respectivamente).

La diferencia estructural no se observa en el núcleo dominante, sino en la expansión de la “cola larga” distributiva: es decir, en el conjunto de categorías de menor frecuencia.

En otras palabras, la IAG no reemplaza las categorías tradicionales predominantes, sino que amplía marginalmente la diversidad conceptual observable, reforzando la activación de categorías menos frecuentes y aportes ciudadanos específicos.

3.8.6 Casos de crecimiento diferencial significativo

La Tabla 18 del Anexo presenta las principales categorías ciudadanas con mayor incremento de frecuencia en el escenario IAG respecto del escenario ML; las de mayor crecimiento son institucionales, encabezadas por «Asamblea Constituyente» (de 0 registros en ML a 1.746 en IAG), seguida de «Tribunal Constitucional: reforma» (+590) y «Toda institución pública» (+258).

El crecimiento de categorías ciudadanas no es homogéneo entre las cuatro taxonomías. El caso más significativo se observa en la taxonomía Instituciones.

En particular, la categoría “Asamblea Constituyente” presenta un incremento sustantivo bajo IAG en comparación con ML. Este fenómeno sugiere que la IAG fue más efectiva en identificar y diferenciar propuestas institucionales explícitas que no fueron capturadas con la misma intensidad por el modelo de ML clásico. El surgimiento de la categoría “Asamblea Constituyente” es consistente con las conclusiones oficiales del proceso participativo 2016 incluidas en el Informe Final sobre el proceso de participación, en el capítulo 2, Bases ciudadanas para una constitución discutida en democracia, acerca de la relevancia de esta categoría a nivel ELA, Provincial y Regional [24].

Asimismo, se observaron incrementos en otras categorías específicas, lo que refuerza la hipótesis de que la IAG amplía la capacidad de detección de demandas ciudadanas emergentes.

3.8.7 Impacto metodológico de la elección del algoritmo

Los resultados anteriores permiten concluir que la clasificación automática no es metodológicamente neutra.

En términos comparativos, el ML mantiene el núcleo dominante, pero reduce la diversidad efectiva y activa menos categorías emergentes, mientras que la IAG mantiene ese núcleo, incrementa la diversidad estructural y amplía significativamente la visibilización de categorías ciudadanas nuevas, especialmente en el ámbito institucional.

La diferencia no altera sustancialmente las prioridades centrales del debate ciudadano, pero sí modifica la representación de conceptos emergentes y demandas institucionales específicas.

En consecuencia, el método de clasificación influye en el grado de granularidad conceptual observable y en la capacidad de detectar aportes ciudadanos que trascienden el marco original de la metodología del proceso constituyente.

4 Conclusiones

4.1 Conclusión general y validación de la hipótesis

La hipótesis planteada sostuvo que la Inteligencia Artificial Generativa (IAG) podía clasificar automáticamente las respuestas “otros” del proceso constituyente 2016 con cobertura y calidad semántica comparable o superior

al enfoque basado en Machine Learning clásico (ML), preservando las prioridades centrales del debate ciudadano identificadas en el corpus original, e implementarse en una aplicación escalable para el procesamiento masivo del corpus.

A partir de los resultados, la hipótesis puede considerarse confirmada en las tres dimensiones declaradas. En cobertura, la IAG clasificó más del 99% del subconjunto “otros”, reduciendo significativamente la categoría residual; ambos métodos alcanzan coberturas comparables sobre el corpus total y las diferencias menores en el porcentaje residual se explican por la depuración previa del corpus ML, no por una ventaja sustantiva de un método. En consistencia temática, el 58,1% de las clasificaciones coincidieron exactamente entre ambos métodos y los indicadores de diversidad distributiva confirman que el núcleo del debate ciudadano se preserva bajo ambos enfoques. En calidad semántica, la evaluación experta independiente no mostró evidencia de inferioridad sistemática de la IAG respecto del ML clásico. La hipótesis queda así confirmada en el plano conceptual —resultados comparables en calidad y estructura temática— y en el operativo, al demostrar viabilidad técnica y escalabilidad institucional.

Adicionalmente, el análisis distributivo reveló que la IAG incrementa moderadamente la diversidad estructural del corpus y amplía la activación de categorías ciudadanas emergentes, especialmente en el ámbito institucional, sin alterar las prioridades centrales del debate.

En consecuencia, la IAG no solo resulta comparable en desempeño, sino que presenta ventajas específicas en cobertura y granularidad conceptual.

