

Ética social del reconocimiento automático del habla

Claudia Gómez-Rosabal¹, María José Castro-Bleda¹, and Pablo Aibar-Ausina²

¹ VRRAIN - Valencian Research Institute for Artificial Intelligence
Universitat Politècnica de València, Valencia, Spain
cgomros@posgrado.upv.es
mcastro@dsic.upv.es

² Departamento de Lenguajes y Sistemas Informáticos
Universitat Jaume I, Castellón, Spain
aibar@uji.es

Abstract. Existe una gran incertidumbre sobre lo que debe considerarse un uso ético de las tecnologías, aunque algunos puntos son claros. Toda tecnología debería tener en cuenta e intentar evitar las posibles consecuencias negativas que puede causar a los usuarios que la utilicen, en cuestiones como su privacidad, la seguridad de sus datos o el uso de roles y estereotipos discriminatorios. En gran medida, la tarea de lograr un uso ético de las tecnologías de voz recae en los desarrolladores encargados de diseñar, entrenar e integrar los modelos que permitan el funcionamiento de los dispositivos de voz.

Muchos estudios actuales apuntan al sesgo que ya existe en los sistemas de voz, debido al género, la raza, la extracción social . . . En la presente comunicación se analizan estos aspectos y como conclusión se propone la realización de un trabajo de recopilación y análisis tanto de los estudios realizados hasta el momento como de los resultados que producen diferentes herramientas de reconocimiento del habla dentro del ámbito del español de las distintas regiones y países, planteando a partir de ellos soluciones y alternativas a los diferentes sesgos que se detecten.

Keywords: reconocimiento automático del habla · inteligencia artificial · bots · ética · sesgo de género · sesgo de raza.

1 Introducción

Los sistemas de Reconocimiento Automático del Habla (RAH), que utilizan sofisticados algoritmos de aprendizaje automático para convertir el lenguaje hablado en texto, se han extendido cada vez más, impulsando populares asistentes virtuales, facilitando el subtítulo automático y permitiendo plataformas de dictado digital para la atención sanitaria. En los últimos años, la calidad de estos sistemas ha mejorado notablemente, debido tanto a los avances en el aprendizaje profundo como a la recopilación de conjuntos de datos a gran escala utilizados para entrenar los sistemas. Sin embargo, es interesante analizar el comportamiento de estas herramientas para diferentes subgrupos de la población [5].

La Inteligencia Artificial (IA) mediante voz es prácticamente omnipresente y cada vez más potente. Las previsiones sugieren que el comercio por voz será un negocio de 80.000 millones de dólares en 2023 [2]. Google informa de que el 20% de sus búsquedas se realizan hoy en día mediante consultas de voz, una cifra que se prevé que aumente al 50%. En 2017, Google anunció que su reconocimiento de voz tenía un 95% de precisión. Aunque es una cifra impresionante, ¿un 95% de precisión para quién? El trabajo que se propone realizar tiene como objetivo analizar precisamente la cuestión de la ética social del reconocimiento del habla en diferentes grupos sociales, mostrar ejemplos, analizar el impacto que pueden tener en la vida de las personas y posibles soluciones.

2 Sesgo de raza

A pesar del auge de las tecnologías del habla, existe la preocupación de que los sistemas de reconocimiento del habla presenten sesgo racial, un problema que ha salido a la luz recientemente en otras aplicaciones de aprendizaje automático, como el reconocimiento de rostros, el procesamiento del lenguaje natural, la publicidad en línea, la atención sanitaria y los servicios infantiles.

En un artículo publicado en el Journal Proceedings of the National Academy of Sciences [5], se evaluaron las disparidades raciales en cinco herramientas comerciales de conversión de voz a texto -desarrolladas por Amazon, Apple, Google, IBM y Microsoft- que potencian algunas de las aplicaciones más populares de la tecnología de reconocimiento de voz. Estas evaluaciones están basadas en entrevistas transcritas realizadas a 42 hablantes blancos y 73 negros. En total, este corpus abarca cinco ciudades de EE.UU. y consta de 19,8 horas de audio emparejadas según la edad y el sexo del hablante. Según los experimentos realizados en este artículo, los cinco sistemas de RAH muestran importantes disparidades raciales, con una tasa media de error de palabra (WER, por sus siglas en inglés “Word Error Rate”) de 35% para los hablantes negros frente a 19% para los blancos.

Los sistemas modernos de reconocimiento automático del habla suelen incluir un modelo lingüístico entrenado con datos de texto y un modelo acústico entrenado con datos de audio. En el estudio indican que las disparidades raciales que se observan surgen principalmente de una brecha de rendimiento en los modelos acústicos, lo que sugiere que los sistemas se confunden con las características fonológicas, fonéticas o prosódicas del inglés vernáculo afroamericano más que con las características gramaticales o léxicas. La causa probable de esta deficiencia es la insuficiencia de datos de audio de hablantes negros al entrenar los modelos.

Estas diferencias de rendimiento sugieren que es considerablemente más difícil para los afroamericanos beneficiarse del uso cada vez más extendido de la tecnología de reconocimiento del habla, desde los asistentes virtuales de los teléfonos móviles hasta la informática de manos libres para los discapacitados físicos. Estas disparidades también pueden perjudicar activamente a las comunidades afroamericanas cuando, por ejemplo, un software de reconocimiento del habla es

utilizado por los empleadores para evaluar automáticamente las entrevistas de candidatos o por los organismos de justicia penal para transcribir automáticamente los procedimientos judiciales.

