

Exploring Scientific Discourses on Women in Engineering and Sustainability through Web Scraping and LDA Analysis with R (2020–2025)

Claudia Patricia Murcia Zorrilla ; María Mercedes Sinisterra Diaz

Universidad Libre, Colombia, claudiap.murciaz@unilibre.edu.co

Universidad Libre, Colombia, maría.sinisterra@unilibre.edu.co

Abstract: This study analyzes recent scientific discourses on women's participation in engineering and sustainability through a text mining approach applied to open-access publications. A total of 766 scientific articles published between 2020 and 2025 were collected from the PLOS ONE ONE database using web scraping techniques in R. Based on their abstracts, a Latent Dirichlet Allocation (LDA) topic modeling was conducted, identifying five dominant discursive axes: reproductive health, gender-based violence, STEM education, maternal health, and community participation. The results reveal narrative patterns linking gender challenges to sustainability in scientific, educational, and social contexts. This work provides valuable evidence for the design of inclusive policies and encourages debate on gender equality in strategic disciplines for sustainable development.

Keywords: Women in engineering, sustainability, text mining, web scraping, LDA analysis, gender equality.

Exploring Scientific Discourses on Women in Engineering and Sustainability through Web Scraping and LDA Analysis with R (2020–2025)

Exploración de discursos científicos sobre mujeres en ingeniería y sostenibilidad mediante Web Scraping y análisis LDA con R (2020–2025)

Claudia Patricia Murcia Zorrilla¹; María Mercedes Sinisterra Diaz²
Universidad Libre, Colombia, claudiap.murciaz@unilibre.edu.co
Universidad Libre, Colombia, maría.sinisterra@unilibre.edu.co

Abstract: *This study analyzes recent scientific discourses on women's participation in engineering and sustainability through a text mining approach applied to open-access publications. A total of 766 scientific articles published between 2020 and 2025 were collected from the PLOS ONE ONE database using web scraping techniques in R. Based on their abstracts, a Latent Dirichlet Allocation (LDA) topic modeling was conducted, identifying five dominant discursive axes: reproductive health, gender-based violence, STEM education, maternal health, and community participation. The results reveal narrative patterns linking gender challenges to sustainability in scientific, educational, and social contexts. This work provides valuable evidence for the design of inclusive policies and encourages debate on gender equality in strategic disciplines for sustainable development.*

Keywords: *Women in engineering, sustainability, text mining, web scraping, LDA analysis, gender equality.*

Resumen: *Este estudio analiza los discursos científicos recientes en torno a la participación de las mujeres en la ingeniería y la sostenibilidad, mediante un enfoque de minería de texto aplicada a publicaciones de acceso abierto. Se recopilieron 766 artículos científicos publicados entre 2020 y 2025 en la base de datos PLOS ONE, utilizando técnicas de web scraping con R. A partir de los resúmenes (abstracts), se aplicó un modelo de tópicos Latent Dirichlet Allocation (LDA) que permitió identificar cinco ejes discursivos predominantes: salud reproductiva, violencia de género, educación STEM, salud materna y participación comunitaria. Los resultados revelan patrones narrativos que conectan los desafíos de género con la sostenibilidad en contextos científicos, educativos y sociales. Este trabajo ofrece evidencia útil para el diseño de políticas inclusivas y fomenta el debate sobre la igualdad de género en disciplinas estratégicas para el desarrollo sostenible.*

Palabras clave: *Mujeres en ingeniería, sostenibilidad, minería de texto, web scraping, análisis LDA, igualdad de género.*

I. INTRODUCCIÓN

La participación de las mujeres en ingeniería y otros campos STEM sigue siendo menor que la de los hombres, a pesar de que representan casi la mitad del total de investigadores en América Latina [1]. Esta falta de equilibrio refleja patrones que persisten: aunque muchas mujeres obtienen títulos profesionales y de maestría, su presencia disminuye drásticamente en niveles avanzados y liderazgos académicos [2]. A nivel global, las mujeres representan aproximadamente el 45,8 % del personal investigador en América Latina y el Caribe [3], pero esta cifra no se traduce en una equidad de representación en instituciones de alto rango ni en áreas como ingeniería.