4.2 Alcances de los resultados

Los resultados permiten establecer cuatro alcances principales. Primero, la clasificación automática transforma información abierta no estructurada en información categorizable y cuantificable, reduciendo significativamente la categoría residual “otros”. Segundo, el núcleo temático dominante del debate ciudadano permanece estable bajo ambos enfoques: las categorías de mayor frecuencia coinciden sustantivamente entre ML e IAG, lo que indica que la elección del algoritmo no altera las prioridades centrales. Tercero, la IAG incrementa la diversidad distributiva global, activando un mayor número efectivo de categorías y asignando una mayor proporción de respuestas a categorías ciudadanas emergentes.

Cuarto, el crecimiento diferencial observado se concentra especialmente en la taxonomía de Instituciones, destacando el incremento de la categoría “Asamblea Constituyente”, coherente con los antecedentes cualitativos del proceso participativo [24].

4.3 Validación cualitativa y complejidad interpretativa

La evaluación experta aplicada a una muestra de 200 casos con discrepancias entre ML e IAG evidenció una baja concordancia inter-evaluador (Kappa de Cohen promedio inferior a 0,10; Kappa de Fleiss = 0,059), lo que no debe interpretarse como falla del modelo sino como manifestación de la naturaleza interpretativa de la clasificación temática constitucional. La asignación de categorías implica juicios conceptuales que varían según la formación disciplinaria y el criterio analítico; no se observó un patrón estable de inferioridad de la IAG respecto del ML clásico, y en múltiples casos esta fue evaluada como mejor o equivalente. Estos resultados refuerzan que la clasificación temática no es una tarea puramente objetiva y que comparar métodos requiere integrar métricas cuantitativas y evaluación cualitativa.

4.4 Impacto metodológico de la elección del algoritmo

Un hallazgo central es que la clasificación automática no es metodológicamente neutra: mientras el ML mantiene mayor concentración en categorías tradicionales consolidadas, la IAG incrementa la diversidad estructural y amplía la visibilización de categorías ciudadanas emergentes, sin modificar el núcleo dominante del debate. La IAG no reemplaza las prioridades centrales, sino que aumenta la granularidad conceptual y la

capacidad de detectar propuestas institucionales específicas; por tanto, el método de clasificación influye en la representación analítica del debate ciudadano y puede impactar la interpretación de procesos participativos complejos.

4.5 Limitaciones

Los resultados deben interpretarse considerando ciertas limitaciones.

La tarea de clasificación es inherentemente interpretativa, lo que se refleja en la baja concordancia entre expertos. Esto implica que las categorías asignadas, tanto por ML como por IAG, representan una lectura posible del corpus, no una verdad objetiva, y que variaciones en el criterio clasificatorio podrían producir distribuciones distintas.

El análisis depende de la taxonomía específica utilizada en el proceso constituyente 2016. Su estructura fue definida por los diseñadores metodológicos del proceso y no está exenta de solapamientos semánticos entre categorías. Modificaciones en dicha estructura, por ejemplo, fusionando categorías semánticamente cercanas o redefiniendo sus límites, podrían alterar las distribuciones observadas y los indicadores de diversidad reportados. Esta dependencia es inherente a cualquier análisis clasificatorio basado en taxonomías predefinidas.

En cuanto al método IAG, el estudio se realizó con un único modelo (gpt-4.1-mini de OpenAI). Si bien la selección respondió a criterios de costo-eficiencia evaluados durante el proceso, no se realizó una comparación sistemática de desempeño entre modelos. No es posible determinar si modelos alternativos, ya sea de la misma familia u otros proveedores, producirían clasificaciones sustantivamente distintas, lo que constituye una limitación para la generalización del método.

El proceso requirió una etapa de control de calidad y depuración para corregir inconsistencias formales detectadas en el output IAG, lo que demuestra que incluso los sistemas automatizados necesitan supervisión metodológica y que los resultados no deben adoptarse sin validación.

El análisis presentado es de carácter agregado y transversal a las tres instancias del proceso constituyente consideradas (local, provincial y regional). No examina la evolución temática entre fases, siendo que cada instancia alimentó a la siguiente: los resultados de los ELA locales sirvieron de base para los cabildos provinciales, y estos para los cabildos regionales. El estudio de la persistencia o transformación de categorías entre fases, incluyendo el comportamiento de categorías ciudadanas emergentes en ese contexto deliberativo secuencial, queda fuera del alcance de este trabajo y se propone como línea de profundización en sección 4.7.

