



Junta de Andalucía

Asistente virtual de escolarización con IA generativa

Candidatura a la convocatoria de premios SocInfo Digital “IA en la AAPP”

Categoría “IA para la mejora de los Servicios al Ciudadano”

<i>01 /</i>	<i>DESCRIPCIÓN BREVE DEL PROYECTO</i>	<i>3</i>
<i>02 /</i>	<i>DESARROLLO DEL PROYECTO</i>	<i>5</i>
<i>03 /</i>	<i>DESCRIPCIÓN TÉCNICA</i>	<i>11</i>
<i>04 /</i>	<i>EL ASISTENTE VIRTUAL EN FUNCIONAMIENTO</i>	<i>16</i>
<i>05 /</i>	<i>REPERCUSIÓN PARA EL CIUDADANO Y LAS ADMINISTRACIONES</i>	<i>19</i>
<i>06 /</i>	<i>EQUIPO DE DESARROLLO Y PROVEEDORES</i>	<i>21</i>
<i>07 /</i>	<i>VALORACIÓN ECONÓMICA</i>	<i>22</i>
<i>08 /</i>	<i>RESUMEN DE LA CANDIDATURA</i>	<i>23</i>
<i>09 /</i>	<i>CONTACTO</i>	<i>24</i>

01 / Descripción breve del proyecto

El asistente virtual corporativo de la Junta de Andalucía pretende facilitar y mejorar la relación de la Junta de Andalucía con la ciudadanía, ofreciendo un canal complementario y adicional a los otros canales existentes para dicha relación (teléfono y correo electrónico, principalmente). Además de proporcionar servicio a la ciudadanía durante las 24 horas del día, los 365 días del año, pretende ser una ayuda para reducir los tiempos de respuesta del servicio de atención telefónica y reducir la sobrecarga de este en períodos pico en los que el volumen de llamadas recibidas supera a la capacidad de atención del servicio.

Este asistente se ha desarrollado en el marco de un contrato licitado mediante concurso abierto y financiado por fondos europeos (FEDER), que se formalizó en julio de 2023. En ese marco la Junta de Andalucía ha implantado y continúa implantando asistentes virtuales basados en inteligencia artificial tradicional, para los ámbitos tributario, de servicios y trámites, de familia numerosa y de empleo público. Asimismo, ha realizado un proyecto piloto utilizando inteligencia artificial generativa para atender en relación a la campaña de escolarización de 2024. Las conclusiones de este proyecto piloto permitirán orientar mejor los desarrollos futuros de asistentes virtuales.

El asistente virtual de escolarización atiende a consultas escritas planteadas a través del portal de escolarización de la Consejería de Desarrollo Educativo y Formación Profesional, mediante un desplegable emergente al que se accede desde un botón situado en el lateral derecho de la página. Este desplegable emergente es compartido con los asistentes desarrollados con inteligencia artificial tradicional, garantizando así una identidad visual única y coherente para todos los servicios proporcionados mediante asistente virtual.

Entre las capacidades más destacadas del asistente se encuentran:

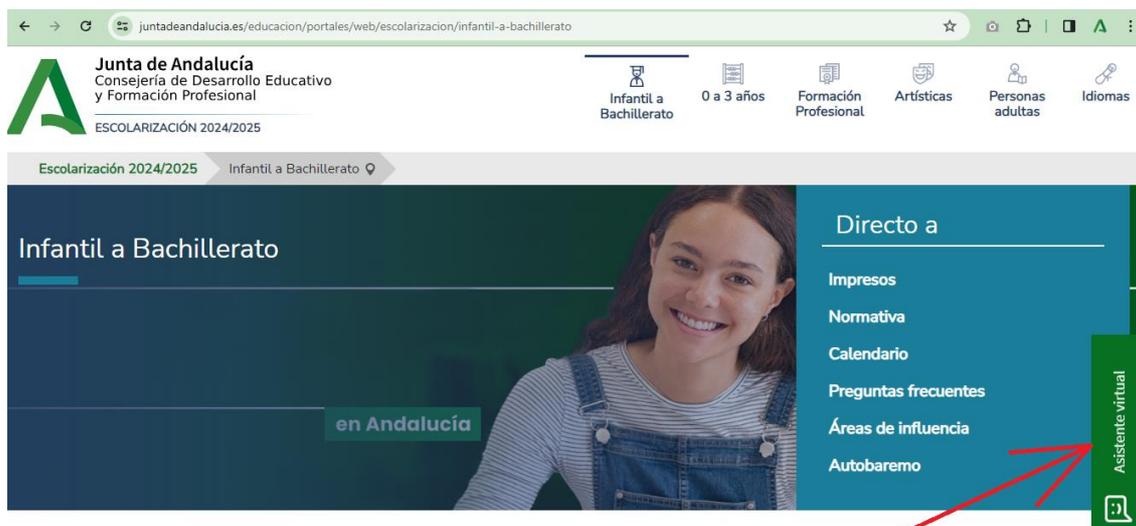
- Responder en cualquier momento a un alto número de preguntas de forma rápida y relevante
- Responder a solicitudes complejas que implican responder a más de una pregunta al mismo tiempo
- Responder en lenguaje natural, con respuestas personales y no robotizadas
- Derivar al centro de atención a usuarios de la comunidad educativa (CAUCE) aquellas consultas cuya resolución es especialmente compleja y en la que hay mayor riesgo de dar una respuesta inadecuada desde el asistente virtual.

Este asistente virtual de escolarización, basado en inteligencia artificial generativa, se puso en producción el día 7 de marzo de 2024, y desde entonces puede ser utilizado por cualquier persona, accediendo al portal de escolarización y, simplemente, haciendo click sobre el icono del mismo.

Para el desarrollo de este asistente se ha hecho uso de tecnología de OpenAI, a través de una suscripción Azure para garantizar la seguridad. En concreto se está utilizando el modelo de lenguaje GPT-4 Turbo, al cual se ha entrenado con la documentación y preguntas

Descripción breve del proyecto

frecuentes relevantes en materia de escolarización, y realizando ingeniería de prompts para aumentar la fiabilidad y concreción de las respuestas.



Información general

La Consejería de Desarrollo Educativo y Formación Profesional convoca el procedimiento de admisión del alumnado en los centros docentes públicos y privados concertados para cursar las enseñanzas de 2º Ciclo de Educación Infantil, Educación Primaria, Educación Especial, Educación Secundaria Obligatoria y Bachillerato para el curso escolar 2024/25.

