

La feminización de la inteligencia artificial y la antropomorfización de los asistentes virtuales

Introducción

En los últimos años, el desarrollo y la implementación de asistentes digitales impulsados por inteligencia artificial han llevado a la creación de sistemas que, en su gran mayoría, presentan características de comportamiento y voces femeninas. Este fenómeno refuerza nociones sexistas, perpetuando desigualdades de género en el contexto tecnológico. A la fecha, se conoce que esta tendencia no es accidental, sino que responde a una combinación de factores socioculturales. Al entender las causas, se pueden abordar mejor las implicaciones de este fenómeno en la sociedad.

Se pretende concientizar sobre los estereotipos de género arraigados en la sociedad y su relación con las funciones de servicio de las asistentes digitales. Además, se busca fomentar una discusión más amplia sobre la representación de género en la tecnología y promover cambios en las expectativas hacia las capacidades y roles de las IA, impulsando la diversificación y equidad en su desarrollo y aplicación.

I. Causas de la Feminización de la IA

Las tendencias actuales en el desarrollo de asistentes virtuales con características y voces femeninas pueden ser explicadas desde distintos ámbitos. En primer lugar, autores como Costa (2019), relacionan este fenómeno con la percepción de la IA en la cultura popular, ya que la representación de la IA en la ciencia ficción a menudo corresponde a una visión sesgada del género que asocia la femineidad con roles sumisos y de cuidado, y la masculinidad con escenarios agresivos o incluso amenazantes.

A su vez, los debates comunes argumentan la preferencia del usuario como justificación para la feminización de las IAs. Los resultados de un estudio llevado a cabo por Vilà (2021), revelan que los asistentes virtuales analizados reproducen patrones de socialización femeninos, construyendo perfiles femeninos exagerados, como el de la madre y la secretaria, hasta llegar a un ideal heteropatriarcal, lo cual antropomorfiza a estos sistemas y activa imaginarios vinculados a ciertos marcos y patrones sociales. Como menciona la autora: "La tendencia a feminizar a los agentes virtuales puede responder a la voluntad de atribuir a los robots cualidades amigables y no amenazantes, asociadas a lo 'agradable' y 'pacífico' de la voz de mujer." (pág. 681).

Por otro lado, se argumenta que la causa de la feminización de los algoritmos detrás de las asistentes virtuales se encuentra implícita en las bases de datos. Según explica Wellner (2020), dichos algoritmos tienen a reproducir y amplificar los sesgos de género existentes en los datos con los que son entrenados. En palabras de la autora: “[desarrolladores] señalan el hecho de que los algoritmos aprenden de conjuntos de datos, y dado que estos reflejan el “mundo”, que tiene un sesgo de género, los algoritmos terminan duplicando la lógica del mundo, incluidos los sesgos” (pág. 147). Experimentos llevados a cabo por Wellner ejemplifican este fenómeno, demostrando que los algoritmos de traducción tienden a asociar profesiones y atributos positivos con formas masculinas y atributos negativos con formas femeninas.

Similarmente, experimentos como el llevado a cabo por Antonopoulou (2023), demuestran que la subrepresentación de mujeres en el proceso de diseño y en los conjuntos de datos de entrenamiento resulta en servicios de IA que discriminan contra los grupos subrepresentados e intensifican su marginación. Como menciona en su artículo:

La investigación actual en inteligencia artificial (IA) arroja luz sobre el sesgo algorítmico integrado en los sistemas de IA. La subrepresentación de las mujeres en el sector de diseño de IA de la industria tecnológica, así como en los conjuntos de datos de capacitación, da como resultado productos tecnológicos que codifican el sesgo de género, refuerzan los estereotipos y reproducen nociones normativas de género y feminidad. (pág. 158)

Otros autores apoyan esta teoría, haciendo énfasis en que el campo de la IA es principalmente liderado por hombres. En palabras de LaFrance (2016): “si los hombres son a menudo los que construyen asistentes digitales, y aquellos asistentes siguen el modelo de las mujeres, [...] eso probablemente refleja lo que algunos hombres piensan de las mujeres” (como se citó en Costa, 2019).

Del mismo modo, Costa (2019) otorga otra propuesta, la cual justifica este fenómeno detrás del entendimiento de las voces femeninas por parte de los usuarios. En su artículo menciona que las voces femeninas se perciben como más adecuadas para los asistentes virtuales porque son menos dominantes y más fáciles de entender a comparación de las de los hombres. Costa trae a la luz artículos como el de Piper (2016), quien afirma que “los parlantes de menor calidad no soportan los graves completos de la voz masculina, solo generando sonidos agudos claramente” (pág. 41)

Por último, autores como Adams (2020) observan la feminización de las tecnologías de IA como un intento de mediar entre las esperanzas y temores opuestos que estas tecnologías generan. Al asociar la tecnología con la feminidad, se busca mitigar el miedo a la inteligencia artificial mientras se promueven sus beneficios potenciales. Adams menciona que las emociones positivas y negativas asociadas con estas tecnologías están intrínsecamente ligadas a su representación de género, mientras que las reacciones afectivas hacia las máquinas inteligentes no se pueden separar fácilmente cuando estas máquinas están feminizadas.