Finalmente, el análisis no profundiza en segmentaciones territoriales, por región, provincia o comuna, ni en dimensiones identitarias como género, pertenencia a pueblos originarios o residencia en el extranjero, que podrían revelar patrones diferenciales en las demandas ciudadanas. Esta limitación es reconocida como oportunidad de desarrollo en sección 4.7.

4.6 Principales contribuciones

Este trabajo aporta en cuatro dimensiones.

En el plano técnico, desarrolla una solución escalable y reproducible basada en IAG para clasificación temática masiva.

En el plano metodológico, ofrece una comparación estructural rigurosa entre ML clásico e IAG aplicada a un corpus constitucional real de gran escala.

En el plano empírico, demuestra que la IAG amplía la activación de categorías ciudadanas emergentes sin alterar las prioridades centrales del debate.

En el plano institucional, evidencia la viabilidad de integrar herramientas generativas en organismos públicos para apoyar procesos deliberativos complejos.

4.7 Proyecciones y trabajos futuros

Los resultados obtenidos permiten proyectar diversas líneas de profundización.

Una línea relevante consiste en analizar la evolución temática entre las etapas local, provincial y regional del proceso constituyente 2016. Dado que cada etapa alimentó la siguiente, resulta pertinente estudiar la persistencia o transformación de categorías, la estabilidad del núcleo temático y el comportamiento de categorías ciudadanas emergentes en un contexto deliberativo secuencial.

Asimismo, se propone desarrollar análisis segmentados utilizando variables territoriales y de pertenencia disponibles (comuna, región, país, residencia en el extranjero y pueblos originarios), con el fin de evaluar posibles diferencias temáticas entre grupos.

Desde una perspectiva metodológica, sería valioso fortalecer la taxonomía incorporando definiciones sintéticas por categoría específica y utilizar IAG para generar resúmenes o “ideas fuerza” asociadas a categorías ciudadanas emergentes.

Asimismo, una línea de profundización pendiente consiste en evaluar la calidad semántica de los casos donde ML e IAG coincidieron exactamente en su clasificación (58,1% del corpus), lo que permitiría estimar qué proporción de clasificaciones compartidas son también semánticamente correctas, complementando la evidencia sobre calidad absoluta de ambos métodos.

Finalmente, se sugiere extender la aplicación de IAG a la quinta pregunta del proceso participativo, orientada a aportes abiertos, con el objetivo de sistematizar propuestas relevantes que aún no han sido estructuradas categorialmente.

4.8 Reflexión final

La presente investigación demuestra que los modelos de Inteligencia Artificial Generativa pueden incorporarse de manera rigurosa en el análisis de procesos participativos complejos. La IAG no sustituye el juicio experto, pero sí amplía significativamente la capacidad de procesamiento, estructuración y análisis de grandes volúmenes de información ciudadana.

En el contexto de procesos constituyentes, donde la legitimidad también se construye a partir de la sistematización transparente de las voces ciudadanas, disponer de herramientas que permitan analizar masivamente aportes abiertos con consistencia conceptual constituye un avance metodológico relevante.

Agradecimientos.

Agradezco el apoyo y el cariño de mis compañeros y colegas de la Biblioteca del Congreso Nacional, así como también el acompañamiento de mi familia, pareja y amigos. Así también, agradezco a mis compañeros de magíster y a mis profesores, el coordinador de magister, los profesionales que participaron con su evaluación experta y muy especialmente a mi profesor guía, quienes me motivaron y guiaron a lo largo de todo este proceso.

Referencias

Lista de todas las referencias usadas en el informe.

