

## CAPÍTULO 16

### P-GPT: UN MODELO DE LENGUAJE DE GRAN TAMAÑO PARA LA GENERACIÓN AUTOMÁTICA DE PROPUESTAS DE TRABAJO FIN DE TÍTULO

JOSÉ LUIS LÓPEZ RUIZ\*, JUAN FRANCISCO GAITÁN GUERRERO\*,  
ANTONIO PEDRO ALBÍN RODRÍGUEZ\*\*, DAVID DÍAZ JIMÉNEZ\*, Y  
MACARENA ESPINILLA\*

*\*Universidad de Jaén; \*\*IES Las Fuentezuelas*

#### INTRODUCCIÓN

La importancia de los modelos de lenguaje de gran tamaño (LLMs por sus siglas en inglés) en la sociedad actual es innegable, ya que han transformado múltiples sectores, incluido el ámbito docente. Estos modelos, tales como GPT, LLaMA o Mistral AI han demostrado ser herramientas valiosas para tareas (Kasneci et al., 2023; Zawacki-Richter et al., 2019) que van desde la generación de textos hasta la asistencia en la enseñanza personalizada, la personalización en la tutorización y la automatización de procesos educativos en general. En un contexto en el que la personalización y la eficiencia son esenciales, los LLMs proporcionan un enfoque adaptable y escalable, facilitando la interacción de docentes y estudiantes con tecnologías avanzadas que promueven el aprendizaje autónomo y estimulan la creatividad.

En el ámbito educativo, aunque los LLMs pueden ofrecer un apoyo considerable a los estudiantes (Firat y Kuleli, 2024; Kasneci et al., 2023), este tipo de modelos también están siendo explorados como una herramienta para mejorar la enseñanza. A través de la automatización de procesos repetitivos, los docentes pueden optimizar su tiempo, permitiéndoles centrarse más en aspectos pedagógicos esenciales.

En la literatura actual, se abordan una gran diversidad de oportunidades (Abdalrazaq et al., 2023; Huber et al., 2024) que se ofrecen al docente como por ejemplo el apoyo en tareas de gestión tales como en el desarrollo de unidades didácticas o planes de estudio, el desarrollo de planes de evaluación basados en competencias, o incluso son capaces de proporcionar una retroalimentación rápida procesando los datos del estudiante. También, existen modelos (Chen et al., 2024) capaces de tutorizar al estudiante de forma personalizada, proporcionando contenido educativo y ejercicios prácticos específicos en función de sus objetivos, compromiso y comprensión de los conceptos.

Aunque los modelos LLMs generales como GPT están proporcionando resultados prometedores, la calidad de su respuesta puede verse muy limitada en tareas

específicas. Aunque existen algunas técnicas como la ingeniería de prompts para mejorar las respuestas, en casos particulares es mucho más apropiado adaptar el modelo (Li et al., 2024) con datos más especializados. A esta técnica se le conoce como fine-tuning, y nos permite no sólo especializar y actualizar el modelo, sino también mejorar problemas de sesgo (Bartl y Leavy, 2024).

En la actualidad, aunque existen diversos modelos open source tales como LLaMA, los modelos GPT tienen ciertas ventajas. La principal ventaja es que se oferta como servicio en la nube, lo que elimina la necesidad de una infraestructura propia para ejecutar modelos LLMs (González-Mayorga et al., 2024). Esto reduce el coste y la complejidad técnica, y permite una mayor escalabilidad. Además, estos modelos son los más demandados debido a la calidad de sus respuestas.

En la educación superior, algunas tareas pueden ser repetitivas, como la creación de propuestas para un Trabajo Fin de Título (TFT). Para evaluar la viabilidad de una solución que automatice este proceso y asegurar que pueda replicarse en otros contextos, es fundamental centrarse en un tipo específico de TFT y en una titulación concreta. En este caso, se ha elegido el Grado en Ingeniería Informática de la Universidad de Jaén (UJA) como ejemplo.

Por tanto, el objetivo principal de este capítulo de libro es desarrollar una herramienta que simplifique la generación de propuestas de TFT para la titulación del Grado en Ingeniería Informática, permitiendo que este enfoque también pueda aplicarse a otros grados universitarios.

A continuación, se definen las principales novedades de este capítulo de libro.

- Desarrollar una metodología innovadora para crear un modelo LLM basado en GPT, capaz de generar de manera eficiente y precisa propuestas personalizadas para TFTs específicos, adaptándose a las necesidades y requisitos académicos particulares de cada proyecto.

- Diseñar un modelo denominado P-GPT para automatizar la generación de propuestas de TFT en el Grado de Ingeniería Informática de la Universidad de Jaén. Este modelo permite optimizar el proceso de creación de propuestas, manteniendo un alto nivel de calidad y alineación con los requisitos académicos.

- Evaluación cualitativa realizada por docentes donde se destaca positivamente la capacidad del modelo P-GPT para generar propuestas coherentes, estructuradas y alineadas con los estándares académicos.

El presente capítulo de libro se encuentra dividido en las siguientes secciones. En primer lugar, se establece la estructura que sigue una propuesta TFT de manera genérica y se explica el proceso de fine-tuning para ajustar un modelo previamente entrenado a una tarea específica. A continuación, se presenta una metodología general para la generación de un modelo LLM basado en GPT-4o aplicado a la estructura de una propuesta en el Grado de Ingeniería Informática de la Universidad de Jaén.

Posteriormente, se presenta la evaluación del modelo P-GPT en términos de calidad de las propuestas generadas, realizada por un grupo de docentes. Se incluye un análisis de resultados que comprende la medición del grado de similitud entre las propuestas generadas y las humanas.