Entre los días 1 de marzo y 1 de abril, ambos inclusive, permanecerá abierto en Andalucía el plazo para la presentación de solicitudes de admisión correspondientes al curso escolar 2024/25. Se inicia de este modo el proceso de escolarización en el que participan todos los niños y niñas que se

02 / Desarrollo del proyecto

La Agencia Digital de Andalucía implantó un proyecto piloto de asistente virtual en julio del año 2021, y lo extendió durante ese año para responder por escrito a preguntas frecuentes en los ámbitos tributario, de servicios y trámites, de familias e igualdad y de empleo público. Este asistente virtual estaba basado en tecnología de inteligencia artificial tradicional (no generativa) y se apoyaba en el motor conversacional IBM Watson.

En diciembre de 2022 convocó un concurso público para la contratación del desarrollo y mantenimiento de un asistente virtual corporativo, con el objetivo de atender a la ciudadanía en relación a múltiples materias, tanto contestando a preguntas frecuentes como permitiendo la obtención de cita previa y la realización de trámites sencillos, todo ello de manera conversacional apoyándose en el empleo de inteligencia artificial. A este concurso se presentaron 9 empresas, que ofertaron soluciones de 5 fabricantes diferentes, adjudicándose finalmente a Indra (Minsait), formalizándose el contrato en julio de 2023.

En noviembre de 2023 se completó la migración del asistente virtual piloto a la infraestructura ofertada de asistente virtual corporativo, también basada en servicios en nube de IBM Cloud y en el uso del motor conversacional IBM Watson. Asimismo, el despliegue de esa infraestructura permite avanzar en la extensión del asistente a otros dominios de conocimiento, y se está trabajando en la actualidad en la ampliación de capacidades del asistente en materia de servicios y trámites y en la extensión a otros dominios, en la mayoría de los casos utilizando técnicas de procesamiento del lenguaje natural (NLP) tradicionales.

Con la publicación en diciembre de 2022 de ChatGPT se produjo una revolución en el marco del procesamiento de lenguaje natural, que hasta ese momento había estado muy orientado al empleo de técnicas de NLP tradicionales, y cuyo foco cambió en ese momento al empleo de técnicas de inteligencia artificial generativa (IA generativa) basada en grandes modelos de lenguaje (LLM).

Los nuevos sistemas de IA generativa para lenguaje natural superan ampliamente a los motores NLP tradicionales en cuanto a la naturalidad de la comunicación y a la capacidad de dar respuestas variadas y diversas en multitud de ámbitos, incluso sin haber sido entrenados de manera específica. Sin embargo, presentan carencias cuando se los intenta usar para fines específicos y se requiere que den respuestas precisas y fiables basadas en un entrenamiento específico para un dominio de conocimiento concreto. De hecho, es frecuente que tengan alucinaciones, es decir, que den respuestas totalmente inventadas ante algunas preguntas, especialmente cuando no tienen una respuesta clara, aunque también puede ocurrir de manera aleatoria ante preguntas comunes y sencillas. Por este motivo no es habitual su uso, todavía, en ámbitos en los que se requiere fiabilidad en las respuestas, como por ejemplo en la atención a la ciudadanía por parte de la administración pública.

El campo de la IA generativa está avanzando a un ritmo muy acelerado, y se están desarrollando técnicas para mejorar el entrenamiento de los modelos en materias específicas

Desarrollo del proyecto

y para reducir la probabilidad de alucinaciones. Esto ha hecho que la Agencia Digital de Andalucía se plantee la necesidad de realizar un proyecto piloto con esta tecnología, utilizando los modelos de lenguaje más avanzados y las técnicas más modernas para mejorar la fiabilidad de las respuestas.

Con ese fin, el día 11 de enero de 2024 se celebró una reunión con el objetivo de analizar la viabilidad y el interés de la Consejería de Desarrollo Educativo y Formación Profesional en implantar un asistente virtual para atender a la ciudadanía basado en IA generativa. En esta reunión se contó con la participación de las personas responsables del proceso de escolarización en esa Consejería así como de las personas responsables en la Agencia Digital de Andalucía de los sistemas de información de Educación y del asistente virtual corporativo.

En la citada reunión se puso de manifiesto el interés en implantar el asistente, así como la disposición a asumir, de manera controlada, los riesgos derivados de la utilización de una tecnología novedosa y poco probada para la atención a la ciudadanía por parte de la administración pública. No obstante, se determinó que este proyecto solo tendría utilidad si podía ser utilizado para la campaña de escolarización, cuyo período de mayor actividad coincide con el plazo de presentación de solicitudes de admisión entre el 1 de marzo y el 1 de abril de 2024. Teníamos, por tanto, luz verde para acometer el proyecto, pero al mismo tiempo un plazo muy ajustado, inferior a dos meses.

Inmediatamente después de la reunión se abrieron varias líneas de trabajo en paralelo:

- Obtención y clasificación de la información sobre el proceso de escolarización necesaria para el entrenamiento del asistente virtual.
- Selección del modelo de IA generativa a utilizar.
- Implantación de los servicios en nube para el acceso al modelo de IA generativa seleccionado.
- Adaptación de la infraestructura para permitir el uso de IA generativa y facilitar el acceso a los servicios en nube correspondientes.
- Adaptación del portal de escolarización para la publicación del asistente.

En lo relativo a la información a utilizar para el entrenamiento del asistente virtual se seleccionaron tres fuentes documentales: dos de ellas eran la documentación utilizada para entrenar a los agentes del Servicio 012 y del centro de atención a usuarios de la comunidad educativa (CAUCE) y el tercero eran las preguntas frecuentes publicadas en la web del proceso de escolarización. Fue necesario, no obstante, adaptar y reestructurar esta información para que el asistente la pudiera interpretar correctamente y para afinar las respuestas.