- [1] Biblioteca del Congreso Nacional de Chile, «Historia de Leyes Modificatorias, Constitución Política de 1980», Historia de la Constitución y de las Propuestas de Nuevas Constituciones. Accedido: 25 de noviembre de 2025. [En línea]. Disponible en: <https://www.bcn.cl/historia-de-la-constitucion/historia-leyes-modificatorias?id=242302>
- [2] C. Heiss, «Legitimacy crisis and the constitutional problem in Chile: A legacy of authoritarianism», *Constellations*, vol. 24, pp. 470-479, sep. 2017, doi: 10.1111/1467-8675.12309.
- [3] M. Bachelet Jeria, «Programa de Gobierno Michelle Bachelet 2014-2018», BND: Colección Digital. Accedido: 25 de noviembre de 2025. [En línea]. Disponible en: <https://www.bibliotecanacionaldigital.gob.cl/bnd/646/w3-article-157782.html>
- [4] Gobierno de Chile, «Boletín 11617-07: Para modificar la Constitución Política de la República». Accedido: 25 de noviembre de 2025. [En línea]. Disponible en: https://www.senado.cl/appsenado/templates/tramitacion/index.php?boletin_ini=11617-07
- [5] Wikipedia, «Estallido social», *Wikipedia, la enciclopedia libre*. 6 de noviembre de 2025. Accedido: 25 de noviembre de 2025. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Estallido_social&oldid=170353449
- [6] Ministerio del Interior, «Ministro Gonzalo Blumel valora acuerdo transversal por la Paz Social y una Nueva Constitución», <http://www.interior.gov.cl/>. Accedido: 25 de noviembre de 2025. [En línea]. Disponible en: <http://www.interior.gov.cl/noticias/2019/11/15/ministro-gonzalo-blumel-valora-acuerdo-transversal-por-la-paz-social-y-una-nueva-constitucion/>
- [7] U. de C. F. de Medicina *et al.*, «Demandas prioritarias y propuestas para un Chile diferente: Sistematización de 1233 cabildos ciudadanos», mar. 2021, Accedido: 27 de noviembre de 2025. [En línea]. Disponible en: <https://repositorio.uchile.cl/handle/2250/178886>
- [8] Biblioteca del Congreso Nacional de Chile, «Documentos para el debate nacional», bcn.cl. Accedido: 25 de noviembre de 2025. [En línea]. Disponible en: <https://www.bcn.cl/debatenacional>
- [9] Biblioteca del Congreso Nacional de Chile, «Proceso Constituyente 2021-2022», bcn.cl. Accedido: 25 de noviembre de 2025. [En línea]. Disponible en: <https://www.bcn.cl/procesoconstituyente/>
- [10] Biblioteca del Congreso Nacional de Chile, «Cabildos 2016. Proceso Constituyente | Biblioteca del Congreso Nacional de Chile», bcn.cl. Accedido: 25 de noviembre de 2025. [En línea]. Disponible en: <https://www.bcn.cl/procesoconstituyente/cabildos2016/>
- [11] Biblioteca del Congreso Nacional de Chile, *BCN Chile presentó herramientas digitales ante comisión temática de la Convención Constitucional*, (10 de noviembre de 2021). Accedido: 27 de noviembre de 2025. [En línea Video]. Disponible en: <https://www.youtube.com/watch?v=0HdRednlEHQ>
- [12] Biblioteca del Congreso Nacional de Chile, «BCN presentó herramientas digitales ante la Convención Constitucional», Biblioteca del Congreso Nacional de Chile. Accedido: 27 de noviembre de 2025. [En línea]. Disponible en: <https://www.bcn.cl/portal/noticias?id=bcn-presento-herramientas-digitales-ante-la-convencion-constitucional>
- [13] Ministerio Secretaría General de la Presidencia, T. Jordán Díaz, P. Figueroa Rubio, R. Araya, y C. Gómez, «Proceso constituyente abierto a la ciudadanía: Guía metodológica para la etapa participativa territorial». 2016. [En línea]. Disponible en: https://www.bcn.cl/procesoconstituyente/cabildos2016/pdf/guia_metodologica_proceso_constituyente_a_buerto_a_la_ciudadania.pdf
- [14] Ministerio Secretaría General de la Presidencia, «Dataset de Proceso Constituyente Abierto a la Ciudadanía», Portal de Datos Abiertos. Accedido: 28 de noviembre de 2025. [En línea]. Disponible en: <https://datos.gob.cl/dataset/proceso-constituyente-abierto-a-la-ciudadania>
- [15] E. G. de E. García de Enterría, «La constitución como norma jurídica», *Anu. Derecho Civ.*, n.º 2-3, pp. 291-342, sep. 1979.