## **MATERIALES**

En esta sección, se describen la estructura de una propuesta TFT y la diversidad de estructuras que existen en el mismo. Además, se explicará el uso de la técnica de fine-tuning en el modelo GPT-4o para adaptar un modelo preentrenado a esta tarea concreta, optimizando la generación de propuestas de TFT alineadas con los estándares académicos del Grado en Ingeniería Informática.

### **Estructura de una propuesta de TFT y diversidad entre universidades**

En esta sección se describe en qué consiste una propuesta de TFT centrándonos en el caso particular de una titulación, destacando la gran diversidad de formatos y estructuras que pueden encontrarse entre distintas universidades.

Las propuestas de TFT son documentos clave en el proceso de finalización de los estudios de grado o máster. Estas propuestas proporcionan una descripción detallada del proyecto que los estudiantes deben desarrollar, estableciendo una hoja de ruta clara que incluye los objetivos específicos o la bibliografía, entre otros. Además, sirven como una guía estructurada para asegurar que el estudiante aborde el proyecto de manera coherente y alineada con los objetivos académicos de su titulación.

La diversidad en las propuestas de TFT es notable, ya que cada universidad, y en algunos casos, cada facultad, establece sus propios criterios y directrices en función de sus planes de estudio y normativas. Algunas instituciones exigen propuestas exhaustivas, con descripciones detalladas sobre los requisitos técnicos y el impacto del proyecto, mientras que otras pueden requerir un formato más simple y genérico. Esta variabilidad en las propuestas implica que la generación de estos documentos no es un proceso uniforme. En algunas facultades, se espera que los docentes personalicen cada propuesta de acuerdo a las necesidades específicas del estudiante o del tema a tratar, mientras que en otras se utilizan plantillas estandarizadas con menos margen para la personalización.

Para evidenciar esta diversidad, se ha centrado el análisis en las estructuras que se definen en la titulación del Grado de Ingeniería Informática en diferentes universidades españolas. Para ello, se han seleccionado cinco universidades diferentes: Universidad de Alicante, Universidad Rey Juan Carlos, Universidad de Granada y la Universidad de Jaén. Estas estructuras se pueden observar en la Tabla 1.

*Tabla 1.* Elementos requeridos en propuestas de TFT para el Grado de Ingeniería Informática en universidades españolas

Universidad de Alicante	Universidad Rey Juan Carlos	Universidad de Granada	Universidad de Jaén
	1. Título		
2. Resumen	2. Tecnologías Específicas	2. Breve descripción incluyendo objetivos	2. Descripción
3. Objetivos concretos	3. Descripción y objetivos	3. Conocimientos previos	3. Conocimientos previos
4. Metodología a emplear	4. Metodología y competencias	4. Hardware y software	4. Objetivos
5. Relación con asignaturas cursadas	5. Medios a utilizar		5. Metodología
6. Bibliografía	6. Bibliografía		6. Documentos a entregar

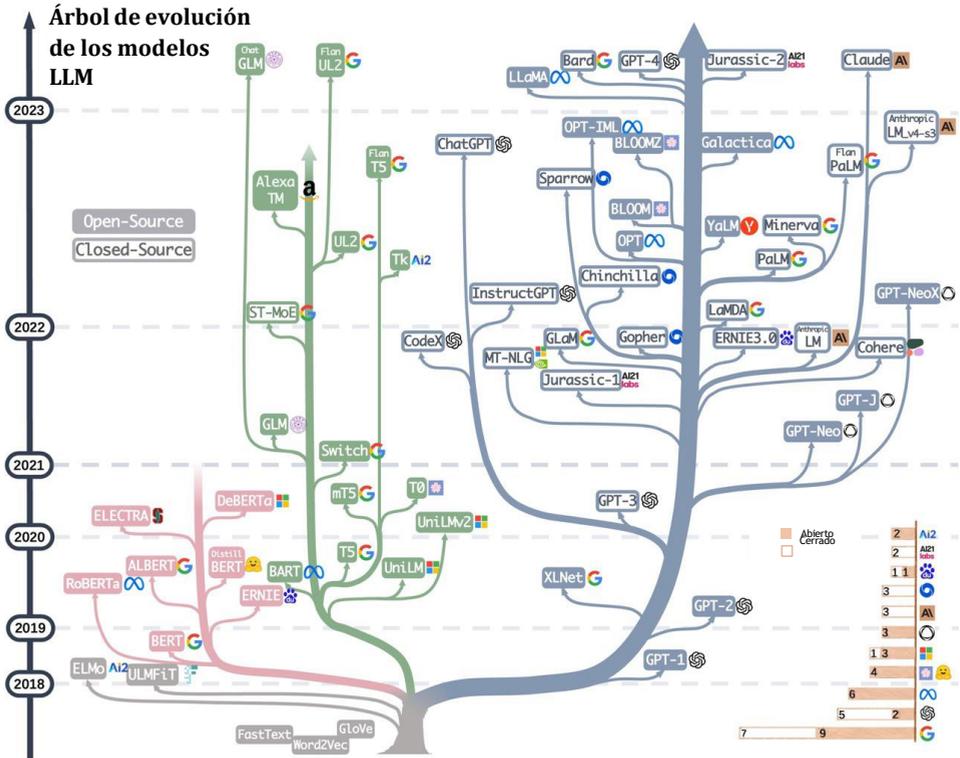
Tras realizar un breve análisis, se puede observar que todas las estructuras requeridas para la definición de una propuesta incluyen al menos un título, una breve descripción o resumen y unos objetivos. En cambio, en la estructura también se puede encontrar la definición de la metodología y los elementos requeridos para su desarrollo como haber cursado asignaturas, tecnología específica o incluso conocimientos previos en general. También, en otros casos se definen competencias, bibliografía y documentos específicos que deben ser entregados por parte de los estudiantes. De las cuatro universidades de ejemplo, la Universidad de Granada es la menos específica, debido a que no define una metodología.