II. Consecuencias e implicaciones sociales

En cuanto a las implicaciones de la feminización de las asistentes virtuales, en primer lugar, la mayoría de los autores afirman que la principal consecuencia es la amplificación del sesgo y la perpetuación de nociones reduccionistas de género y feminidad. Como menciona Wellner (2020): “Debido al mecanismo de retroalimentación, los resultados con sesgo de género se retroalimentan al sistema, profundizando así los sesgos”. (pág. 130)

Similarmente, como menciona Antonopolou (2023), las IAs antropomórficas son percibidas como entidades sociales y las interacciones de los usuarios con ellas afectan las interacciones interpersonales en el mundo real. Por lo tanto, codificar comportamientos serviles y permisivos en IAs feminizadas normaliza el comportamiento sexista y abusivo hacia las mujeres virtuales, lo que potencialmente se expande hacia las mujeres reales.

Por otro lado, el análisis de Woods (2018) revela que los roles estereotípicamente femeninos que desempeñan las asistentes digitales, tal como el de la cuidadora, madre o esposa, ayudan a disimular los modos de vigilancia y facilitan la aceptación de la lógica de mercado en la economía de plataformas. Al adoptar una persona femenina esencialista, caracterizado como “domesticidad digital,” estos objetos inteligentes orientan a los usuarios a participar productivamente en el “capitalismo de vigilancia” es decir, la mercantilización de datos personales, normalizando el intercambio de datos íntimos. En palabras del autor:

[...] equipar a la IA con una personalidad femenina funciona para desarmar retóricamente a los usuarios con ansiedades sobre el intercambio íntimo de datos, lo que a su vez permite la condensación de datos y poder en manos de unos pocos capitalistas de vigilancia selectos. (pág. 335)

III. Soluciones y alternativas en el desarrollo de IA

En cuanto a soluciones para afrontar este fenómeno, autores sugieren que es crucial realizar cambios tanto en los conjuntos de datos como en los propios algoritmos para asegurar una representación justa y equitativa de todos los usuarios. Por su parte, Wellner (2020) propone como solución la revisión y limpieza de los conjuntos de datos utilizados para entrenar los algoritmos, asegurando que sean representativos y no contengan sesgos de género. Se propone un cambio de enfoque desde el "aprendizaje automático" (machine learning) hacia la "educación automática" (machine education), donde los algoritmos no solo aprenden patrones de datos, sino que también entienden conceptos de lo bueno y lo malo. Esto implica enseñar a los algoritmos principios éticos y morales, como se sugiere en Wellner & Rothman (2019).

Por otro lado, autores como Hall (2023) proponen un enfoque más social, recalcando la importancia del desarrollo de soluciones interdisciplinarias, involucrando a más mujeres y minorías de género en los equipos de IA. Hall afirma que tratar el sesgo algorítmico solo como un problema técnico es insuficiente, ya que puede resultar en resultados discriminatorios. Se recalca que la supervisión humana, la integración de la ética en el diseño, y la promoción de la diversidad y la inclusión son esenciales para evitar perpetuar estereotipos pasados y promover una IA más justa. Según afirma Johnson (2019):

Aumentar la inclusión de género en el desarrollo de tecnologías de inteligencia artificial introducirá perspectivas importantes y diversas, reducirá la influencia de los sesgos cognitivos en el diseño, la capacitación y la supervisión de los algoritmos de aprendizaje y, por lo tanto, mitigará la preocupación por la gestión de riesgos relacionados con los sesgos (como se citó en Hall, 2023)

Similarmente, el artículo escrito por Antonopoulou (2023) sostiene que la participación activa de mujeres y grupos subrepresentados en el diseño de IA es crucial para desarrollar tecnologías que sean más inclusivas y menos sesgadas. El autor resalta la necesidad de diversificar los parámetros de diseño y utilizar datos de entrenamiento alternativos para crear IA que reflejen una gama más amplia de identidades y experiencias.

También existen propuestas de modificación de los algoritmos de aprendizaje automático, como la presentada por Marinucci (2022), la cual propone la implementación de recursos semánticos integrados que combinen conocimientos simbólicos y "subsimbólicos", permitiendo crear redes de conocimiento compuestas para entender relaciones implícitas que afectan a grupos sociales específicos.

