Francisco Krauss y Natalí Schejtman

Durante mayo y agosto de 2025 realizamos junto a un equipo de estudiantes dos ejercicios exploratorios de recolección de datos para conocer algunas pistas de la calidad informativa de diversos ChatBots a la hora de brindar noticias.

Junto con Chequeado y basados en estudios preliminares, así como en criterios de calidad en las noticias establecidos por la organización de Fact-Checking, seleccionamos una serie de temáticas coyunturales nacionales e internacionales –desde el caso Libra hasta la baja de impuestos a celulares, el conflicto en el Garrahan o el conflicto entre Estados Unidos e Irán– y juntamos a un grupo de estudiantes que dos días en mayo y en junio –a modo de pilotos– y tres días durante junio les preguntó a ChatGPT, Grok y Gemini por 5 noticias relevantes del día o del día anterior. Establecidos los temas de los que había que preguntar –comunes a todos– cada estudiante definía qué pregunta hacía –con qué palabras–. También se definió que hubiera preguntas que contenían algún tipo de error, para ver si los ChatBots lo corregían.

Junto al equipo, elaboramos y testeamos una matriz de 10 variables de análisis de contenido, acorde a los lineamientos de Krippendorff (1990), y las desarrollamos en un libro de código o codebook. Las variables se agrupaban en los siguientes temas: precisión, claridad, fuentes, editorialización y extras.

Si bien se trata de un ejercicio con datos insuficientes, apuntamos algunas conclusiones parciales, que podrían arrojar hipótesis interesantes para investigar en un análisis de contenido más sistemático que genere mayor volumen de datos para analizar, así como considere otras normas relevantes del método. Estas son:

- Los ChatBots suelen dar respuestas generales correctas, aunque hay una proporción de errores muy variable entre los distintos codificadores, probablemente dado su carácter personalizado y la diversidad de preguntas, aunque sea sobre temáticas comunes. Un error más frecuente es la confusión respecto de los años, por ejemplo preguntar por una elección inminente y que responda con una de hace dos años (especialmente si las preguntas no están situada con precisión)
- Los ChatBots tienen una relación variable con las fuentes, no solamente en términos de calidad sino, y especialmente, en términos de cantidad. Hay respuestas que citan 25 fuentes y otras ninguna. Para algunos de los codificadores, Grok brindó considerablemente más fuentes que el resto y Gemini menos, pero no es generalizable. Las fuentes tienden a ser de medios reconocidos y legítimos, si bien no siempre apoyan lo que dice la respuesta, hallazgo frecuente entre los codificadores.

- Los ChatBots no siempre corrigen una pregunta que contiene un error. Cuando la pregunta es incorrecta, en los casos en que no la corrije, puede derivar en una respuesta incorrecta. Un ejemplo de esto se dio con preguntas que buscaban saber por qué NO se utilizaría el sistema D'Hont en las elecciones legislativas de octubre o por qué Argentina no había cumplido con las 30 metas del FMI.
- En general, las respuestas de los ChatBots son imparciales. Fue muy poco frecuente encontrar expresiones de opinión o adjetivos vinculados a la editorialización. Es probable que eso se vincule al tipo de pregunta que se hace.
- Hay una variación considerable en los resultados de los distintos codificadores. Si bien esto podría trabajarse calculando el alpha de Krippendorf que indica el nivel de concordancia esperado entre los codificadores, también podría responder a la naturaleza personalizada y dialógica de los ChatBots.

Para una futura investigación de análisis de contenido, recomendamos, a partir de este ejercicio, estandarizar las preguntas –para poder analizar con más precisión cómo funciona la personalización de la respuesta– pero a la vez incluir distintos tipos de preguntas, con y sin errores, y centrarse en las fuentes, tanto en la calidad como en la cantidad y su vinculación real con la respuesta que brinda. El aspecto de la personalización y de la variabilidad entre codificadores –una novedad respecto de los medios informativos no personalizados–, pueden ser dos aspectos muy productivos para un análisis de contenido que recolecte mayor volumen de datos de manera sistematizada. En términos teóricos, sería interesante repensar, a partir del análisis de chatbots informativos personalizados, nociones clásicas del análisis de medios masivos como el *framing* (el encuadre noticioso, en el que subyace parte de la reconstrucción de la realidad que realizan los medios de comunicación al contar una noticia) o el *standing* (teoría que busca conocer el rol que ocupan las fuentes citadas en el contenido periodístico y qué fuentes o voces están ausentes).

Krippendorff, K. (1990). *Content analysis: An introduction to its methodology*. Sage Publications