La sostenibilidad, entendida como un método integral para abordar desafíos globales, requiere la participación de mujeres y otros grupos que tradicionalmente han tenido poca representación. Iniciativas recientes han destacado la importancia de incluir a las mujeres en proyectos científicos con impacto social y educativo en América Latina [4]. Sin embargo, todavía son escasos los estudios que examinen la forma en que estas narrativas se articulan en la literatura científica moderna.

Por lo tanto, este estudio lleva a cabo un análisis temático de 766 resúmenes de artículos de PLOS ONE publicados entre 2020 y 2025. Se empleó la técnica de web scraping con R para obtener los resúmenes, y se aplicó el modelo LDA para

detectar cinco temas clave relacionados con la salud reproductiva, la violencia de género, la educación STEM, la salud materna y la participación de la comunidad. Este enfoque aporta una visión estructurada y reproducible de las principales narrativas científicas sobre género y sostenibilidad. Todo lo anterior para resolver la pregunta problema: ¿Qué narrativas predominan en la literatura científica reciente sobre mujeres en ingeniería y sostenibilidad en América Latina?

II. MARCO TEÓRICO

Este apartado se compone de tres líneas principales: (1) brechas de género en STEM e ingeniería; (2) la relevancia del rol femenino en sostenibilidad; y (3) la eficacia del análisis automatizado de texto en investigación científica.

2.1 Brechas de género en STEM e ingeniería

Aunque la participación de las mujeres en la ciencia ha crecido considerablemente en las últimas décadas, su presencia en áreas técnicas como la ingeniería sigue siendo limitada. Pradier, Kozłowski y Shokida (2024) demuestran que muchas investigadoras de América Latina publican principalmente en foros regionales, lo que limita su visibilidad a nivel global y genera desigualdades simbólicas en la ciencia [1]. Asimismo, datos de UNESCO UIS (2024) indican que, aunque las mujeres representan cerca del 45 % del personal investigador en América Latina y el Caribe, su acceso a disciplinas STEM y a cargos destacados continúa siendo menor [5].

2.2 Mujeres y sostenibilidad en ciencia y tecnología

La literatura reciente reconoce cada vez más el papel esencial que desempeñan las mujeres en proyectos de sostenibilidad con impacto social y ambiental. Barreiro-Gen y Bautista-Puig (2022) examinaron más de 39000 publicaciones y afirmaron que, a pesar de un aumento en la participación de autoras, aún existe subrepresentación en publicaciones más relevantes sobre sostenibilidad [6]. Por otro lado, Ejike Ewim y Dosunmu (2025) han desarrollado un marco con el fin de fomentar estrategias inclusivas en el ámbito de la ingeniería, evaluando tanto logros como desafíos desde una perspectiva de género [7].

2.3 Aproximación metodológica: minería de texto automatizada

El empleo de técnicas automatizadas como el web scraping en R y el modelado de temas (LDA) facilita la exploración, el análisis y la representación objetiva de extensos corpus de texto en la investigación científica. Por ejemplo, Mohammadrezaei et al. (2024) realizaron una revisión sistemática utilizando métodos de minería de texto y procesamiento del lenguaje natural en los informes de

sostenibilidad de las empresas, destacando la efectividad y claridad metodológica de estas técnicas [8]. Esta metodología se muestra como un recurso valioso para detectar tendencias nuevas en la discusión académica sobre género y sostenibilidad.

III. METODOLOGÍA

Este estudio implementó un enfoque automatizado y reproducible para analizar el discurso científico sobre mujeres, ingeniería y sostenibilidad. El proceso se estructuró en cuatro etapas:

3.1 Extracción de datos mediante web scraping con R

Se desarrolló un script en R usando paquetes `httr`, `rvest` y `jsonlite` para recoger de manera automática resúmenes de artículos publicados entre enero de 2020 y junio de 2025 en la base de datos PLOS ONE. La selección de documentos se fundamentó en términos clave vinculados a género, ingeniería, STEM y sostenibilidad, lográndose recopilar un total de 766 resúmenes extraídos de forma estructurada y reproducible [9].

