**Agenciamiento de una ruta estratégica de investigación en la FCSH**

**Documento diagnóstico - Antecedentes**

**Elaborado por:** Cristian Montoya Garcés, estudiante de sociología de la Universidad de Antioquia.

**1. Contexto del uso del la IA y porque es importan**

La integración de la Inteligencia Artificial (IA) en las investigaciones en ciencias sociales representa tanto un desafío como una oportunidad para comprender los fenómenos sociales en el mundo contemporáneo. Al emplear algoritmos para la gestión, procesamiento y estructuración de grandes volúmenes de datos, la IA, en particular los modelos de lenguaje, posee el potencial de identificar patrones y entidades que facilitan la conversión de estos datos en información relevante para la investigación. Sin embargo, este avance también plantea un desafío, ya que requiere que las ciencias sociales, con todas sus disciplinas, lideren esta implementación ajustando sus métodos de investigación tradicionales para maximizar los beneficios y minimizar los riesgos asociados al uso de IA en su ejecución.

**2. Fuentes**

Dentro de las fuentes que se utilizaron se encuentra los los proyectos de investigación de los grupos de investigación de la Facultad de Ciencias Sociales y Humana de las Universidad de Antioquia. Para ello, se identificaron 128 proyectos identificados, de 2014 al 2023, y se descargaron un total de 106. cada uno de estos archivos se descargó en formato PDF y fueron etiquetados con nombre del proyecto, año y grupo de investigación.

**3. Criterios de selección**

El criterio para seleccionar los proyectos a descargar, se acordó solo integrar como fuente aquellos proyectos que no estuvieran bajo el estado de rechazado. Adicionalmente, también se tuvo en cuanta la temporalidad (2014-2023) y el estado de los datos dado que algunos no tenían la información completa en los archivos.

**4. Hablar sobre el procedimiento**

El procedimiento para llevar a cabo el uso de técnicas computacionales e inteligencia artificial en la búsqueda de entidades académicas se dividió en cuatro fases fundamentales. En primer lugar, se realizó la descarga de los datos desde la plataforma universitaria designada para esta tarea. Los proyectos seleccionados fueron obtenidos de la base de datos del CISH, específicamente de la sabana de proyectos, y fueron debidamente renombrados con el año, la etiqueta del grupo de investigación y el título del proyecto. Posteriormente, se llevó a cabo la limpieza y reestructuración de los datos utilizando el lenguaje de programación Python, el cual proporciona herramientas para el procesamiento del lenguaje natural (NLP). Esto incluyó la extracción de la información de los archivos PDF utilizando la librería Pdfminer para convertirlos en archivos de texto (txt), el uso de expresiones regulares para identificar patrones o palabras clave y la utilización de la librería Pandas para organizar los datos en un formato tabular similar al de los archivos Excel, lo que permitió una manipulación más eficiente de los mismos. A continuación, se procedió a la construcción de los criterios de búsqueda, tarea llevada a cabo por un equipo interdisciplinario de profesores de todos los departamentos de la facultad. Estos criterios fueron diseñados para dirigir el trabajo de la IA en el reconocimiento de las entidades académicas, asegurando que se incluyeran las diversas formas de nombrarlas en cada área para obtener resultados óptimos. Por último, se desarrolló un algoritmo que siguiera un flujo de trabajo definido, el cual recibía el prompt y los datos y proporcionaba una o varias de las entidades solicitadas. Este algoritmo se construyó utilizando Python y la API de OpenAI con el modelo GPT-3.5, permitiendo así la integración efectiva de la inteligencia artificial en el proceso de búsqueda de entidades académicas.

**5. Resultados**

Los resultados fueron un conjunto de entidades que hacen referencia a elmentos propios de la investigación en ciencias sociales como la temporalidad, el objeto de la investigación, el territorio, y su población. Cada una de estas cuatro entidades, se puso en un un archivo der Excel junto la información respectiva del proyecto. Esta información, luego sería analizada por el equipo de profesionales en ciencias sociales, para su posterior validación y uso en la construcción de categorías y sub-categorías.

**6. Dificultades**

La primera dificultad, es el acceso a la información. Si bien, la Vicerrectoría de Investigación se supone que cuenta las bases de datos con la información de los proyectos, no hubo una respuesta positiva para la obtención de los mismos y esto constituyó un problema debido a que gran parte del tiempo se gastó en la descarga y etiqueta de los proyectos de forma manual. Otra dificultad, es el tema de los recursos destinados para dicha tarea. Dentro del proceso del desarrollo de los algoritmos, no se contó con el tiempo suficiente para entregar un producto mas diversificado y escalable a futuro. El estudiante, bajo su convenio de pasante, contaba con horas limitadas para desarrollar con mas detalle dicha labor, y esto dificultó el desarrollo investigativo, convirtiéndose en factor a tener en cuenta en futuros proyectos. Teniendo en cuenta su contraprestación monetaria, la contratación de estudiantes, bajo la modalidad de pasantes o auxiliares administrativos, si bien aporta al proceso formativo del mismo, también posibilita generar espacios de precarización del trabajo.