Por tanto, esta diversidad supone una limitación en la generación de propuestas automáticas basadas en LLM, ya que no hay una estructura común. Por tanto, es necesario establecer métodos que permiten generar propuestas de calidad enfocadas en una titulación impartida en una institución universitaria en particular.

En este documento nos centraremos en la estructura del Grado de Ingeniería Informática de la UJA para generar propuestas automáticas de Trabajo Fin de Grado (TFG). Por tanto, es esencial definir los diferentes aspectos de una propuesta TFG para esta titulación y son los siguientes:

- Estructura: título, descripción, conocimientos previos (opcional), objetivos, metodología y documentos a entregar (obligatoria autorización para publicación en abierto e informe favorable del tutor).
- Modalidad: proyecto de ingeniería o teórico/experimental.

Figura 1. Árbol de evolución de los modelos LLM desde 2018-2023



Fuente: adaptación de Yang et al. (2023)

### Técnica de Fine-Tuning

El fine-tuning (Bartl y Leavy, 2024) es una técnica en el campo del aprendizaje automático que consiste en ajustar un modelo previamente entrenado en un conjunto de datos general a una nueva tarea específica. En lugar de entrenar un modelo desde cero, se utiliza un modelo ya preentrenado, en el que ha aprendido representaciones generales a partir de grandes cantidades de datos. En este caso particular, partiendo del modelo GPT-4o general.

El proceso de fine-tuning se basa en la idea de que los modelos preentrenados ya han capturado una vasta cantidad de conocimientos sobre el lenguaje, imágenes u otros tipos de datos en sus primeras capas. Estos modelos son capaces de identificar patrones complejos, tales como relaciones semánticas entre términos, lo cual es transferible entre tareas diversas. Al aplicar el fine-tuning, se ajustan principalmente las últimas capas del modelo, donde las representaciones específicas de la tarea se refinan, mientras que las primeras capas, que capturan características más generales, permanecen casi intactas.

Esta técnica ha ganado popularidad porque reduce la necesidad de disponer de grandes cantidades de datos específicos para entrenar un modelo desde cero. Lo que resulta de utilidad en situaciones donde los datos son escasos o costosos de obtener. Además, mejora el rendimiento de los modelos en tareas específicas y reduce el tiempo y los recursos necesarios para entrenar modelos efectivos.

## **MÉTODO**

En la presente sección, describiremos el proceso general utilizado para la generación automática de propuestas TFG mediante un modelo LLM. Así, en primer lugar, se especifica el modelo seleccionado y, posteriormente, se detallará el proceso de creación del modelo LLM específico para la generación automática de estas propuestas, con un enfoque basado en ingeniería de prompts y fine-tuning con el objetivo de adaptar el modelo a unas necesidades académicas concretas.

### **Selección del Modelo LLM**

Para la creación de un modelo capaz de generar propuestas de TFG, es necesario estudiar y analizar los diferentes LLM que se encuentran a nuestra disposición con el objetivo de seleccionar el modelo más apropiado para nuestro contexto particular.

Inicialmente, en 2018 los modelos de representaciones estáticas de palabras, como FastText, GloVe y Word2Vec, sentaron las bases para el desarrollo de los modelos de lenguaje contextuales, como BERT y GPT. Los LLMs no solo producen representaciones de palabras, sino que utilizan la arquitectura específica para procesar secuencias enteras de texto, capturando dependencias a largo plazo y generando representaciones contextuales dinámicas.

Actualmente, de todos los modelos disponibles, LLaMA es un modelo abierto que está siendo muy utilizado debido a sus grandes resultados. Sin embargo, para poder desplegar los modelos más complejos tales como LLaMA 3.1 405B o 70B, es necesario mantener una gran infraestructura que sea capaz de almacenar y ejecutarlos en tiempo real. En cambio, hay otros modelos que se encuentran como un servicio en la nube a un coste asequible y con una calidad igual o mayor, y con la posibilidad de integrarse en nuestras propias herramientas. Estos son los modelos GPT.

Por dicho motivo, en este capítulo de libro se ha seleccionado el modelo más reciente de la familia GPT, la versión 4o.

### **Personalización del prompt**

La ingeniería de prompts es un elemento esencial para mejorar la respuesta de los modelos generales LLM. Esta disciplina es especialmente relevante, debido a que nos permite diseñar adecuadamente el prompt para dirigir el modelo hacia la solución correcta de un problema.

Mediante la adecuada construcción de un prompt, es posible generar resultados alineados con los objetivos específicos del usuario. Asimismo, la ingeniería de prompts constituye una técnica fundamental para adaptar LLMs a tareas específicas, sin la necesidad de realizar un ajuste completo o exhaustivo del modelo mediante fine-tuning. Es un enfoque más ágil y flexible que permite sacar el máximo provecho del modelo sin requerir recursos computacionales adicionales.

Para diseñar este tipo de prompts se recomiendan las siguientes tareas clave:

- Definir un rol para indicar cómo debe actuar el modelo.
- Contextualización de la pregunta.
- Se debe ser muy claro y detallado.