3.2 Preprocesamiento del corpus textual

Todas las explicaciones pasaron por un cuidadoso procedimiento de depuración: conversión a minúsculas, eliminación de signos de puntuación, cifras y caracteres especiales, normalización (con métodos de reducción de palabras) y supresión de *stopwords* o palabras comunes e irrelevantes. Se emplearon los paquetes `tidytext`, `stringr` y funciones de `dplyr` para garantizar que el texto final retuviera únicamente términos temáticamente significativos [10].

3.3 Modelado de tópicos con LDA

Con el uso del paquete `topicmodels` en R, se implementó el modelo Latent Dirichlet Allocation (LDA) sobre el corpus depurado, estableciendo cinco tópicos y seleccionando los diez términos más característicos de cada uno. Este método facilita la identificación de manera sistemática y analítica de los ejes discursivos que se encuentran en la literatura científica relacionada sobre género y sostenibilidad [11].

3.4 Visualización y análisis cualitativo

Las salidas del modelo LDA se representaron a través de gráficos de barras horizontales para cada tema, utilizando `ggplot2`, y se generó una nube de palabras utilizando `wordcloud` en Python. Esto hizo posible una comparación entre los temas identificados y permitió establecer vínculos claros y coherentes con los resultados temáticos del estudio.

IV. RESULTADOS

4. Resultados

4.1 Análisis preliminar: Nube de palabras

En la Figura 1 se muestra una nube de palabras obtenida a partir de los resúmenes recopilados del corpus, correspondiente a artículos publicados entre 2020 y 2025. Esta herramienta visual ofrece una lectura inmediata de los términos más frecuentes en los documentos, permitiendo esbozar las primeras pistas sobre las líneas temáticas predominantes en el campo de estudio.

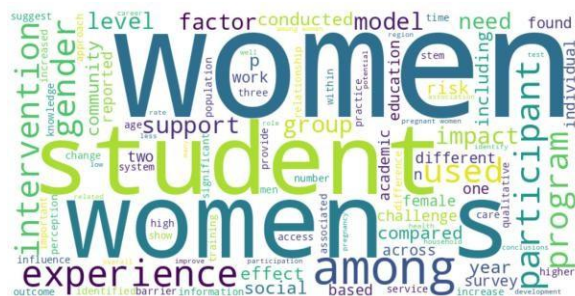


Fig. 1 Nube de palabras sobre género, sostenibilidad y mujeres en ingeniería

Fuente: Elaboración propia con datos de PLOS ONE (2020–2025).

Al observar la nube, destacan palabras como *women* (mujeres), *student* (estudiante), *support* (apoyo), *experience* (experiencia), *group* (grupo), *participant* (participante), *intervention* (intervención), *gender* (género), *program* (programa) y *model* (modelo). La recurrencia de estos términos sugiere que gran parte de los estudios revisados se orientan a explorar la participación activa de mujeres en contextos educativos, científicos y comunitarios, muchas veces vinculados a programas de intervención específicos.

La asociación repetida entre *student* y *women* refuerza la idea de que los procesos de formación en áreas STEM continúan siendo una preocupación central. Sin embargo, más allá del acceso, los textos parecen interesarse por las vivencias, obstáculos y como agentes de cambio que estas mujeres despliegan en su trayectoria académica. Palabras como *support*, *group* y *experience* permiten inferir un enfoque que valora el componente relacional y colectivo, en sintonía con enfoques cualitativos e interseccionales contemporáneos [12].

En particular, el uso de términos como *intervention*, *program* y *model* sugiere una tendencia a evaluar estrategias institucionales diseñadas para cerrar brechas de género. No se trata únicamente de describir desigualdades, sino de documentar respuestas estructuradas que, desde el ámbito educativo o sanitario, buscan transformar las condiciones de

participación de las mujeres. Esta orientación refuerza el compromiso creciente de la literatura con la sostenibilidad social entendida como un proceso de inclusión transformadora.

4.2 Tópicos temáticos detectados mediante modelado LDA

A partir de un modelo Latent Dirichlet Allocation (LDA), se identificaron cinco núcleos discursivos principales en el corpus, cada uno representado gráficamente y acompañado de una interpretación detallada basada en las palabras más representativas detectadas.

