

PrevenIA: Un agente conversacional para la información sobre el suicidio

Don Pablo Ascorbe Fernández 

paascorb@unirioja.es

Dpto. de Matemáticas y Computación, Universidad de La Rioja

Palabras clave: Chatbot, Suicidio, Generación Mejorada por Recuperación, Español

Introducción

El suicidio se sitúa como la principal causa de muerte por factores externos en España, con 4227 casos registrados en 2022, lo que supone una media de 11 muertes diarias [1]. Además, se cree que cada suicidio consumado va acompañado de aproximadamente 20 intentos, por cada intento, 14 personas han contemplado el suicidio; a su vez, al menos 6 supervivientes del fallecido sufren el impacto directo de la pérdida [2]. Estas estadísticas subrayan por qué la Organización Mundial de la Salud identifica el suicidio y los intentos de suicidio como una pandemia global, instando a todos los estados miembros a dar prioridad a su mitigación [2].

Con el reciente auge de los agentes conversacionales y su potencial para proveer información, surge PrevenIA, un robot de cibercharla enfocado en informar sobre el suicidio a familiares o amigos cercanos a personas con ideación suicida. No obstante, es necesario que toda respuesta generada por el sistema sea apropiada, por ello, la estructura interna es un sistema de Generación Mejorada por Recuperación, comúnmente conocido como *Retrieval Augmented Generation* (RAG) [3]. Estos sistemas extraen o recuperan la información de un corpus documental que contiene esencialmente todo el contenido apto para ser respondido por el agente.

Este proyecto cuenta con el apoyo de la subención PID2020-115225RB-I00 financiada por MCIN/AEI/ 10.13039/501100011033, y con los fondos para las estrategias 2023 del Ministerio de Sanidad que fueron aprobados en el CISNS de fecha 23 de junio de 2023 como apoyo a la implantación del Plan de Acción de Salud Mental

Estructura del sistema

PrevenIA es un sistema compuesto por tres capas. Las dos primeras son filtros que utilizan modelos de clasificación, primero para determinar si la pregunta está relacionada con el tema y después para comprobar si el usuario está buscando información o presenta riesgo y es necesario redirigirlo a los recursos que le puedan prestar la ayuda necesaria. La última capa es el sistema RAG, compuesto por dos módulos, uno para determinar los contextos más próximos a la pregunta recibida por el usuario y que contienen la información para responderla y otro, que es el que contiene el modelo de lenguaje, para generar la respuesta esperada.

Evaluación

La evaluación es una parte fundamental del desarrollo de cualquier sistema, con agravante en un contexto sensible como es el de la salud mental. En consecuencia, PrevenIA ha sido evaluado manual y automáticamente. La evaluación manual ha sido llevada a cabo por especialistas en el suicidio, evaluando las respuestas generadas por múltiples modelos del lenguaje y creando un banco de datos preparado para los sistemas pregunta/respuesta enteramente en español. La evaluación automática se ha realizado utilizando métricas como BertScore o Rouge y modelos de lenguaje como evaluadores, todo ello contra el banco de datos mencionado previamente. En general, cada método de evaluación tiene sus ventajas e inconvenientes y coinciden en cuál es el mejor modelo de entre los posibles candidatos.

Trabajo futuro

Quiero terminar esta relación evidenciando que todavía queda mucho por hacer. Los cimientos van tomando forma, no obstante, falta trabajo hasta que la aplicación esté disponible para su despliegue en usuarios finales. Muchos son los aspectos necesarios que hay que ajustar o desarrollar, como pueden ser: incorporar una memoria a corto-medio plazo para entender mejor las preguntas, entrenar modelos de lenguaje con los documentos validados o desarrollar una aplicación web donde alojar la herramienta, entre otros.

Bibliografía

- [1] Instituto Nacional de Estadística. Defunciones según la causa de muerte año 2022. Technical report, 2023.
- [2] WHO. Suicide worldwide in 2019: global health estimates, 2021.

- [3] Patrick Lewis et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.