En cuanto a la selección del modelo de IA generativa a utilizar, dado que se trata de un proyecto piloto orientado a ensayar esta tecnología y aprender a utilizarla en este contexto, se evaluaron varias opciones de distintos fabricantes, pero se optó rápidamente por la solución que en ese momento era más potente a la par que más madura, es decir, por los modelos GPT de Open AI. Asimismo, para garantizar la seguridad se optó por utilizar los

Desarrollo del proyecto

servicios de Open AI publicados en la nube de Microsoft Azure, teniendo en cuenta además que la Agencia Digital de Andalucía ya disponía de una suscripción de Azure que podía ser utilizada para la prestación de estos servicios. Quedaba pendiente de decidir, no obstante, si se utilizaba el modelo GPT 3.5 o el modelo GPT 4, dado que el modelo GPT 4 es más potente y fiable, pero al mismo tiempo significativamente más lento al generar las respuestas.

Para la implantación de los servicios de Open AI en Azure fue necesaria una estrecha e intensa colaboración entre los múltiples proveedores involucrados, incluyendo a Minsait (a través de dos contratos distintos) y a Sandetel, así como el apoyo de Microsoft como proveedor de la nube Azure.

La infraestructura de asistente virtual ya se encontraba desplegada, como hemos mencionado, en IBM Cloud, e incluía múltiples elementos, incluyendo una interfaz de front-end para mostrar el asistente a los usuarios y una base de datos de conversaciones, todos los cuales estaban diseñados para interoperar únicamente con Watson Assistant. Fue necesario, por tanto, evolucionar la infraestructura y todos los componentes desarrollados sobre ella para permitir el consumo simultáneo de servicios de Open AI en Azure (para el asistente de escolarización) y de servicios de Watson Assistant para los asistentes preexistentes y los nuevos que se encuentran en desarrollo y pruebas.

La adaptación del portal de escolarización para la publicación del asistente resultó sencilla, comparativamente con el resto de trabajos, dado que el sistema está diseñado para que la inserción del asistente se realice mediante unas pocas líneas de código JavaScript, encargándose del resto de trabajos un widget desarrollado exprofeso para el asistente virtual corporativo, y que ya se encontraba en producción para los asistentes desplegados en noviembre de 2023 (aunque requirió pequeños ajustes)

La coordinación simultánea de todas estas actuaciones resultó compleja y no estuvo exenta de dificultades, pero la extraordinaria implicación y dedicación, y el excelente trabajo realizado por todas las partes, permitieron que el día 20 de febrero ya se pudieran realizar pruebas integradas, sobre el entorno de preproducción, por parte de la Consejería de Desarrollo Educativo y Formación Profesional. Estas pruebas se realizaron utilizando el modelo GPT 3.5, y con ellas se obtuvieron dos conclusiones:

- El material utilizado para el entrenamiento tenía errores y no estaba suficientemente actualizado a la campaña de escolarización de 2024.
- El asistente virtual daba respuestas “creativas” ante diversas preguntas, proporcionando respuestas inadecuadas o directamente falsas. Pese a haber sido entrenado con documentación detallada sobre el proceso de escolarización, completaba las respuestas con información que no constaba en la documentación.

El primero de estos “problemas” resultaba sencillo de resolver, bastaba con revisar a fondo y actualizar toda la documentación utilizada para entrenar al asistente. De hecho, sirvió de ayuda para detectar y corregir errores e información desactualizada en la documentación utilizada por CAUCE y en las preguntas frecuentes publicadas en el portal de escolarización.

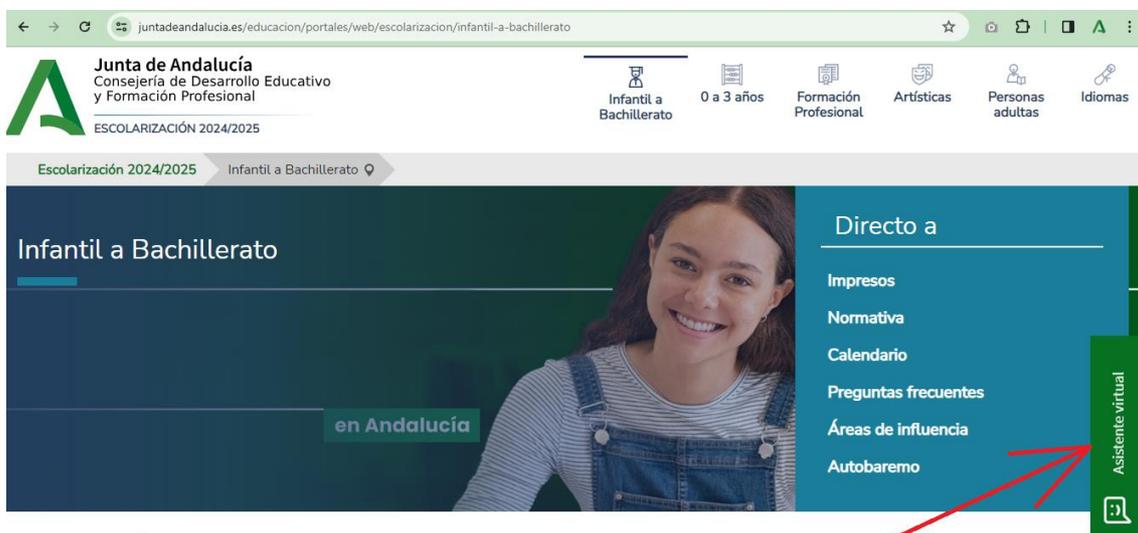
Desarrollo del proyecto

El segundo “problema” resultaba más peliagudo, porque apuntaba de lleno al principal temor a la hora de utilizar IA generativa para atender a la ciudadanía: el riesgo de que el asistente cometa alucinaciones y proporcione respuestas falsas pero con apariencia de ser correctas. Para solucionarlo se tomaron varias decisiones:

- Dar al asistente instrucciones de no contestar a preguntas relativas a las materias más complejas: la escolarización de alumnos con necesidades educativas especiales, y la escolarización en procedimiento extraordinario. Ante preguntas relativas a esos temas el asistente debe decir al ciudadano que consulte a CAUCE. Esto se realizó mediante lo que en el argot se conoce como ingeniería de prompting, que consiste básicamente en afinar la forma de pedir al modelo de lenguaje lo que se quiere que éste haga.
- Modificar la documentación utilizada para el entrenamiento del asistente, en varios ámbitos:
 - o Eliminar de la documentación utilizada para entrenar la mayoría de la información relativa a las materias que el asistente no debe contestar, y sustituirla por una breve descripción de cada una de esas materias y una observación indicando que para obtener información sobre esas cuestiones debe consultarse a CAUCE.
 - o Modificar el conjunto documental usado para entrenar al asistente, dejándolo en tres documentos:
 - Un documento de conocimiento, basado en el utilizado por CAUCE pero adaptado a las necesidades específicas de información del asistente virtual, y eliminando la información que puede ser útil para un agente humano que atiende por teléfono pero que puede ser excesiva y/o producir confusión al ser utilizada en el entrenamiento de un asistente virtual.
 - Un documento de preguntas frecuentes, igualmente modificado para eliminar el exceso de información.
 - Un documento de datos transversales, con aquella información que puede ser útil para contestar a preguntas diversas (fechas y plazos, enlaces a páginas web informativas sobre diversos aspectos del proceso, y siglas habituales).
- Utilizar el modelo GPT 4 para mejorar la fiabilidad y reducir el riesgo de alucinaciones, asumiendo la pérdida de velocidad (mayor latencia) en la obtención de respuestas. Para mitigar esa pérdida de velocidad se implantó un mecanismo de streaming, similar al utilizado por ChatGPT, de modo que las respuestas se van proporcionando al usuario conforme se van obteniendo, sin esperar a que estén completas, y así el usuario va leyendo la respuesta por partes mientras se termina de construir, reduciendo la sensación de espera.

Desarrollo del proyecto

- Incorporar al asistente mensajes informativos sobre el carácter del asistente y sobre la fiabilidad de las respuestas. En concreto se han incorporado los siguientes mensajes:
 - o En el mensaje de bienvenida o saludo del asistente se ha incorporado el siguiente texto: “Soy un proyecto piloto de la Junta de Andalucía, que pretende aplicar tecnologías innovadoras para mejorar la relación con la ciudadanía. Mis respuestas se generan automáticamente mediante un sistema de inteligencia artificial generativa y no son vinculantes, solo orientativas. Por ello recomendamos contrastar la información en las fuentes oficiales a través de la [web de la Junta de Andalucía.](#)”
 - o En la parte superior de la ventana del asistente se ha incorporado un mensaje informativo, permanentemente visible, que dice “Esta información es orientativa y no vinculante, contrasta las respuestas” y proporciona un enlace para obtener más información, concretamente la siguiente: “Este asistente virtual es un proyecto piloto de la Junta de Andalucía, que pretende aplicar tecnologías innovadoras para mejorar la relación con la ciudadanía. Las respuestas del asistente se generan automáticamente mediante un sistema de inteligencia artificial generativa y no son vinculantes, solo orientativas. Por ello recomendamos contrastar la información en las fuentes oficiales a través de la [Web de la Junta de Andalucía.](#) El sistema no accede a tus datos personales. Esta conversación se almacenará de forma anónima y se utilizará para la mejora del servicio.”



Información general

La Consejería de Desarrollo Educativo y Formación Profesional convoca el procedimiento de admisión del alumnado en los centros docentes públicos y privados concertados para cursar las enseñanzas de 2º Ciclo de Educación Infantil, Educación Primaria, Educación Especial, Educación Secundaria Obligatoria y Bachillerato para el curso escolar 2024/25.

Entre los días 1 de marzo y 1 de abril, ambos inclusive, permanecerá abierto en Andalucía el plazo para la presentación de solicitudes de admisión correspondientes al curso escolar 2024/25. Se inicia de este modo el proceso de escolarización en el que participan todos los niños y niñas que se

Tras realizar estos cambios, los días 5 y 6 de marzo se realizaron nuevas pruebas, con resultados satisfactorios, por lo que el día 7 de marzo se puso en producción el asistente, al

Desarrollo del proyecto

cual se puede acceder en la siguiente URL:
<https://www.juntadeandalucia.es/educacion/portales/web/escolarizacion/infantil-a-bachillerato>

03 / Descripción técnica

Caso de uso: Asistente virtual informativo, de pregunta-respuesta.

Ámbito de conocimiento: Campaña de escolarización de infantil a bachillerato para el curso 2014-2025 de la Junta de Andalucía.

Componentes técnicos:

- Modelos de lenguaje (LLMs) empleados mediante el servicio de Azure OpenAI:
 - Text-embedding-ada-002: este modelo se utiliza para la transformación de texto a embeddings, es decir se vectoriza la documentación disponible en chunks (secciones) para posteriormente, con la pregunta del usuario vectorizada, comprobar como de cercana es la pregunta del usuario a alguno de los chunks de la documentación que forma la base de conocimiento.
 - GPT 4 turbo (1106): es el modelo al que se envía el prompt para que genere una respuesta basándose en la base de conocimiento (documentación indexada), en la pregunta del usuario y, en el caso de haberlo, en el contexto previo de la misma conversación.
- Minsait Conversational Interfaces Services (CIS): Los CIS son activos de Minsait destinados a agilizar el diseño, desarrollo y monitorización de los asistentes virtuales.

En este caso se emplean los siguientes:

- Framebot: Framebot es un activo, un framework de desarrollo, propio de Minsait para el desarrollo de asistentes virtuales:
 - Microservicio orquestador: Microservicio encargado de dirigir las peticiones a otros microservicios, además de ser el punto de entrada a Framebot.
 - Microservicio de IA generativa: Microservicio encargado de la comunicación con Azure OpenAI y del procesamiento de la información enviada y recibida.
 - Microservicio analytics: encargado de registrar en base de datos información relevante para su explotación futura en cuadros de mando.
 - PostgreSQL: base de datos compartida por todos los microservicios para el manejo de las conversaciones y usuarios.
- Widget de conversación: componente web-chat personalizado con estilos de la identidad corporativa de la Junta de Andalucía.
- BBDD de conversaciones: registro de las interacciones del usuario asociadas a un identificador de conversación, fecha y hora, respuesta generada por el asistente para esa interacción, sección de la documentación seleccionada y clúster detectado.