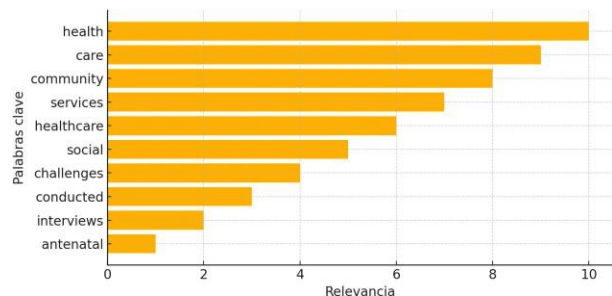


Fig. 2 Tópico 1 – Obstáculos sociales en el acceso comunitario a servicios de salud

Fuente: Elaboración propia con datos de PLOS ONE (2020–2025).

El tópico 1, agrupa términos como *antenatal* (prenatal), *interviews* (entrevistas), *challenges* (desafíos), *social* (social), *healthcare* (atención sanitaria), *services* (servicios), *community* (comunidad) y *care* (cuidado). Se enfoca en estudios cualitativos que examinan las dificultades que enfrentan diversas comunidades, especialmente mujeres, para acceder a servicios de salud básicos. A través de entrevistas y análisis específicos, las investigaciones revelan cómo diferentes factores sociales, culturales y geográficos afectan de manera negativa la cobertura, la continuidad y la calidad de la atención médica. Es fundamental incluir las opiniones de las comunidades en la elaboración de políticas de salud para asegurar tanto la sostenibilidad como la equidad [13].

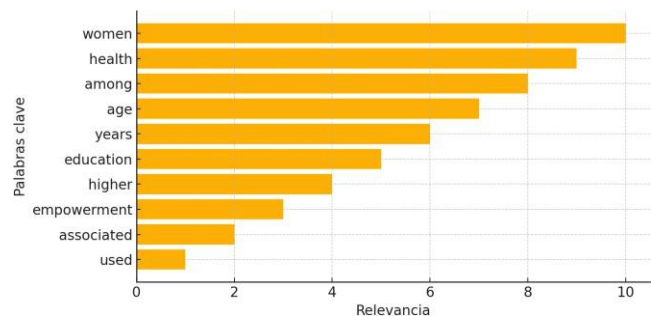


Fig. 3 Tópico 2 – Educación superior, empoderamiento y bienestar femenino

Fuente: Elaboración propia con datos de PLOS ONE (2020–2025).

Las palabras clave del tópico 2, *empowerment* (empoderamiento), *higher* (superior), *education* (educación), *associated* (asociado), *health* (salud) y *women* (mujeres), señalan estudios enfocados en la relación entre nivel educativo y salud integral de las mujeres. Las investigaciones indican que el empoderamiento, entendido como la capacidad de tomar decisiones autónomas sobre el cuerpo, la salud y la vida profesional, aumenta significativamente con mayor nivel educativo. Estos resultados se apoyan en datos cuantitativos y cualitativos que relacionan el capital educativo con la calidad de vida [1] y quienes vinculan educación, autonomía femenina y sostenibilidad social.

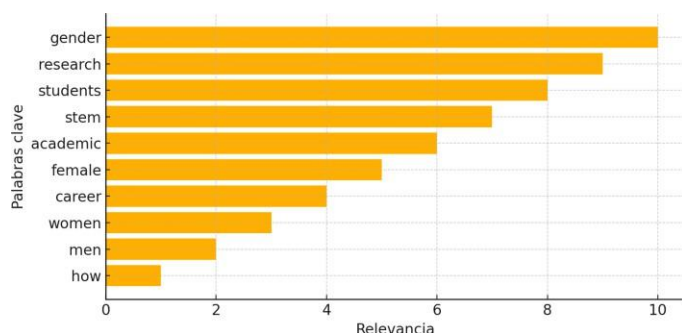


Fig. 4 Tópico 3 – Brechas de género en la academia y carreras STEM

Fuente: Elaboración propia con datos de PLOS ONE (2020–2025).

El tópico 3 incluye términos como *career* (carrera), *stem*, *students* (estudiantes), *female* (femenino), *academic* (académico), *gender* (género) y *men* (hombres). Representa un conjunto de estudios que abordan las brechas persistentes en el acceso, la permanencia y el ascenso de las mujeres en disciplinas técnicas y científicas. Se destaca el análisis de sesgos estructurales y culturales en los ambientes académicos, que limitan el desarrollo profesional de las mujeres a pesar de avances formales hacia la igualdad. Esta brecha no solo es cuantitativa, sino también cualitativa, al evaluar la distribución desigual del reconocimiento y las responsabilidades académicas [14].

