Académicos de Informática UV y UAI desarrollan herramientas de IA para extraer información confiable de encuestas abiertas

Un grupo de académicos de la Escuela de Ingeniería Informática de la Universidad de Valparaíso y de la Universidad Adolfo Ibáñez, liderados por el investigador Enrique Canessa, se encuentran desarrollando herramientas de la inteligencia artificial (IA) y de machine learning, con el objetivo de extraer información confiable de encuestas abiertas, donde las personas se expresan libremente, sin límites ni opciones prefijadas.

Con financiamiento de la Agencia Nacional de Investigación y Desarrollo (ANID) y del proyecto Fondecyt 1240075, se encuentran investigando este desafío. Así lo explica el profesor Canessa: «Las encuestas están por todas partes, pero marcar una alternativa no siempre cuenta lo que realmente pensamos, nuestro proyecto usa inteligencia artificial para escuchar mejor».

«Casi cada semana aparece una encuesta nueva. La última de Cadem, del pasado 7 de octubre, indica que siete de cada diez chilenos tienen un temor alto de ser víctima de un delito. Y la encuesta CEP de abril-mayo mostró que la delincuencia es, con cincuenta y siete por ciento, lo que más se menciona entre los principales problemas del país. Son datos valiosos para orientar decisiones. Pero hay una pregunta incómoda que casi nunca nos hacemos: ¿Sabemos qué quiso decir cada persona con su respuesta?», señala.

El problema de las alternativas o respuestas cerradas

Según el investigador «para procesar miles de respuestas rápidamente, las encuestadoras suelen usar preguntas cerradas: la persona elige entre las opciones que la propia encuestadora le entrega. Es práctico y fácil de procesar, pero no siempre es la mejor forma de entender un tema a fondo».

«Tomemos una pregunta clásica: "¿Se siente seguro en su barrio?" Para una persona, sentirse seguro es que no le hagan un portonazo al llegar a casa. Para otra, poder volver caminando de noche desde el paradero. Para una tercera, que no haya bandas delictuales en la esquina. Las tres pueden responder "no", y en la tabla de resultados quedan sumadas como si dijeran lo mismo. Hasta la propia Cadem lo deja entrever: los delitos que más temor generan van desde el robo con violencia (35%) y las encerronas (30%) hasta los homicidios (27%). Son miedos bien distintos bajo una misma palabra», indica.

Al respecto, explica que «lo mismo pasa con conceptos como "educación de calidad": un estudiante, una profesora y un político probablemente imaginan cosas diferentes. Y ninguno tiene la "verdad" del asunto. Si queremos conversar en serio sobre un tema, primero hay que saber qué entiende cada persona».

«La alternativa son las preguntas abiertas: en vez de marcar una casilla, cada persona explica con sus propias palabras qué significa para ella "sentirse seguro en su barrio". Es la mejor manera de profundizar, pero tiene su dificultad: aparecen miles de respuestas distintas y alguien tiene que entenderlas y resumirlas. Ahí podría ayudar la inteligencia artificial (IA), aunque no es una varita mágica: todo modelo de IA mira el mundo con ciertos "lentes" y puede introducir sesgos en su análisis», explica.

El proyecto

Sobre la iniciativa cuenta que «llevamos varios años investigando este desafío. Partimos de una idea clave: muchos conceptos son “difusos”, es decir, no tienen bordes claros. Cuánto es “harto” o cuánto dura “un ratito”: cada persona lo entiende a su manera, y está bien. Con “seguridad” o “educación de calidad” pasa lo mismo».

«Por eso hemos desarrollado herramientas de IA y machine learning que permiten extraer información confiable de encuestas abiertas, donde las personas se expresan libremente, sin límites ni opciones prefijadas. En este punto surgen dos preguntas difíciles: ¿A cuántas personas preguntar y cuándo ya entendimos? «, añade.

«La estadística tradicional responde cuántas personas hay que encuestar para que una diferencia sea realmente significativa; por ejemplo, para saber si de verdad son más quienes dicen “no me siento seguro” que quienes dicen “sí”. Nuestra duda va más allá: ¿Hemos escuchado suficientes personas para captar todos los significados de sentirse seguro?», señala.

Para el académico «es como armar un puzle sin ver la caja: ¿Cuántas piezas necesitamos para saber qué imagen hay? Con modelos matemáticos hemos logrado responder, de forma aproximada, dos cosas: Antes de encuestar: cuántas personas conviene entrevistar, según la cantidad de respuestas distintas que esperamos y otros parámetros. Después de encuestar: si las respuestas recogidas realmente nos dan una comprensión adecuada del concepto, o si todavía nos falta escuchar a alguien».

Para qué sirve

El investigador asegura que «estas herramientas ayudan, por un lado, a la ciencia: a estudiar cómo se organiza nuestra memoria de los significados (memoria semántica: cómo guardamos las palabras y conceptos en la cabeza, y cómo los recordamos y usamos). Por otro, tienen un uso muy práctico: entender más

profundamente lo que la gente quiere decir en encuestas de respuestas abiertas».

«La próxima vez que las personas lean un titular del tipo “7 de cada 10 chilenos…”, vale la pena preguntarse: ¿Qué estaba pensando cada uno de esos 7 cuando respondió?», enfatiza.