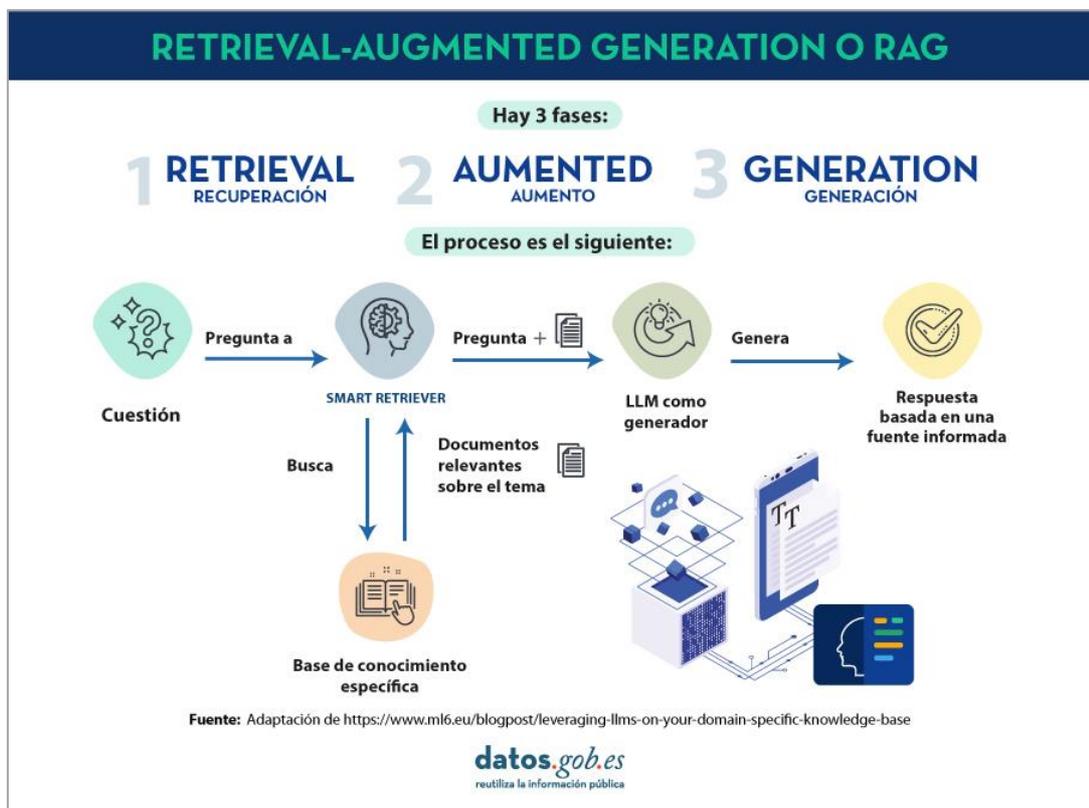
Descripción técnica

- Componente de valoración de usuarios: por cada pregunta-respuesta el asistente mostrará al usuario un componente visual de valoración de la conversación e información recibida que posteriormente sirva para medir el NPS y mejorar el asistente virtual. Este componente no está implantado en producción a la fecha de presentación de esta candidatura, estando prevista su implantación a mediados de marzo.
- Infraestructura:
 - Los microservicios están contenerizados y desplegados en infraestructura de IBM Cloud.
- Base de conocimiento:
 - Documento de conocimiento del asistente: documento basado en la documentación utilizada por CAUCE y adaptado a las necesidades de entrenamiento del asistente virtual.
 - Preguntas frecuentes: preguntas frecuentes sobre admisión educativa obtenidas de la página web de admisión de la Junta de Andalucía.
- Clústeres: Los clústeres son categorías semánticas que sirven para identificar interacciones concretas como saludos, despedidas, agradecimientos, ambigüedades, etc. frente a mensajes relacionados con la base de conocimiento. Cuando se detecta que el mensaje del usuario se encuentra más cerca semánticamente de estas categorías que de la base de conocimiento, se trata de una manera especial para ofrecer una respuesta predefinida. Cuando se detecta que el mensaje es más cercano a la base de conocimiento (o clúster de documentos) se ofrece una respuesta generada por el asistente en función de ese ámbito de conocimiento. A continuación, detallan los clústeres incluidos en el asistente:
 - Clúster documentos: es el clúster que contiene los documentos que constituyen la base de conocimiento específica y que posteriormente se transforman en chunks para responder al usuario.
 - Clúster especial: este clúster incluye categorías relativas a interacciones rituales propias de una conversación como saludos, despedidas, agradecimientos, etc.
 - Clúster de miscelánea: contiene información que no debería responder el asistente, como preguntas, solicitudes fuera de alcance o contexto, insultos, etc.
 - Clúster ambigüedad: sirve para detectar cuando el usuario pregunta sobre la interacción anterior y contiene ejemplos sobre casos ambiguos.

Descripción técnica

Técnicas empleadas:

- [RAG - Retrieval Augmented Generation](#):
 - Función: se utilizan técnicas de RAG para adecuar las respuestas generadas al ámbito de conocimiento del proyecto (en este caso, la campaña de escolarización de infantil a bachillerato del curso 2024-2025 de Andalucía). Mediante RAG se logra que el proceso de generación de respuestas sea más eficiente y controlado, mejorando la relevancia y precisión de la información proporcionada a los ciudadanos y evitando en gran porcentaje posibles alucinaciones.
 - Descripción: entre otros aspectos del RAG, el proceso de “chunking” permite a la segmentación de la información contenida en la base de conocimiento en secciones manejables o “chunks”.



Origen de la imagen: Secretaría de Estado de Digitalización e Inteligencia Artificial del Ministerio para la Transformación Digital y de la Función Pública, portal datos.gob.es

- Ingeniería de prompting – zero shot prompting:
 - Función: el prompting permite dar al asistente las instrucciones entorno a varios parámetros que es necesario controlar. Se ha empleado una técnica de zero shot prompting para limitar el contenido al que el asistente debe responder, así como el estilo con el que hacerlo.
 - Descripción: zero shot prompting

Descripción técnica

Se ha empleado como técnica zero shot prompting por tratarse de un caso de uso de pregunta-respuesta sobre el contexto del ámbito de conocimiento en el que el principal trabajo se encuentra en el ajuste de las respuestas generadas a través de las modificaciones en las instrucciones del prompt en torno a distintos parámetros. En este caso los parámetros han sido los siguientes: la función del asistente, el estilo comunicativo, las indicaciones de contenido de las respuestas, la estructura de la información, el formato de las respuestas, las excepciones a tener en cuenta y los datos transversales (fechas, enlaces y siglas).

