

Caso de éxito

Asistente virtual de información “MAKU” para la Unidad Nacional de Protección

Programa de Fortalecimiento de Habilidades y
Herramientas de Inteligencia Artificial para el Sector
Público

Nombre Autor (a): Andrés López Gómez.
2025

1. Nombre del Caso/Iniciativa

Asistente virtual de información “MAKU” para la Unidad Nacional de Protección.

2. Entidad(es) Responsable(s)

Unidad Nacional de Protección (UNP) de Colombia, Oficina Asesora PI – GGT.

3. Sector Administrativo (Salud, Justicia, Educación, etc.)

Seguridad, Ministerio del Interior.

4. Área de Aplicación

Optimización de Servicios y Atención al Ciudadano

5. Problema Público Abordado

La tarea que se buscó optimizar fue la atención de consultas ciudadanas sobre trámites y normatividad de la UNP, que se realizaba de forma manual a través de teléfono, correo electrónico y ventanilla. El problema principal era la acumulación de preguntas frecuentes repetitivas, los tiempos de respuesta prolongados (sobre todo fuera del horario laboral), la sobrecarga de correos y llamadas, la dificultad de las personas para encontrar por sí mismas la resolución o norma aplicable, el riesgo de respuestas inconsistentes entre funcionarios y las barreras de acceso para personas con baja alfabetización digital o discapacidad visual al no existir un canal conversacional con respuesta en voz.

6. Solución de IA Implementada

Durante la Semana de la Innovación se diseñó y construyó un prototipo funcional de asistente virtual llamado MAKU. Primero se recopilaron y depuraron resoluciones, instructivos, preguntas frecuentes y guías de trámites, que se condensaron en un corpus estructurado (corpus.txt) con información oficial de la UNP. Luego se configuró un backend en FastAPI (Python) que recibe la pregunta del usuario, consulta el corpus y construye un prompt específico para un modelo de lenguaje de Google Gemini, indicándole que responda sólo con base en esa información institucional y en tono acorde con la entidad. El modelo genera una respuesta breve y clara en español que el backend utiliza tanto en formato de texto como para producir audio mediante el servicio de Text-to-Speech de Google Cloud. Cada interacción se registra con fecha y

hora en un archivo de log para trazabilidad. Paralelamente se desarrolló un frontend en React + TypeScript que ofrece un campo sencillo para escribir la pregunta, muestra la conversación en pantalla y reproduce el audio de la respuesta utilizando Wavesurfer.js para visualizar la onda sonora. En la fase de laboratorio se realizaron pruebas con casos reales de preguntas frecuentes, ajustando los prompts, el formato de salida y validando la coherencia de las respuestas con la normatividad y lineamientos de la entidad.

7. Tecnologías Utilizadas (ej. PLN, Visión por Computador, Machine Learning)

Se empleó principalmente el modelo de lenguaje Google Gemini (gemini-2.5-flash-lite) como motor de comprensión de lenguaje natural y generación de respuestas, restringido a un corpus propio de la UNP. Además, se utilizó el servicio de Text-to-Speech de Google Cloud para transformar las respuestas en audio en español de Colombia, y se dejó configurado, a nivel de laboratorio, el servicio de Speech-to-Text para una futura fase en la que la ciudadanía pueda formular preguntas por voz. Estas capacidades de procesamiento de lenguaje natural y voz se integraron en una arquitectura basada en FastAPI, React, TypeScript, Vite, Wavesurfer.js y herramientas de registro y despliegue en Python.

8. Resultados

En la situación de referencia, todas las consultas se atendían por canales tradicionales (teléfono, correo, ventanilla), con alta carga de trabajo para los funcionarios y sin un canal automatizado disponible en cualquier horario. El prototipo MAKU, en entorno de pruebas, muestra que las preguntas frecuentes sobre la UNP, su estructura, el significado de la SER, los niveles de riesgo, las poblaciones protegidas o la Línea Vida 103 pueden ser respondidas de forma inmediata en cuestión de segundos, con respuestas homogéneas basadas en fuentes oficiales. Esto implica un ahorro de tiempo para el personal, que puede concentrarse en casos complejos, y una reducción del tiempo de respuesta percibido por la ciudadanía. Cualitativamente, se mejora la calidad del servicio gracias a la consistencia en la información entregada, la incorporación de audio como alternativa al texto para personas con dificultades de lectura o discapacidad visual y la posibilidad de auditar y mejorar el sistema mediante el registro de interacciones. Aunque aún no existen métricas masivas en producción, las pruebas internas permiten proyectar una disminución significativa de consultas

repetitivas en canales tradicionales y un aumento potencial de la satisfacción ciudadana cuando el sistema se despliegue formalmente.

9. Factor de Sostenibilidad y Escalabilidad

La solución fue concebida con una arquitectura modular y reutilizable que facilita su réplica en otras dependencias de la UNP o en entidades públicas con retos similares. Para escalarla basta con que cada área identifique y organice sus documentos oficiales, estructure un corpus análogo, ajuste los prompts al lenguaje institucional propio y conecte el mismo backend de IA generativa y servicios de voz. Al basarse en servicios en la nube y tecnologías de software libre o de amplio uso, la solución puede mantenerse con recursos relativamente acotados, siempre que exista un proceso de actualización del corpus cuando cambie la normatividad. El beneficio para la ciudadanía y para la misión institucional radica en ampliar la capacidad de respuesta de la UNP, ofrecer información clara y confiable sobre servicios de protección, reducir la incertidumbre en momentos sensibles, eliminar barreras de acceso para personas con dificultades de lectura y reforzar la transparencia al fundamentar las respuestas en documentos oficiales controlados.