- [16] RAE, «Definición de Constitución1 - Diccionario panhispánico del español jurídico - RAE», Diccionario panhispánico del español jurídico - Real Academia Española. Accedido: 11 de diciembre de 2025. [En línea]. Disponible en: <https://dpej.rae.es/lema/constituci%C3%B3n1>
- [17] Ministerio Secretaría General de la Presidencia, «Decreto 100 | Fija el texto refundido, coordinado y sistematizado de la Constitución política de la República de Chile», www.bcn.cl/leychile. Accedido: 11 de diciembre de 2025. [En línea]. Disponible en: <https://www.bcn.cl/leychile/navegar?idNorma=242302>
- [18] M. Ríos Tobar, «Mecanismos de cambio constitucional en el mundo. Análisis desde la experiencia comparada», UNDP. Accedido: 14 de diciembre de 2025. [En línea]. Disponible en: <https://www.undp.org/es/latin-america/publicaciones/mecanismos-de-cambio-constitucional-en-el-mundo-analisis-desde-la-experiencia-comparada>
- [19] G. Pisarello, *Procesos constituyentes: caminos para la ruptura democrática*. Madrid: Trotta. Accedido: 21 de diciembre de 2025. [En línea]. Disponible en: <https://www.prosaypolitica.cl/libro/procesos-constituyentes-caminos-para-la-ruptura-democratica/>
- [20] Biblioteca del Congreso Nacional, «Constitución Política de la República de Chile, de 16-NOV-2015 a 04-ENE-2017», www.bcn.cl/leychile. Accedido: 21 de diciembre de 2025. [En línea]. Disponible en: <https://www.bcn.cl/leychile/navegar?idNorma=242302&idVersion=2015-11-16>
- [21] Ministerio Secretaría General de la Presidencia, «Síntesis de resultados etapa participativa - Archivo Michelle Bachelet Jeria 2014-2018». Accedido: 22 de diciembre de 2025. [En línea]. Disponible en: <https://archivospresidenciales.archivonacional.cl/index.php/sintesis-de-resultados-etapa-participativa>
- [22] P. Figueroa Rubio y T. Jordán Díaz, «Convergencia Deliberativa: La participación ciudadana en el cambio constitucional, Chile 2016», *Polis Rev. Latinoam.*, n.º 58, mar. 2021, doi: <http://dx.doi.org/10.32735/S0718-6568/2021-N58-1579>.
- [23] C. Fierro, C. Fuentes, J. Pérez, y M. Quezada, «200K+ Crowdsourced Political Arguments for a New Chilean Constitution», en *Proceedings of the 4th Workshop on Argument Mining*, I. Habernal, I. Gurevych, K. Ashley, C. Cardie, N. Green, D. Litman, G. Petasis, C. Reed, N. Slonim, y V. Walker, Eds., Copenhagen, Denmark: Association for Computational Linguistics, sep. 2017, pp. 1-10. doi: 10.18653/v1/W17-5101.
- [24] Ministerio Secretaría General de la Presidencia *et al.*, «Informe Final sobre el proceso de participación y diálogos constitucionales», Consejo Ciudadano de Observadores, ene. 2017. [En línea]. Disponible en: <https://www.bcn.cl/procesoconstituyente/cabildos2016/pdf/informe-final-cco-16-de-enero-de-2017.pdf>
- [25] A. Cruz, Z. Elkins, R. Gardner, M. Martin, y A. Moran, «Measuring constitutional preferences: A new method for analyzing public consultation data», *PLOS ONE*, vol. 18, n.º 12, p. e0295396, dic. 2023, doi: 10.1371/journal.pone.0295396.
- [26] I. Chalkidis, M. Fergadiotis, y I. Androutsopoulos, «MultiEURLEX - A multi-lingual and multi-label legal document classification dataset for zero-shot cross-lingual transfer», en *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, y S. W. Yih, Eds., Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, nov. 2021, pp. 6974-6996. doi: 10.18653/v1/2021.emnlp-main.559.
- [27] E. Gunes y C. K. Florczak, «Multiclass Classification of Policy Documents with Large Language Models», 2023, doi: 10.48550/ARXIV.2310.08167.
- [28] «Realizan visita a sede de Cabildo Regional ubicada en Estación Mapocho», Ilustre Municipalidad de Santiago. Accedido: 24 de diciembre de 2025. [En línea]. Disponible en: <https://www.munistgo.cl/realizan-visita-a-sede-de-cabildo-regional-ubicada-en-estacion-mapocho/>
- [29] S. T. SAMIUC, «Kappa de Cohen», SAMIUC. Accedido: 24 de febrero de 2026. [En línea]. Disponible en: <https://www.samiuc.es/estadisticas-variables-binarias/medidas-de-concordancia/kappa-de-cohen/>