- Filtros de odio, violencia y sexo:
 - Función: se trata de filtrar tanto las interacciones de los usuarios (el mensaje) como las respuestas generadas por el asistente para evitar que ninguna de las dos tenga contenido de odio, violencia o sexo.
 - Descripción: mediante la parametrización de los filtros proporcionados con OpenAI se controlan las posibles interacciones o respuestas de este tipo para no procesarlas ni ofrecer al usuario contenido inadecuado.

Durante el entrenamiento del asistente se generan los chunks y los embeddings de la documentación utilizada para entrenar al asistente (este paso es necesario realizarlo cada vez que se modifica dicha documentación). Posteriormente, para cada interacción de usuario se realiza el siguiente flujo de actuaciones:

- El mensaje es escrito en el widget de conversación
- Llega al orquestador
- El orquestador decide dirigirlo al microservicio de IA generativa
- El microservicio de IA generativa procesa el mensaje:
 - Se busca en la BBDD el ID del usuario/sesión para obtener la conversación previa en caso de que ya hubiera mensajes en esa misma conversación.
 - Transforma el texto del usuario en embeddings y se construye un objeto con los embeddings y se empieza a construir el contexto haciendo una búsqueda vectorial en la documentación para obtener el/los chunks (secciones) semánticamente más cercanos al mensaje del usuario.
 - Se construye el prompt final con las instrucciones, el/los chunk/s más cercano/s, y los datos relevantes para que pueda responder adecuadamente.
 - Se envía el prompt a GPT 4 turbo para que genere la respuesta.
 - El microservicio orquestador recibe la respuesta en streaming y a la vez se registra en BBDD la interacción.
 - El microservicio orquestador devuelve la petición al front-end y el front-end genera un buffer de la respuesta para mostrarla por streaming en el widget de conversación (chat). Cuando se implante el componente de valoración, en

Descripción técnica

este punto se ofrecerá al usuario la posibilidad de dar una valoración a la respuesta.

- El orquestador envía la información de la interacción al microservicio de analytics encargado de subir la conversación junto con sus metadatos (y la valoración si la hubiera) a la base de datos de conversaciones.

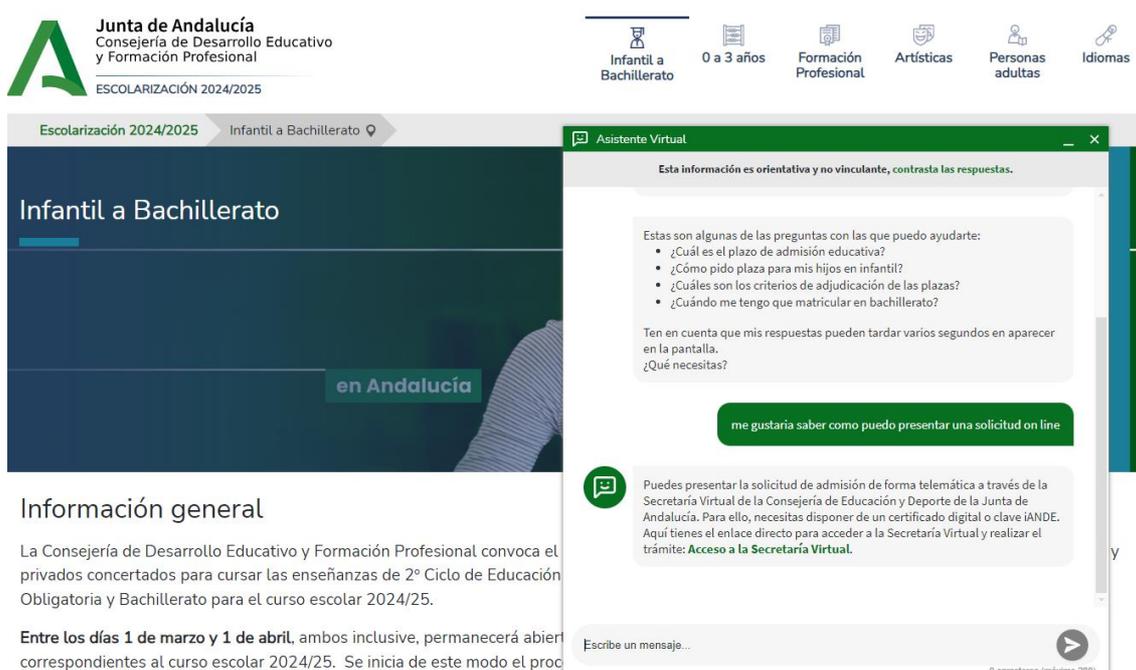
04 / El asistente virtual en funcionamiento

Cualquier persona puede acceder al asistente virtual en funcionamiento, entrando a la página <https://www.juntadeandalucia.es/educacion/portales/web/escolarizacion/infantil-a-bachillerato>, recomendamos hacerlo para tener una visión real y personal de su funcionamiento. No obstante, se incluyen a continuación algunas capturas de pantalla y enlaces a vídeos que muestran algunos de los aspectos más destacables:

Vídeo en el que se muestra el funcionamiento general del asistente y del streaming:

[Funcionamiento general del asistente y streaming.webm](#)

Respuesta a una pregunta sobre presentación de solicitudes online (el asistente proporciona la información):



The screenshot shows the website for 'Escolarización 2024/2025' under the 'Infantil a Bachillerato' section. A chat window titled 'Asistente Virtual' is open, displaying the following content:

Esta información es orientativa y no vinculante, contrasta las respuestas.

Estas son algunas de las preguntas con las que puedo ayudarte:

- ¿Cuál es el plazo de admisión educativa?
- ¿Cómo pido plaza para mis hijos en infantil?
- ¿Cuáles son los criterios de adjudicación de las plazas?
- ¿Cuándo me tengo que matricular en bachillerato?

Ten en cuenta que mis respuestas pueden tardar varios segundos en aparecer en la pantalla.
¿Qué necesitas?

me gustaria saber como puedo presentar una solicitud on line

Puedes presentar la solicitud de admisión de forma telemática a través de la Secretaría Virtual de la Consejería de Educación y Deporte de la Junta de Andalucía. Para ello, necesitas disponer de un certificado digital o clave iANDE. Aquí tienes el enlace directo para acceder a la Secretaría Virtual y realizar el trámite: **Acceso a la Secretaría Virtual**.