Por tanto, inicialmente se va a diseñar un prompt con el objetivo de buscar respuestas precisas en el modelo GPT-4o general. Para ello, en base a las recomendaciones, se ha elaborado un prompt personalizado para generar respuestas enfocadas en la generación de propuestas de TFG para el Grado en Ingeniería Informática de la UJA. A continuación, se incluye el prompt que se ha diseñado.

### *Prompt propuesto*

Actúa como profesor del Grado de Ingeniería Informática en la Universidad de Jaén.

### *Objetivo*

generar una propuesta de Trabajo Fin de Grado en relación a la descripción proporcionada y a su modalidad.

### *Estructura*

la propuesta debe tener un título, unos conocimientos previos, unos objetivos, una metodología y la documentación a entregar.

### *Modalidad*

[Proyecto de Ingeniería o Teórico/Experimental].

### *Descripción*

[Breve resumen de la propuesta].

Para comprobar la efectividad de la personalización del prompt, se han analizado dos propuestas diferentes: sin personalización y con personalización.

Dada la extensión de las respuestas que han proporcionado los modelos, se ha incluido un ejemplo de cada tipo en el enlace (<https://github.com/jossellr/P-GPT->

Resultados/tree/main/Propuestas%20seccion%202.3). A continuación, en la Tabla 2 se incluye un breve análisis de las dos respuestas.

**Tabla 2.** Comparativa entre la generación automática de propuestas de TFG con y sin personalización mediante el modelo GPT-4o

Prompt		
[MODALIDAD] = Proyecto de Ingeniería	<p>Genera una propuesta de Trabajo Fin de Grado sobre el siguiente tema</p> <p>Actúa como profesor del Grado de Ingeniería Informática en la Universidad de Jaén.</p>	
[DESCRIPCIÓN] = En la actualidad las plataformas web sanitarias son una herramienta esencial para la mejora de calidad de vida en pacientes debido a que proporcionan una mejor gestión de los datos y permiten realizar tratamientos personalizados por parte del personal sociosanitario. En el ámbito sociosanitario, los pacientes con diabetes generan una gran cantidad de datos a través de dispositivos IoT de medición continua de glucosa. Por tanto, en este Trabajo Fin de Grado se propone el desarrollo de una plataforma web centralizada enfocada en el ámbito sociosanitario para gestionar datos de personas diabéticas tipo I. Esta aplicación permitirá gestionar los datos de glucosa desde diferentes roles: paciente, personal sociosanitario y familiares. Para la implementación el alumno deberá utilizar las tecnologías webs que mejor se adapten al proyecto.	<p>Objetivo: generar una propuesta de Trabajo Fin de Grado en relación a la descripción proporcionada y a su modalidad. Estructura: la propuesta debe tener un título, unos conocimientos previos, unos objetivos, una metodología y la documentación a entregar.                      Modalidad: [MODALIDAD]                      Descripción: [DESCRIPCIÓN]</p>	
	Sin personalización	Con personalización
Extensión media (palabras)	512	492
¿El texto tiene coherencia?	✓	✓
¿Define la propuesta siguiendo la estructura de la UJA?	X	✓
¿La respuesta es concisa?	X	X
¿Podría utilizarse para una propuesta real?	X	X

Las respuestas obtenidas para los dos prompt son coherentes, pero en el caso de la personalización el modelo nos ha proporcionado una respuesta con los elementos de la estructura indicada. No obstante, sigue incluyendo un resumen al finalizar la propuesta. Además, ambas respuestas incluyen objetivos que no se indican en la descripción y complican el TFG, como la incorporación de sensores de glucosa en el sistema.

Para el caso de la respuesta sin personalización, se incluyen otros aspectos como la revisión de la literatura, referencias, resultados esperados, conclusiones y trabajos futuros y, recomendaciones adicionales.

Por último, en ambos casos las respuestas son demasiado extensas, siendo la media de TFGs reales en el Grado de Ingeniería Informática de 224 palabras (sin incluir la descripción).

Tras el análisis general concluimos que el prompt con personalización funciona mejor, pero es necesario aplicar métodos adicionales para obtener resultados más concisos y de mayor calidad.

### **P-GPT: modelo fine-tuning para generar propuestas de TFG**

Tras la evaluación de los modelos generales se ha demostrado la necesidad de incorporar técnicas diferentes que nos permitan generar respuestas con mejores resultados. Por ello, se va a aplicar la técnica fine-tuning para generar un modelo más específico a partir de propuestas reales. Este modelo lo denominaremos P-GPT.

Para poder especializar al modelo, se han recogido datos del histórico de TFGs del Grado en Ingeniería Informática de la UJA disponibles en la plataforma web de la Escuela Politécnica de la Universidad de Jaén (<http://eps-anterior.ujaen.es/TFGtemporal/index.php>).

De todo el histórico de TFGs recuperados, se ha realizado un proceso de filtrado eliminando propuestas incompletas o con errores y se ha adaptado el contenido de algunas secciones. Por ejemplo, en los casos en los que se requiere en la entrega de la documentación la memoria en papel o en un disco compacto. Una vez realizado este procedimiento, se han obtenido un total de 951 propuestas de TFG para utilizarlas en el proceso fine-tuning de especialización.

Para poder especializar al modelo GPT-4o, es necesario convertir toda la información en forma de conversaciones, incluyendo ejemplos del paso de mensajes entre el usuario y el modelo. Para ello, es esencial proporcionar qué solicita el usuario y cuál es la respuesta idónea. El modelo requiere registros en formato JSON, los cuales son almacenados en formato JSON Lines. Todo el formato viene especificado en la documentación (OpenAI, 2024).

