

Informe inicial TFG:

En el panorama actual del mundo de las publicaciones científicas, nos encontramos con varios problemas a la hora de publicar un artículo nuevo en una revista y que sea aceptado por la comunidad, para empezar se necesita pagar una suma de alrededor de 250€ en lo que se conoce como Article Processing Charges únicamente para su publicación, lo cual no es necesariamente un problema puesto que para publicar y mantener en la red estos artículos se tiene un coste obligado asociado. Ahora bien, el problema principal surge cuando el autor de dicho artículo pretende publicar en una revista de alta categoría donde independientemente de la calidad del trabajo se cargan sumas más de 10 veces mayores, donde además entran en juego otros peligros como la negativa o el ser el primero en publicar la idea o investigación al tardar incluso meses en obtener una resolución.

El impacto de un artículo se define la mayoría de las veces con el paso del tiempo, según factores como por ejemplo las citas o referencias que se le haga, pero en el momento de la publicación el salir en una revista de alta categoría o no puede marcar enormemente la validación de la comunidad científica. La alternativa de publicar en otras plataformas como Diamond Open Access cuyos costes son mucho menores no tiene el mismo valor pues se enfrenta a diversos problemas como la ausencia de editores que revisen el trabajo, la jerarquía según la calidad del trabajo o el prestigio de la propia plataforma en la comunidad.

El objetivo de este TFG es buscar una posible solución a toda esta serie de problemas apoyándose en la Inteligencia Artificial, si bien crear y entrenar un modelo desde cero sería una cantidad de trabajo que sobrepasa las capacidades del proyecto, se pueden aprovechar la variada cantidad de modelos de lenguaje grande “LLM” existentes y ya entrenados y altamente complejos para predecir el posible impacto a futuro de un artículo científico en el momento de su publicación basándose en el *abstract*.

Para ello el proceso ideado inicialmente consistirá en los siguientes pasos: Probar varios LLMs online utilizando indicaciones manuales para evaluar hasta qué punto puede generar respuestas adecuadas para predecir el factor de impacto (IF). Analizar el efecto del *prompt engineering*. Definir si es más adecuado especificar un área científica de interés. Identificar un conjunto de artículos de prueba y recuperar los datos del factor de impacto correspondientes, entre las posibles fuentes están Web of Science o Scopus. Explorar opciones para instalar y modificar localmente los LLMs. Considerar enfoques como la Generación Aumentada por Recuperación (RAG) y el ajuste o reentrenamiento del modelo.

Los tiempos adecuados a cada una de las ideas son ajustables pues la adecuación de cada estrategia de cara al objetivo es impredecible en el momento actual, idealmente gran parte del cuerpo será desarrollado en el LLM en local probando qué estrategia será la mas adecuada.

Bibliografía:

ECS. «For-Science or For-Profit?» Accedido 10 de marzo de 2025.
<https://www.electrochem.org/for-science-or-for-profit>.

