

La compañía admitió en una nota oficial hace apenas una semana que “expertos del lenguaje” escuchan el 0,2% de las conversaciones que mantienen los usuarios con su asistente virtual tras una filtración de audios en idioma neerlandés a una televisión belga. Lo hizo sin especificar en qué lenguas se realizan las escuchas, aunque según ha podido saber llevan años realizando esta práctica con grabaciones privadas en diferentes idiomas, entre ellos español y portugués. Google España no ha querido hacer aclaraciones al respecto y se remite al comunicado de la firma.

Google Assistant. Es decir, para que el asistente sea capaz de entender mejor lo que los usuarios hablan. Esta labor despierta algunas dudas especialmente por la fuga que puede haber de datos sensibles. Sobre todo cuando se conocen las condiciones de trabajo de los transcriptores, al menos de buena parte de ellos: escuchan unas 5.000 grabaciones a la semana por revisor, no son empleados de Google, sino de una empresa tercera, tienen un contrato por obra y servicio o incluso de freelance y difícilmente llegan a ganar el salario mínimo interprofesional.

Según explican los trabajadores consultados, los proyectos para revisar lo que entiende el software o para transcribir desde cero el audio del usuario (cuando le habla a Google Assistant) van cambiando: los hay de meras búsquedas en Google, de Google Maps, mensajes e incluso peticiones a Google Home. Estos expertos en idiomas son empleados de una empresa tecnológica subcontratada por Google, con la que tienen que firmar un contrato de confidencialidad. Entre ellos, los hay que todavía trabajan en este tipo de audiciones (hasta la semana pasada, cuando saltó el escándalo y todos los proyectos se suspendieron) y otros antiguos transcriptores que se dedicaron a esa labor durante años.

Los juristas consultados por EL PAÍS explican que esta práctica en principio es legal, ya que Google lo señala en sus términos y condiciones. Pese a ello, despierta algunas dudas el procedimiento. “A falta de una investigación oficial, lo que parece que existe es un problema de transparencia, en el sentido de que el usuario no es consciente de qué se está realizando o qué se podría estar haciendo a partir de sus interacciones con el asistente virtual”, asegura Moisés Barrio, letrado del Consejo de Estado y abogado experto en derecho digital.

Entre los transcriptores consultados, todos coinciden en que la mayoría de las grabaciones se inician por el comando ‘Ok Google’. “En los casos en que el micrófono salta de forma errónea, no podíamos transcribir lo que se dice. En algunas ocasiones se escucha: ‘Google, esto no es para ti’. Pero otras no se dan cuenta y continúan su conversación, que se oye en el audio, aunque no lo escribíamos. Tampoco podíamos anotar cuando se dicen documentos personales [de identidad, bancarios…]. Lo que sí teníamos que transcribir son los números de teléfono móvil y direcciones”, explican varios antiguos trabajadores de la firma que prefieren mantener el anonimato.

Google sostiene que “los fragmentos de audio no se asocian con las cuentas de usuario como parte del proceso de revisión”, según afirmó en un comunicado oficial tras la filtración de audios en Bélgica. Algo que confirman los transcriptores. “El sistema solo te muestra el audio, su duración y un apartado donde escribir o revisar lo ya escrito. No tenemos acceso a ningún dato personal más allá de lo que se escucha”, aseguran los expertos en idiomas. Sin embargo, esto no quiere decir que la compañía no pueda identificar los datos con usuarios concretos. Portavoces de Google en España no han querido hacer aclaraciones a EL PAÍS sobre estas prácticas y se remiten a la nota oficial.

Mensajes íntimos

Pero, ¿qué tipo de búsquedas o conversaciones les llega a estos transcriptores? Todo lo que cualquier persona puede buscar en Internet. “Hay conversaciones familiares, entre amigos, discusiones e incluso charlas íntimas de parejas. También mucha consulta de direcciones y de lugares o tiendas, así como mensajes de tipo sexual y búsquedas de webs pornográficas”, reconocen. Con todo, la privacidad de los usuarios se pone en duda, ya que en los audios se escuchan datos sensibles en un entorno que Google no controla por completo.

En los casos de audios pertenecientes a interacciones con Google Home, la activación del sistema también lo marca el ‘Ok Google’. Sin embargo, en no pocas ocasiones la grabación comienza sin ese aviso. “Pasa mucho que la persona dice algo parecido y se activa el micrófono. Algunos se dan cuenta e incluso se asustan, aunque otros siguen con su conversación personal sin saber que se ha activado”, explican los transcriptores. En estas grabaciones, lo que se escucha principalmente son peticiones para una casa conectada: “Sobre todo para encender o apagar la luz, poner música, abrir el garaje, encender el aire acondicionado…”

La primera fisura —conocida— en la coraza que custodia los datos personales que los usuarios confían a Google se produjo con la filtración de un millar de estas grabaciones a la televisión belga VRT NWS. Esto provocó el reconocimiento de esta práctica a través del comunicado oficial de la firma, en el que David Mosees, responsable de producto de búsquedas de Google, decía: “Los expertos en idioma revisan y transcriben un pequeño conjunto de consultas para ayudarnos a comprender mejor cada idioma. Es una parte crítica y necesaria del proceso de creación de tecnología de voz”.

Esto implica que al menos una parte de esas interacciones no son completamente privadas. Es decir, se rompe la asunción habitual, y lo que a menudo afirman las empresas que gestionan asistentes virtuales como Google, Amazon, Samsung y Apple, de que las conversaciones entre un usuario y su asistente virtual se producen solo a través de la inteligencia artificial. Esto es, que los únicos que escuchan al usuario son robots. “A mis amigos les digo que no le hablen a los asistentes virtuales de cualquier compañía. Que nunca se sabe dónde pueden acabar esos audios…”, reconoce uno de los transcriptores.

Tras lo ocurrido, la firma californiana lanzó una caza de brujas para dar con el responsable de la fuga de audios en Bélgica y paralizó todos los trabajos. “Uno de los revisores ha violado nuestra política de seguridad de datos al filtrar información confidencial de audios. Nuestro equipo de seguridad y privacidad está investigándolo y tomaremos medidas. Estamos realizando una revisión completa de nuestras salvaguardas en este espacio para evitar que vuelva a ocurrir una conducta indebida como esta”, aseguraba Mosees en la nota. Una acción que es la que provoca que todos los transcriptores consultados pidan mantener el anonimato para no sufrir consecuencias.

Fuente: ElPais.com