En su artículo, se habla sobre juntar dos tipos de información diferentes sobre el género. Una es sobre las palabras que la gente asocia libremente con el género, y la otra es sobre cómo las palabras están relacionadas en bases de datos de palabras. Se busca juntar toda esta información en un mismo lugar para llegar a un mejor entendimiento del género y cómo la gente piensa en él. Para hacerlo, se usa una especie de "mapa" de conocimiento para organizar y conectar toda la información de manera clara y útil. Todo con el propósito mejorar la equidad y transparencia en los algoritmos de aprendizaje automático y abordar la dificultad teórica de definir sesgos y estereotipos, especialmente de género. Esta integración puede ser útil para desarrolladores en la creación de los sistemas inteligentes detrás de las asistentes virtuales. Según Marinucci (2022):

Todos los debates éticos y los marcos para mitigar los prejuicios representan una superestructura de decisiones y responsabilidades humanas que la máquina tendrá que aprender. Para ser efectiva, esta superestructura debe depender de la conciencia. Y, en consecuencia, la explicabilidad de la estructura cognitiva profunda y los mecanismos internos de la toma de decisiones automática e inconsciente como recurso cognitivo de todos los seres humanos. Solo así las aplicaciones de la IA podrán realmente obtener su valor positivo en nuestra sociedad, en lugar de ser una fuente de peligros. Este enfoque sinérgico nos permitiría aprender de la IA algo sobre nosotros, aunque no necesariamente agradable, en lugar de solo darle a la IA nuestros datos para que aprenda. (p.758)

Conclusión

A lo largo del texto, se analizó cómo la feminización de los asistentes digitales refleja y amplifica los sesgos de género, perpetuando estereotipos y afectando negativamente los comportamientos sociales hacia las mujeres. Este fenómeno, influenciado por percepciones culturales y sesgos en los datos de entrenamiento, normaliza el comportamiento sexista y contribuye a la mercantilización de datos personales. Para mitigar estos problemas, es esencial diversificar los equipos de desarrollo e implementar enfoques éticos en el diseño de IA, así como modificar los conjuntos de datos y los mismos algoritmos para asegurar una representación más justa y equitativa. Solo a través de estas acciones será posible avanzar hacia un futuro donde la tecnología sea una herramienta de progreso y no un reflejo de los prejuicios existentes.

Referencias:

- Adams, R. (2020). Helen A'Loy and other tales of female automata: a gendered reading of the narratives of hopes and fears of intelligent machines and artificial intelligence. *AI & Society*, 35(3), 569-579. <https://doi-org.udlap.idm.oclc.org/10.1007/s00146-019-00918-7>
- Antonopoulou, C. (2023). Algorithmic bias in anthropomorphic artificial intelligence: Critical perspectives through the practice of women media artists and designers. *Technoetic Arts: A Journal of Speculative Research*, 21(2), 157-174. https://doi-org.udlap.idm.oclc.org/10.1386/tear_00109_1
- Costa, P., & Ribas, L. (2019). AI becomes her: Discussing gender and artificial intelligence. *Technoetic Arts: A Journal of Speculative Research*, 17(1/2), 171-193. https://doi-org.udlap.idm.oclc.org/10.1386/tear_00014_1
- Hall, P., & Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review*, 47(7), 1264-1279. <https://doi-org.udlap.idm.oclc.org/10.1108/OIR-08-2021-0452>
- Marinucci, L. (2022). Exposing implicit biases and stereotypes in human and artificial intelligence: state of the art and challenges with a focus on gender. *AI & Society*, 38, 747-761. <https://doi.org/10.1007/s00146-022-01474-3>
- Vilà, C. (2021). Reproducción de discursos dominantes en la tecnología: Análisis crítico del discurso aplicado a los asistentes de voz virtuales. *Discurso & Sociedad*, 15(3) 679-700. <https://dialnet.unirioja.es/servlet/articulo?codigo=8168329>
- Wellner, G. P. (2020). When AI is Gender-biased: The Effects of Biased AI on the Everyday Experiences of Women. *Humana Mente*, 12(37), 127-150.
- Woods, H. S. (2018). Asking more of Siri and Alexa: feminine persona in service of surveillance capitalism. *Critical Studies in Media Communication*, 35(4), 334-349. <https://doi-org.udlap.idm.oclc.org/10.1080/15295036.2018.1488082>

Sobre el autor:

Daniela Renée Ramírez Gutiérrez, estudiante de la licenciatura en Ciencia de Datos e integrante del Programa de Honores en la Universidad de las Américas Puebla. Ha desarrollado proyectos de investigación del área de biotecnología con acreditaciones nacionales e internacionales. Previamente, certificó estudios en las áreas de emprendimiento e ingenierías por el programa POP del Bachillerato Internacional (IB).

Contacto: daniela.ramirezgz@udlap.mx

Tutor académico:

Miguel Ángel Reyes Cortés, Doctor en Estadística e Investigación Operativa por la Universidad de La Coruña. Además, Maestro en Técnicas Estadísticas, Maestro en Demografía, y Licenciado en Física por la Benemérita Universidad Autónoma de Puebla. Actualmente, es Coordinador de la Licenciatura en Ciencia de Datos UDLAP.

Contacto: miguel.reyes@udlap.mx