3 Sesgo de género

Los beneficios de la IA son significativos: puede impulsar la eficiencia, la innovación y el ahorro de costes en la mano de obra y en la vida cotidiana. Sin embargo, la IA plantea problemas de parcialidad, automatización y seguridad humana que podrían agravar las históricas desigualdades sociales y económicas.

Un área particular que merece mayor atención es la manera en que los robots y asistentes de voz promueven estereotipos de género injustos. En todo el mundo, varios robots de servicio orientados al cliente, como el personal automatizado de los hoteles, camareros, guardias de seguridad o cuidadores de niños, presentan nombres, voces o apariencias de género. En Estados Unidos, Siri, Alexa, Cortana y Google Assistant -que en conjunto suman un 92,4% de la cuota de mercado de los asistentes para teléfonos inteligentes- han tenido tradicionalmente voces femeninas [4]. A medida que los bots artificiales y los asistentes de voz se hacen más frecuentes, es crucial evaluar cómo representan y refuerzan los estereotipos laborales de género existentes y cómo la composición de sus equipos de desarrollo afecta a estas representaciones.

El género ha provocado históricamente importantes disparidades económicas y sociales. Incluso hoy en día, los estereotipos relacionados con el género conforman las expectativas normativas para las mujeres en el lugar de trabajo; existen importantes investigaciones académicas que indican que la ayuda y el altruismo se perciben como rasgos femeninos, mientras que el liderazgo y la autoridad se asocian con la masculinidad.

Los debates sobre el género son vitales para crear una IA socialmente beneficiosa. A pesar de tener menos de una década, los modernos asistentes de voz requieren un escrutinio particular debido a la adopción generalizada por parte de los consumidores y a la tendencia social a antropomorfizar estos objetos asignándoles un género. Para abordar las representaciones de género en los bots de IA, los desarrolladores deben centrarse en la diversificación de sus equipos de ingeniería; deben desarrollarse estándares para toda la industria sobre el género en los bots de IA; y las empresas tecnológicas deben aumentar la transparencia. Los asistentes de voz no serán el último bot de IA popular, pero cuanto antes normalicemos el cuestionamiento de la representación de género en estos productos, más fácil será continuar con estas conversaciones a medida que surja la futura IA.

4 Conclusiones

Hoy en día aún existe una gran incertidumbre acerca de lo que debe considerarse un uso ético de las tecnologías, pero algunos aspectos sí suscitan un amplio consenso: ninguna tecnología debería producir consecuencias negativas para los

usuarios, ni en términos de su privacidad, ni en términos de seguridad de sus datos o acerca de cómo se perciben a sí mismos como individuos [3].

Muchos estudios apuntan a la existencia de un sesgo en los actuales sistemas de voz relacionados tanto con la raza, como con el género o algunos otros aspectos de tipo social. Se propone realizar un trabajo que profundice en el estudio concreto de estos posibles sesgos en el ámbito hispanoparlante mediante la realización de un trabajo de recopilación y análisis de los estudios realizados hasta el momento y de los resultados que producen diferentes herramientas de reconocimiento del habla en las distintas regiones y países de habla hispana. Concretamente, los objetivos específicos del trabajo son:

- Recopilar los estudios realizados hasta el momento sobre los prejuicios de raza, género o minorías, poniendo el foco en las distintas zonas en las que se habla el español.
- Realizar un estudio empírico sobre como dichos prejuicios se trasladan a las tecnologías de reconocimiento del habla disponibles. Para ello:
 - Se utilizarán distintos corpora de voz de hablantes de distintas zonas (regiones o países) donde se hable español [1].
 - Se seleccionarán las herramientas de voz más utilizadas (asistente de Google, Siri u otros similares).
 - Se estudiarán las diferencias en cuanto a la tasa de reconocimiento que se alcanza con esas herramientas de voz.

En gran medida, la tarea de lograr un uso ético de las tecnologías de voz recae en los desarrolladores encargados de diseñar, entrenar e integrar los modelos que permitan el funcionamiento de los dispositivos de voz. Por supuesto, depende de múltiples organismos (por ejemplo, la Comisión Europea) trabajar en estándares, pautas y buenas prácticas que determinen los requisitos mínimos que los desarrolladores y otras partes interesadas deben mantener para que se considere que hacen un uso ético de las tecnologías de voz. Nuestro objetivo final será, a partir de los estudios bibliográficos y empíricos, plantear soluciones y alternativas para reducir y eliminar estos sesgos, con especial énfasis en el sesgo del habla del español de distintas regiones.

References

1. Corpus de Referencia del Español Actual (2008), <http://asines.org/corpus/>
2. Bajorek, J.: Voice recognition still has significant race and gender biases. *Harvard business review* (2019)
3. Campbell, A.: Ethics in Voice Technologies (2021), <https://www.compriseh2020.eu/ethics-in-voice-technologies/>
4. Chin, C., Robison, M.: How AI bots and voice assistants reinforce gender bias. *Tech. rep., Brookings and the Italian Institute for International Political Studies* (2020)
5. Koenecke, A., Nam, A., Lake, E., Nudell, J., Quartey, M., Mengesha, Z., Toups, C., Rickford, J.R., Jurafsky, D., Goel, S.: Racial disparities in automated speech recognition. *Proceedings of the National Academy of Sciences* **117**(14), 7684–7689 (2020). <https://doi.org/10.1073/pnas.1915768117>, <https://www.pnas.org/doi/abs/10.1073/pnas.1915768117>