En este caso, se ha optado por entrenar al modelo con solicitudes del usuario generadas de forma automática utilizando el prompt personalizado y añadiendo como resumen del tema la propia descripción de la propuesta. Por otro lado, la respuesta está compuesta por cada uno de los elementos requeridos para cualquier propuesta del título.

Todos los TFGs procesados y transformados en registros solicitud-respuesta, se han unificado en un único archivo JSON Lines y se le ha suministrado nuestro modelo basado en un modelo GPT 4o general.

## EVALUACIÓN Y ANÁLISIS DEL MODELO P-GPT

En esta sección se presentan los resultados obtenidos en la evaluación del modelo P-GPT. Para ello, en primer lugar, se presentarán cada uno de los casos que se han definido para generar las propuestas. A continuación, se analizarán los resultados que hemos obtenido a través de cuestionarios cumplimentados por profesorado que imparte docencia en el Grado de Ingeniería Informática.

### Diseño experimental

Para poder evaluar el modelo, es necesario desafiarlo en diferentes contextos. Por ello, se han definido un total de cuatro contextos diferentes para comprobar la efectividad del modelo. Debido a los dos tipos de modalidad de TFG en la UJA, dos de las propuestas son proyectos de ingeniería y las otras dos teóricos/experimentales. A continuación, en la Tabla 3, se listan los diferentes contextos utilizados en la experimentación.

*Tabla 3.* Contextos definidos para evaluar el modelo P-GPT

Nombre del contexto	Tipo	Descripción
Contexto 1	Proyecto de Ingeniería	En la actualidad las plataformas web sanitarias son una herramienta esencial para la mejora de calidad de vida en pacientes debido a que proporcionan una mejor gestión de los datos y permiten realizar tratamientos personalizados por parte del personal sociosanitario. En el ámbito sociosanitario, los pacientes con diabetes generan una gran cantidad de datos a través de dispositivos IoT de medición continua de glucosa. Por tanto, en este Trabajo Fin de Grado se propone el desarrollo de una plataforma web centralizada enfocada en el ámbito sociosanitario para gestionar datos de personas diabéticas tipo I. Esta aplicación permitirá gestionar los datos de glucosa desde diferentes roles: paciente, personal sociosanitario y familiares. Para la implementación el alumno deberá utilizar las tecnologías webs que mejor se adapten al proyecto.
Contexto 2	Proyecto de Ingeniería	Aunque en la actualidad los sistemas de posicionamiento global (GPS) son muy precisos, en el interior de edificios, en zonas urbanas densas o en entornos con obstáculos, la señal de los satélites no llega con la calidad suficiente para garantizar una localización precisa. Por ello, en la literatura se han desarrollado sistemas de posicionamiento en interiores. Por tanto, en este Trabajo Fin de Grado el estudiante desarrollará un sistema de localización en interiores basado en la tecnología Ultra Wide Band (UWB). Para ello, se deberá estudiar la tecnología UWB, analizar las técnicas de posicionamiento basadas en esta tecnología y desarrollar un sistema capaz de localizar a una persona de forma precisa en una vivienda.

Tabla 3. Contextos definidos para evaluar el modelo P-GPT (continuación)

Nombre del contexto	Tipo	Descripción
Contexto 3	Teórico/experimental	En este Trabajo Fin de Grado el estudiante desarrollará un modelo de reconocimiento de actividades humanas de personas mayores en el hogar. Para ello, se analizarán las técnicas de reconocimiento de actividades humanas y se propondrá un modelo que permita reconocer las actividades de las personas mayores en su hogar haciendo uso de datos que provienen de sensores tales como sensores de movimiento, apertura y cierre, humedad y temperatura, entre otros. El modelo propuesto se evaluará haciendo uso de conjuntos de datos reales y se comparará con otros modelos propuestos en la literatura.
Contexto 4	Teórico/experimental	En este Trabajo Fin de Grado se propone el desarrollo de un modelo de inteligencia artificial para predecir la glucosa intersticial en pacientes dietéticos tipo I. Para ello, el estudiante deberá realizar una búsqueda bibliográfica sobre los diferentes tipos de modelos y los conjuntos de datos disponibles, para poder realizar una nueva propuesta de modelo. Para comprobar la efectividad del mismo, deberá compararlo con los modelos existentes.

Una vez definidos, el modelo P-GPT ha generado las propuestas para los cuatro contextos. A continuación, en la Tabla 4 se incluye el resultado obtenido para la propuesta número 3. Por razones de espacio, el resto de propuestas generadas junto con su prompt se encuentran disponibles en el siguiente enlace (<https://github.com/jossellr/P-GPT-Resultados/tree/main/Propuestas%20generadas>).