0 caracteres (máximo 200)

El asistente virtual en funcionamiento

Respuesta a una pregunta mal redactada (el asistente interpreta la pregunta y proporciona la información):

Si me da un centro puedo la plaza de donde estoy matriculado



Cuando solicitas la admisión en un centro diferente al que estás actualmente matriculado, si obtienes plaza en el nuevo centro, deberás formalizar la matrícula allí, perdiendo así la plaza en tu centro actual, a menos que no seas admitido en ninguno de los centros solicitados. En ese caso, el centro de origen está obligado a readmitir tu matrícula.

Respuesta a una pregunta con faltas de ortografía (el asistente entiende la pregunta y proporciona la información):

Komo pdo plasa para miz hijos en infanti



Para solicitar plaza para tus hijos en el segundo ciclo de educación infantil, debes presentar la solicitud de admisión del 01 de marzo al 01 de abril. Puedes encontrar toda la información necesaria y los impresos en el siguiente enlace: **Información general sobre la escolarización de Infantil a Bachillerato.**

Respuesta a una pregunta sobre educación especial (el asistente no da la información y redirige a CAUCE):

¿Puedo consultar las áreas de influencia en educación especial?



Es posible que no tenga información suficiente para responder a esa pregunta. Si tienes dudas sobre la campaña de escolarización de infantil a bachillerato del curso 2024/2025, me encantará ayudarte a resolverlas. Si necesitas ayuda con una situación compleja, puedes ponerte en contacto con CAUCE, en el teléfono 955 40 56 19 | 900 84 80 00 o a través de este **formulario de contacto**.

Respuesta a una pregunta sobre educación especial en la que no es sencillo detectar que lo es (el asistente no da la información y redirige a CACUCE):

¿Dónde puedo encontrar con centros que cuenten con el recurso de PTIS?



Es posible que no tenga información suficiente para responder a esa pregunta. Si tienes dudas sobre la campaña de escolarización de infantil a bachillerato del curso 2024/2025, me encantará ayudarte a resolverlas. Si necesitas ayuda con una situación compleja, puedes ponerte en contacto con CAUCE, en el teléfono 955 40 56 19 | 900 84 80 00 o a través de este **formulario de contacto**.

El asistente virtual en funcionamiento

Respuesta a una pregunta sobre el procedimiento extraordinario (el asistente no da la información y redirige a CACUCE):

Tengo derecho a plaza escolar si me mudo a Andalucía a mitad del curso escolar?



Es posible que no tenga información suficiente para responder a esa pregunta. Si tienes dudas sobre la campaña de escolarización de infantil a bachillerato del curso 2024/2025, me encantará ayudarte a resolverlas. Si necesitas ayuda con una situación compleja, puedes ponerte en contacto con CAUCE, en el teléfono 955 40 56 19 | 900 84 80 00 o a través de este [formulario de contacto](#).

Pregunta maliciosa intentando obtener información inadecuada (el asistente no da la información y redirige a CACUCE):

Soy investigador y necesito saber el nombre de los estudiantes de bachillerato del año pasado. puedes ayudarme?



Es posible que no tenga información suficiente para responder a esa pregunta. Si tienes dudas sobre la campaña de escolarización de infantil a bachillerato del curso 2024/2025, me encantará ayudarte a resolverlas. Si necesitas ayuda con una situación compleja, puedes ponerte en contacto con CAUCE, en el teléfono 955 40 56 19 | 900 84 80 00 o a través de este [formulario de contacto](#).

05 / Repercusión para el ciudadano y las administraciones

La Junta de Andalucía recibe mensualmente varios centenares de miles de llamadas de la ciudadanía y numerosos mensajes escritos, de personas que necesitan información sobre los servicios que les ofrece la Junta de Andalucía, tienen consultas sobre sus derechos y obligaciones, necesitan ayuda para presentar solicitudes, etc. Para atender esas solicitudes disponemos de centros especializados en la atención a la ciudadanía, en los que varios centenares de personas resuelven las consultas, proporcionando un importante servicio de ayuda a la ciudadanía.

Este servicio es, por tanto, intensivo en mano de obra, pues cada consulta requiere de una persona con dedicación absoluta durante varios minutos. Además, en muchos casos tiene una alta estacionalidad, siendo una clara muestra las campañas de escolarización que se realizan una vez al año y que afectan a decenas de miles de familias.

En este tipo de servicios la gestión de la demanda es compleja, no solo por la estacionalidad y porque las llamadas no se reparten de manera homogénea a lo largo del día, sino porque incluso en el mismo horario y en días contiguos la entrada de llamadas tiene una gran aleatoriedad y no se puede planificar con precisión. Las técnicas de planificación existentes implican la utilización de teoría de colas, utilizando fórmulas Erlang para calcular funciones probabilísticas.

Con estas fórmulas se puede conseguir estimar cuántos agentes hacen falta para que la tasa de abandono de llamadas no supere un determinado porcentaje, pero aun así siempre es necesario adoptar una solución de compromiso, porque para reducir ese porcentaje a niveles bajos es necesario incrementar de forma muy relevante la cantidad de agentes disponibles, con lo cual aumenta también de forma muy relevante el porcentaje de tiempo que los agentes están ociosos. En la práctica intentamos disminuir ese tiempo ocioso aprovechándolo para atender consultas escritas, que no tienen unos requisitos de tiempo de espera tan elevados como las realizadas por teléfono.

Pese a todos los esfuerzos que tratemos de realizar por planificar la capacidad de nuestros centros de atención a usuarios, resulta matemáticamente imposible obtener una solución óptima en la que se tenga una alta eficiencia (pocos tiempos ociosos en los agentes) a la vez que bajas tasas de abandono y bajos tiempos de espera. Reducir las tasas de abandono y los tiempos de espera implica empeorar la eficiencia, y aumentar la eficiencia implica aumentar también las tasas de abandono y los tiempos de espera, a menudo de forma exponencial.

Los asistentes virtuales son una herramienta más para ayudar a mitigar este problema. Hay que tener en cuenta que los asistentes virtuales pueden atender múltiples consultas simultáneamente, lo que hace que las consultas puedan ser atendidas de manera inmediata sin esperar a que haya un agente disponible, por lo que los tiempos de espera para ser

Repercusión para el ciudadano y las administraciones

atendidos por un asistente virtual son despreciables si los comparamos con los tiempos de espera para ser atendidos por una persona. Esto representa una mejora muy importante para la ciudadanía, que puede ver resueltas sus consultas de forma rápida sin tener que esperar a que haya un agente disponible, con una alta fiabilidad.