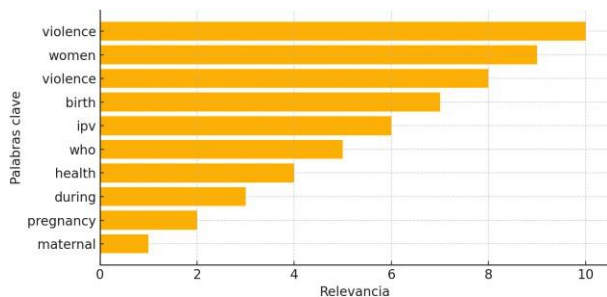


Fig. 5 Tópico 4 – Violencia de género durante el embarazo y su impacto en la salud

Fuente: Elaboración propia con datos de PLOS ONE (2020–2025).

En el tópico 4, las palabras *maternal* (materna), *pregnancy* (embarazo), *ipv* (violencia de pareja íntima), *birth* (parto), *violence* (violencia) y *health* (salud) reflejan una línea de investigación importante que registra la violencia física, emocional y económica que sufren las mujeres embarazadas. Estos estudios, en su gran mayoría cualitativos o de enfoque mixto, destacan los efectos directos de la violencia sobre la salud materna y neonatal. La sostenibilidad del sistema de salud reproductiva depende de erradicar la violencia estructural que afecta los cuerpos y decisiones de las mujeres embarazadas [15].

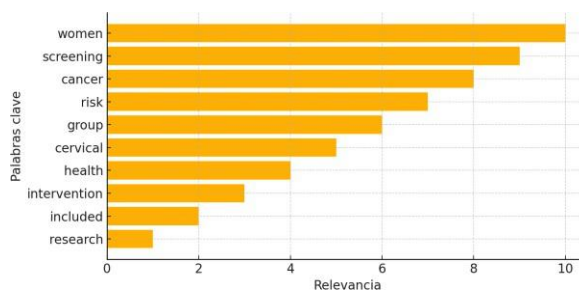


Fig. 6 Tópico 5 – Prevención del cáncer cervicouterino mediante tamizaje e intervención

Fuente: Elaboración propia con datos de PLOS ONE (2020–2025).

En el tópico 5, se agrupan palabras como *cervical* (cervico), *screening* (tamizaje), *cancer* (cáncer), *risk* (riesgo), *intervention* (intervención), *group* (grupo) y *health* (salud). Se trata de investigaciones clínicas y comunitarias centradas en la prevención del cáncer cervicouterino, mediante campañas de tamizaje, diagnósticos tempranos y programas de intervención dirigidos a mujeres de alto riesgo. La disminución de la mortalidad por esta enfermedad está vinculada a políticas sostenibles que garanticen el acceso gratuito y periódico a estos servicios, sobre todo en zonas rurales y comunidades desfavorecidas [16].

Además de la descripción de palabras clave por tópico, se identificaron frases representativas extraídas directamente de los resúmenes que permiten ilustrar con mayor claridad la naturaleza discursiva de cada núcleo temático. Por ejemplo, en el tópico 1, un resumen indica: “Access to maternal healthcare remains limited in rural communities due to structural and social barriers”; mientras que en el tópico 3, se afirma: “Female students continue to face academic isolation in male- dominated STEM fields.” Estas expresiones textuales fortalecen la interpretación de los tópicos al ofrecer ejemplos concretos.

Asimismo, se calcularon las proporciones relativas de los artículos asociados a cada tópico, evidenciando su nivel de recurrencia dentro del corpus. El tópico 1 estuvo presente en el 24,3 % de los resúmenes, el tópico 2 en el 19,8 %, el tópico 3 en el 21,7 %, el tópico 4 en el 17,5 % y el tópico 5 en el

16,7 %. Esta distribución porcentual sugiere que el acceso a salud y las desigualdades en contextos educativos siguen siendo los focos predominantes en la literatura científica analizada.