Tabla 4. Respuesta del modelo P-GPT al contexto número 3

Título: Desarrollo de un modelo de reconocimiento de actividades humanas de personas mayores en el hogar
Conocimientos previos
El estudiante debe tener conocimientos en las siguientes materias: <ul style="list-style-type: none"> <li>- Inteligencia artificial</li> <li>- Minería de datos.</li> </ul>
Objetivos:
Los objetivos del presente trabajo son: <ul style="list-style-type: none"> <li>- Analizar las diferentes técnicas de reconocimiento de actividades humanas.</li> <li>- Analizar diferentes conjuntos de datos para el reconocimiento de actividades humanas.</li> <li>- Diseñar y desarrollar un modelo para reconocer actividades humanas de las personas mayores en el hogar a partir de datos obtenidos de sensores instalados en un entorno inteligente.</li> <li>- Implementar el modelo y realizar pruebas y un análisis comparativo con el estado del arte.</li> <li>- Elaborar una memoria en la que se presente el trabajo desarrollado, así como los manuales de instalación de ejecución del prototipo desarrollado.</li> </ul>

**Tabla 4.** Respuesta del modelo P-GPT al contexto número 3 (continuación)

---

**Metodología:**

Para la consecución de los objetivos anteriores se propone la siguiente metodología de trabajo:

- Análisis de técnicas de reconocimiento de actividades humanas.
  - Estudio de la extracción de características de los distintos tipos de sensores utilizados en los conjuntos de datos disponibles.
  - Análisis de las diferentes aproximaciones usadas para el reconocimiento de actividades humanas y selección de la más adecuada que se utilizará en el modelo que se implemente.
  - Diseño y desarrollo del modelo propuesto. Implementación y pruebas del modelo.
  - Análisis comparativo de los resultados obtenidos por el modelo propuesto y el estado del arte.
  - Elaboración de una memoria con todos los aspectos de análisis, diseño, implementación y pruebas realizadas.
- 

**Documentación a entregar:**

El estudiante entregará una memoria final que recoja todos los trabajos desarrollados. Dicha memoria se presentará en formato digital en un fichero PDF, con la estructura habitual en estos casos y siguiendo las plantillas y guías de estilo propuestas en la EPS de la Universidad de Jaén. La memoria incluirá todos los apartados necesarios para seguir adecuadamente el proceso de desarrollo del trabajo, junto con el material adicional que sea conveniente, en forma de apéndices o anexos. Junto con la memoria se incluirá el correspondiente código del modelo desarrollado, así como de su manual de instalación y uso, con el objeto de verificar su funcionamiento. Igualmente, podrá acompañar a la memoria todo tipo de documento audiovisual que permita comprobar el funcionamiento del sistema en situaciones tipo.

---

De forma general se puede apreciar una respuesta más concisa y personalizada al contexto de la titulación, debido a la inclusión de asignaturas relacionadas con el tema y que son impartidas en el Grado de Ingeniería Informática de la UJA. Los objetivos y metodología son adecuados para la propuesta y la documentación a entregar menciona algunos detalles importantes como las guías de estilo de la Escuela Politécnica de Jaén y a diferencia del modelo general, no define la estructura de la memoria. Por último, la media de las cuatro propuestas es de 238 palabras, siendo 224 la media en el caso de las propuestas reales. Por tanto, podemos decir que en mayor medida ha conseguido emular a los ejemplos suministrados. No obstante, se analizará con más detalle a través de los resultados obtenidos a través de cuestionarios.

### **Evaluación cualitativa por docentes**

Una vez generadas las propuestas, se ha procedido a realizar un cuestionario a un grupo de docentes del Grado de Ingeniería Informática.

El objetivo de esta encuesta fue, por un lado, comprobar si creen que es útil aplicar la IA para automatizar las propuestas de TFG y, por otro lado, evaluar la validez y utilidad de las propuestas generadas automáticamente por el modelo, así como conocer la percepción de los docentes respecto a la calidad de las mismas.

El cuestionario estaba compuesto por tres preguntas iniciales para comprobar el grado de utilidad del modelo P-GPT para los docentes y ocho preguntas por cada propuesta para evaluarlas. En total, se han incluido en el cuestionario seis propuestas, cuatro generadas por el modelo y otras dos (<https://github.com/jossellr/P-GPT-Resultados/tree/main/Propuestas%20adicionales>) elaboradas por los autores del documento. El objetivo era incluir dos propuestas reales para comprobar si los docentes eran capaces de distinguir una propuesta real de una respuesta generada por IA. Las preguntas del cuestionario se encuentran reflejadas en la Tabla 5.

*Tabla 5.* Listado de preguntas y tipo de respuestas utilizadas para el cuestionario

Cuestionario inicial de automatización	
Pregunta	Respuesta
1. ¿Elaborar propuestas de TFG lo consideras una tarea repetitiva?	
2. Reutilizas parte de otras propuestas de TFG para elaborar una nueva?	Totalmente de acuerdo De acuerdo
3. ¿Utilizarías una herramienta basada en Inteligencia Artificial capaz de generar una propuesta completa (título, conocimientos previos, objetivos, metodología y documentación a entregar) en relación a una pequeña descripción del TFG?	Ni de acuerdo ni en desacuerdo En desacuerdo Totalmente en desacuerdo
Cuestionario por propuesta	
Pregunta	Respuesta
1. ¿Cree que la propuesta de TFG es clara y comprensible de forma general?	
2. Los conocimientos previos requeridos son adecuados para la propuesta	Totalmente de acuerdo
3. Los objetivos son claros y concisos	De acuerdo
4. La metodología se encuentra bien definida y tiene coherencia	Ni de acuerdo ni en desacuerdo En desacuerdo
5. Los documentos y entregables son suficientes	Totalmente en desacuerdo
6. Considera que las propuestas de TFG son viables para ser desarrolladas por estudiantes de grado en el tiempo estipulado?	
7. ¿Qué nivel de dificultad cree que tiene la propuesta generada?	Muy difícil Difícil Normal Fácil Muy fácil
8. Cree que esta propuesta ha sido generada por Inteligencia Artificial (IA) o por un humano?	Generada por IA Generada por un humano No estoy seguro