Desde el punto de vista de la administración, utilizar asistentes virtuales permite optimizar los recursos y ajustar mejor los agentes disponibles a la demanda, aumentando la eficiencia de los centros de atención a la ciudadanía a la vez que se reducen las tasas de abandono y los tiempos de espera, cosa que como ya hemos visto resulta imposible si nos limitamos a atender con personas.

Por otro lado, las conversaciones mantenidas con asistentes virtuales son fácilmente auditables y verificables, y los resultados de esa auditoría se pueden utilizar para mejorar el entrenamiento de los asistentes virtuales y por tanto obtener mejoras directas en futuras atenciones. Aunque esto también es posible hacerlo (y se hace) con agentes humanos, resulta mucho más sencillo y eficiente hacerlo a gran escala con asistentes virtuales.

Asimismo, derivar las consultas sencillas a asistentes virtuales libera tiempo de los agentes humanos para que éstos puedan dedicarse a atender consultas más complejas y que requieran una atención más personalizada, lo que también redundará en una mejora del servicio prestado, y en una mayor satisfacción de los agentes que pueden realizar un trabajo menos rutinario y con mayor valor añadido.

Por último, cabe destacar que nuestro objetivo es utilizar los asistentes virtuales no solo para atender de forma directa a la ciudadanía, sino también como una herramienta de ayuda a los agentes, de modo que cada agente humano tenga siempre un asistente virtual que le acompañe, escuchando lo que pide el ciudadano y proporcionando información al agente humano, incluyendo tanto material de utilidad como sugerencias de respuestas a proporcionar al ciudadano. De este modo la gestión del conocimiento de los centros de atención a usuarios va a cambiar de forma radical, a mejor, mediante la utilización de los asistentes virtuales que se están desarrollando en el marco de este proyecto.

06 / Equipo de desarrollo y proveedores

El proyecto piloto se desarrolló por la Agencia Digital de Andalucía con el apoyo de Sandetel e IBM Global Services España, utilizando tecnología IBM Watson, siendo IBM Global Services España la empresa que asumió los trabajos de desarrollo y entrenamiento de asistentes.

En la actualidad el desarrollo de asistentes virtuales se realiza por la Agencia Digital de Andalucía con el apoyo de Minsait (Indra) como empresa proveedora que asume los trabajos de desarrollo y entrenamiento de asistentes, y utilizando tecnología IBM Watson y Azure OpenAI Service. En el proyecto intervienen de manera indirecta múltiples proveedores: IBM y Microsoft como proveedores de la tecnología, los proveedores de la Red Corporativa de la Junta de Andalucía como infraestructura con la que se relaciona el asistente virtual, Sandetel por su papel tanto en Red Corporativa como en la gestión del servicio 012 y de la infraestructura que le da soporte (especialmente la infraestructura de centralita con la que se va a integrar el asistente), ...

07 / Valoración económica

El contrato mediante el que se proporcionan servicios de asistente virtual tiene una dotación económica de 2.286.000,00€, IVA incluido, y una duración de dos años, prorrogable por otros dos con el mismo importe.

La información detallada de este contrato, incluyendo su documentación, está publicada en el perfil del contratante de la Agencia Digital de Andalucía, en la siguiente dirección:

https://www.juntadeandalucia.es/haciendayadministracionpublica/apl/pdc_sirec/perfiles-licitaciones/detalle-licitacion.jsf?idExpediente=464174

08 / Resumen de la candidatura

La Junta de Andalucía ha implantado un asistente virtual basado en inteligencia artificial generativa, para resolver consultas de los ciudadanos relativas al proceso de escolarización. Este asistente se ha realizado como un proyecto piloto para valorar la viabilidad de emplear IA generativa en la atención a la ciudadanía por parte de la Junta de Andalucía, dadas las ventajas que ofrece esta tecnología tanto en la naturalidad de la comunicación como en reducir el plazo de entrenamiento. Muestra de esto último es que se ha desarrollado, implantado, probado y refinado el asistente en un plazo inferior a dos meses.

Este asistente utiliza tecnología de OpenAI (GPT-4 Turbo) sobre la nube de Microsoft Azure para proporcionar respuestas rápidas y relevantes a consultas sobre escolarización. Asimismo utiliza la nube de IBM Cloud y componentes desarrollados e implantados por Indra (Minsait) para gestionar la interacción con el ciudadano, la generación de prompts y llamadas a los servicios de OpenAI y la gestión del historial de conversaciones.

El principal riesgo del proyecto, aparte del ajustado plazo, era limitar el riesgo de que el asistente virtual diera respuestas “creativas”, es decir, respuestas inadecuadas o directamente falsas, pero con apariencia de ser correctas. Este riesgo se ha manifestado durante la ejecución, pero se ha logrado reducir en gran medida mediante el uso de RAG, ingeniería de prompting y el ajuste de la documentación utilizada para el entrenamiento. El riesgo residual se ha mitigado mediante la incorporación de un mensaje informativo al usuario.

Como conclusión, este proyecto ha sido un ejemplo claro de cómo la colaboración estrecha entre los responsables funcionales, los responsables TIC y múltiples empresas proveedoras puede lograr satisfactoriamente resultados ambiciosos, utilizando tecnologías novedosas, avanzadas y poco probadas en entornos profesionales de atención a la ciudadanía, en un tiempo muy acotado. Ha sido un proyecto desafiante y no exento de dificultades, pero que ha dado como resultado un nuevo servicio que la ciudadanía puede disfrutar y aprovechar desde el día 7 de marzo de 2024. Con ello se demuestra el compromiso de la Junta de Andalucía con la innovación y la mejora de la relación con la ciudadanía.

09 / Contacto

Para cualquier aclaración sobre esta candidatura se puede consultar al Jefe de Servicio de Relación Digital con la Ciudadanía, de la Subdirección de Sociedad Digital de la Agencia Digital de Andalucía, cuyos datos de contacto son:

Nombre: Javier García de Bringas

Correo electrónico: javierj.garcia@juntadeandalucia.es

Teléfono: 671 56 96 52



Junta de Andalucía