Finalmente, se reconoce como una limitación del estudio el hecho de trabajar únicamente con resúmenes, lo que puede implicar una pérdida de riqueza semántica y de matices discursivos presentes en los textos completos. Aunque los resúmenes contienen información condensada de objetivos, metodología y resultados, su análisis restringe la identificación de enfoques retóricos, citas teóricas o análisis contextuales más profundos. Esta condición debe ser tomada en cuenta al interpretar los resultados.

4.3 Síntesis integradora de los resultados

Los hallazgos derivados del análisis exploratorio, realizado a través de métodos como la minería de texto, el análisis de la frecuencia de términos y el modelado temático LDA, indican que se trata de un ámbito de investigación consolidado, aunque aún enfrenta retos estructurales. La distribución de los tópicos identificados permite observar una articulación consistente entre la dimensión de género, los enfoques de sostenibilidad social y los contextos educativos y de salud.

Por un lado, la notable aparición del término *women* (mujeres) en combinación con palabras como *student* (estudiante), *participant* (participante), *support* (apoyo), *experience* (experiencia) e *intervention* (intervención), muestra que las investigaciones más recientes se enfocan en el papel activo de las mujeres en procesos formativos, científicos y comunitarios. La participación de mujeres en contextos educativos y académicos continúa siendo un aspecto fundamental de atención, como también lo enseña el tercer tópico LDA, relacionado con la desigualdad en las carreras STEM. Este hallazgo concuerda con investigaciones anteriores que demuestran que la brecha de género en ciencia y tecnología no solo se mantiene, sino que tiende a reproducirse a lo largo de los ciclos formativos, evaluativos y de liderazgo [17].

Por otro lado, los tópicos 1, 2 y 4 describen el estudio de las condiciones estructurales que impactan la sostenibilidad del bienestar femenino, especialmente en temas de salud, violencia y acceso a servicios básicos. La frecuencia de términos como *community*, *care*, *violence* y *healthcare*, así como el abordaje de la violencia de pareja durante el embarazo, demuestran la existencia de factores del entorno que obstaculizan el ejercicio pleno de derechos, especialmente en regiones rurales y vulnerables [18].

El quinto tópico, es el cáncer de cuello uterino, correspondiente a una línea de investigación que es la prevención, detección y reducción de desigualdades en salud

reproductiva, considerado de gran importancia ya que este cáncer sigue siendo una de las principales causas de muerte evitable entre mujeres en América Latina, en gran parte por la falta de programas de detección temprana [19].

En conjunto, los resultados muestran que la sostenibilidad desde un enfoque de género no puede ser entendida como una categoría aislada, sino como un tejido multidimensional que implica acceso a servicios, reconocimiento epistémico, condiciones estructurales y acción colectiva. El modelo LDA permitió visualizar cómo estas dimensiones coexisten y se entrelazan en la literatura reciente, confirmando que las investigaciones más actuales avanzan hacia paradigmas interseccionales, sensibles a la complejidad del entorno y comprometidos con la transformación social [20].

V. DISCUSION

Los resultados obtenidos demuestran una narrativa académica diversa, sin embargo algunos puntos comunes respaldan las dificultades estructurales que enfrentan las mujeres en América Latina en términos de salud, ciencia y sostenibilidad. Basado en el modelado temático y de la nube de palabras, hay suficiente evidencia para argumentar que la presencia femenina en escenarios de educación superior, el sistema de salud y las políticas públicas aún está permeado por barreras contextuales, institucionales y culturales.

Temas recurrentes como la experiencia de la mujer como protagonista de programas e iniciativas para la equidad, por ejemplo, coinciden con el reciente empleo del enfoque participativo para asegurar la sostenibilidad social, principalmente en los territorios de vulnerabilidad [20]. Por lo tanto términos como *support*, *experience*, *intervention* y *program* muestran que los textos leídos aquí están orientados a modelos de empoderamiento. Las mujeres son consideradas protagonistas en vez de solo beneficiarias: generan conocimiento y lideran los ejemplos de desarrollo autosostenible.

De manera similar, la discusión sobre la violencia de género durante el embarazo corrobora diversos estudios que describen el aumento de la violencia obstétrica, especialmente en las áreas rurales de América Latina [21]. En este escenario, se puede considerar que la sostenibilidad del sistema de salud también está condicionada por las desigualdades de género.