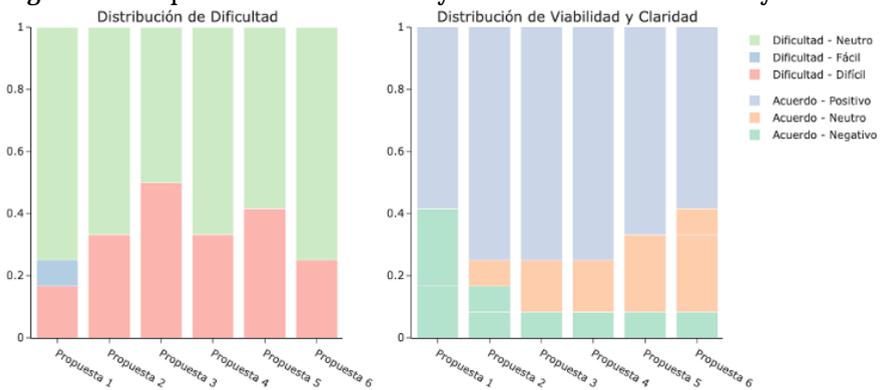
### 4.3 Análisis comparativo

A continuación, se analizan los resultados obtenidos de la encuesta donde han participado un total de 12 docentes que imparten docencia en el Grado de Ingeniería Informática de la Universidad de Jaén.

De acuerdo con el cuestionario inicial, cabe manifestar como la viabilidad de la metodología presentada en este documento es apoyada por los encuestados a través de la percepción de repetitividad en la elaboración de TFG, así como su alineación con las casuísticas de reutilización de material y de confianza frente a la utilización de IA para la generación de propuestas. En concreto, se ha obtenido que el 75% de los encuestados está a favor, mientras que el 8,33% se decanta por métodos tradicionales y no considera de valor la proposición de este tipo de herramienta; el 16,7% restante postula una percepción neutral respecto a los beneficios y limitaciones de esta práctica.

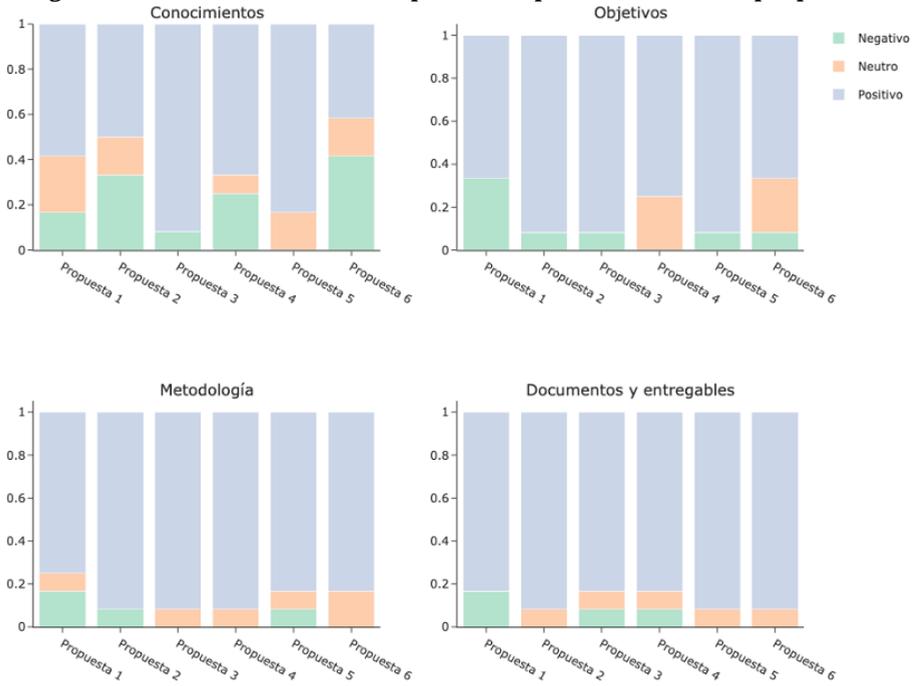
Con el objetivo de evaluar la calidad de las propuestas, el análisis de las preguntas 1-6-7 del cuestionario por propuesta, desgrena equilibrio entre dificultad y normalidad en las propuestas presentadas, mientras que hay un consenso generalmente positivo entre la claridad y viabilidad de las mismas (Figura 2). Nótese que las propuestas 1-4 se corresponden con aquellas generadas por IA, mientras que las propuestas 5 y 6 hacen referencia a aquellas generadas por un docente.

Figura 2. Comparación de dificultad y acuerdo sobre viabilidad y claridad



De forma más específica (véase Figura 3), se tiene que la adecuación de los conocimientos previos en las propuestas generadas por IA concentra opiniones de acuerdo. Mientras tanto, las propuestas humanas presentan una mayor variabilidad, por lo que se advierte menos uniformidad en este tipo de propuestas al ser susceptibles de criterios individualistas. Sin embargo, en la claridad de los objetivos, a pesar de una positiva aceptación en general, existe cierta discordancia con una de las propuestas generadas por IA, poniendo de nuevo en manifiesto criterios de precisión diferenciados. Por su parte, la metodología y los documentos y entregables de las propuestas son aceptados en mayor grado por los encuestados.

Figura 3. Acuerdo sobre los componentes que conforman las propuestas



Adicionalmente, se ha evaluado la capacidad de los encuestados para discernir entre aquellas propuestas generadas por IA y las generadas por un docente (véase Tabla 6). En general, los encuestados han reflejado que la IA ha simulado de manera bastante aproximada las propuestas que generaría un docente, puesto que los encuestados se han decantado por la generación humana en un 47% de media, mientras que en un 24% de las veces han dudado de la procedencia de la propuesta; únicamente se ha percibido como IA en el 29% de las ocasiones. En relación a esto último, puede observarse cómo las propuestas generadas por IA presentan un acierto relativamente bajo, a excepción de la propuesta 4 que sí presenta un mayor porcentaje de encuestados que la han detectado, además del 25% que han manifestado duda; las propuestas 2 y 3 muestran cómo los encuestados las perciben, en un alto porcentaje, como generadas por un humano. Por su parte, las propuestas generadas realmente por un docente presentan cierta variabilidad, mientras que la propuesta 5 es claramente correctamente detectada, la propuesta 6 ha generado un porcentaje de duda superior.

Tabla 6. Porcentaje de las respuestas asociadas a la pregunta 8 del cuestionario de propuestas

	Generada por IA	Generada por Humano	Ns/Nc
Propuesta 1	0.33	0.33	0.33
Propuesta 2	0.17	0.58	0.25
Propuesta 3	0.17	0.67	0.17
Propuesta 4	0.58	0.16	0.25
Propuesta 5	0.16	0.75	0.08
Propuesta 6	0.33	0.33	0.33
Media	0.29	0.47	0.24

Consecuentemente, los resultados del cuestionario inicial demuestran el apoyo significativo mostrado por parte de los encuestados hacia la metodología propuesta. Sin embargo, se aprecian notables diferencias en la percepción por parte de los docentes en el momento de comparación de las propuestas con diferentes procedencias. Asimismo, a pesar de que este estudio positivamente sugiere cómo la IA puede imitar en gran medida el estilo humano, se considera cómo diferentes aspectos en el proceso de especialización del modelo P-GPT pueden requerir de ajustes futuros para lograr una alineación completa en términos de precisión y subjetividad, no sin olvidar la subjetividad inherente al criterio humano.

## CONCLUSIONES

Este capítulo de libro ha presentado un innovador proyecto para la generación automática de propuestas de TFT mediante el uso de Inteligencia Artificial, con el objetivo de asistir a los docentes en su elaboración. En primer lugar, se ha logrado una metodología general aplicable a la generación de propuestas de TFG, ejemplificada en el Grado de Ingeniería Informática de la Universidad de Jaén, pero adaptable a otras titulaciones.

Como segundo logro, se ha creado un modelo de lenguaje de gran escala denominado P-GPT, basado en GPT-4o y entrenado mediante fine-tuning con prompts personalizados. La evaluación del modelo por parte de docentes universitarios ha demostrado resultados positivos en cuanto a la calidad de las propuestas generadas, con la excepción del apartado de conocimientos previos, que requiere mejoras. El estudio comparativo entre propuestas generadas por Inteligencia Artificial y por docentes ha revelado la dificultad de los docentes para distinguir entre ambas, lo que subraya la calidad del contenido generado por P-GPT.

Los resultados sugieren una tendencia positiva hacia la adopción de la Inteligencia Artificial Generativa en contextos educativos, aunque se enfatiza la necesidad de evaluaciones continuas para mejorar su implementación. Es importante destacar que esta herramienta no pretende reemplazar la labor docente, sino simplificarla, ofreciendo propuestas iniciales que pueden ser modificadas y adaptadas

según las necesidades específicas. El proyecto demuestra el potencial de la Inteligencia Artificial Generativa para optimizar los procesos académicos, al tiempo que subraya la importancia del criterio y la experiencia del docente en la supervisión y personalización final de las propuestas.

## REFERENCIAS

Abd-alrazaq, A., AlSaad, R., Alhuwail, D., Ahmed, A., Healy, P. M., Latifi, S., Aziz, S., Damseh, R., Alabed Alrazak, S., y Sheikh, J. (2023). Large language models in medical education: Opportunities, challenges, and future directions. *JMIR Medical Education*, 9, e48291.

Bartl, M., y Leavy, S. (2024). From 'showgirls' to 'performers': Fine-tuning with gender-inclusive language for bias reduction in LLMs.

Chen, E., Lee, J.-E., Lin, J., y Koedinger, K. (2024). GPTutor: Great personalized tutor with large language models for personalized learning content generation. En *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (Vol. 13, pp. 539–541). ACM.

Firat, M., y Kuleli, S. (2024). GPT vs. Google: A comparative study of self-code learning in ODL students. *Journal of Educational Technology and Online Learning*, 7(3), 308–320.

González-Mayorga, H., Rodríguez-Esteban, A., y Vidal, J. (2024). El uso del modelo GPT de OpenAI para el análisis de textos abiertos en investigación educativa. *Pixel-Bit. Revista de Medios y Educación*, (69), 227–253.

Huber, S. E., Kiili, K., Nebel, S., Ryan, R. M., Sailer, M., y Ninaus, M. (2024). Leveraging the potential of large language models in education through playful and game-based learning. *Educational Psychology Review*, 36(1).

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., y Kasneci, G. (2023). ChatGPT for good.

Li, Q., Fu, L., Zhang, W., Chen, X., Yu, J., Xia, W., Zhang, W., Tang, R., y Yu, Y. (2024). Adapting large language models for education: Foundational capabilities, potentials, and challenges.

OpenAI. (2024). *Guía para aplicar Fine-tuning en modelos GPT*. Recuperado de <https://platform.openai.com/docs/guides/fine-tuning>

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Yin, B., y Hu, X. (2023). Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond. <https://doi.org/10.48550/arXiv.2304.13712>

Zawacki-Richter, O., Marín, V. I., Bond, M., y Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? *International Journal of Educational Technology in Higher Education*, 16(1).