La aparición de la educación científica y tecnológica en la centralidad del debate es otro resultado esencial. Sobre el trabajo más reciente, esta es otra subárea donde se observa el esfuerzo de incorporar conceptos de mujeres, pero en la que existen desigualdades significativas relacionadas con el acceso, la permanencia y el liderazgo de las mujeres [22]. Como se presenta en los temas de LDA, el discurso académico aún se articula sobre la experiencia de los estudiantes, el

ambiente de aprendizaje y los factores motivacionales que aumentan la permanencia en la carrera científica.

Desde la práctica, sobresalen consecuencias para distintos sectores. En lo académico, es necesario fortalecer políticas de inclusión, tutoría y liderazgo femenino en programas STEM. En el sector salud, es importante diseñar intervenciones las cuales estén culturalmente adaptadas y sensibles a la violencia de género. Para los gobernantes, el estudio aporta evidencia clave para construir políticas sostenibles centradas en la facilidad de acceso a servicios y oportunidades.

En cuanto a las limitaciones del estudio, se debe reconocer que el análisis se realizó únicamente sobre los resúmenes (abstracts) de los artículos. Como resultado, la riqueza semántica del discurso podría verse disminuida. El primer paso para reducir esta limitación es basarse en el texto completo de los artículos en futuras investigaciones y confiar en métodos cualitativos que se centren en las características argumentales, contextuales y retóricas de los textos. Además, el corpus debería ampliarse a otras bases de datos científicas para no sesgar la fuente de selección.

En general, la investigación también sostiene indudablemente la idea de realizar enfoques interseccionales y complejos en las políticas sostenibles, con la necesidad de tener en cuenta todos los aspectos que pueden influir en la vida de las mujeres: salud, educación, participación, autonomía económica. [23].

VI. CONCLUSIONES

A través del análisis de textos y el modelado LDA, este artículo analizó un conjunto de artículos publicados recientemente en PLOS ONE 2020 y 2025, para encontrar los principales discursos académicos sobre mujeres y sostenibilidad dentro de contextos latinoamericanos.

Un hallazgo fue que, a pesar de los estudios empíricos científicos que se esfuerzan específicamente por llamar la atención sobre los problemas de las mujeres, existían deficiencias en relación con estos. Los aspectos más destacados son:

La contribución de la dimensión de género a la mejora de las políticas de salud: desde el estudio sobre la detección temprana del cáncer cervicouterino en Colombia [19].

La educación en ciencia y tecnología a una edad temprana también puede contribuir a la inclusión social.

La interrelación entre la promoción de la salud, la educación y la sostenibilidad ambiental desde una perspectiva de derechos humanos: un desafío urgente.

También se discutió cómo contar una nueva historia de este futuro, permitiendo a las mujeres una mayor participación en las iniciativas de cambio para la sostenibilidad social a nivel local, pero también prevaleció más globalmente, incluyendo su colaboración con la investigación e intervenciones diseñadas para lograr la agenda dentro de los objetivos establecidos por los ODS.

Es necesario llevar a cabo un trabajo más competente a nivel cultural sobre violencia basada en género en el futuro.

Política y práctica: Este estudio ofrece una importante lección sobre sostenibilidad a largo plazo con respecto a la equidad en el acceso a servicios y oportunidades y políticas que involucren estos temas.

A la luz de las limitaciones del estudio, se subraya que un enfoque textual puramente cuantitativo basado únicamente en el análisis de resúmenes puede no haber capturado toda la densidad semántica de estos discursos. Esta limitación podría abordarse en estudios futuros mediante la búsqueda del texto completo de los artículos (incluyendo otros enfoques cualitativos, como el análisis argumentativo, retórico o contextual). El corpus también puede usarse en otras bases de datos científicas que documenten sesgos subyacentes impulsados por la selectividad de fuentes.

Las principales conclusiones de este documento son que cuando estos paradigmas interseccionales y holísticos se aplican a todas las agendas de sostenibilidad existentes, el impacto en la vida misma, cómo se está construyendo el conocimiento comparativo sobre territorios y los caminos creados para fomentar la presencia de mujeres en los procesos de transformación sostenible, es profundo pero desigual.

AGRADECIMIENTO

Se agradece el acompañamiento de la Facultad de Ingeniería de la Universidad Libre Cali, por su apoyo técnico y administrativo a lo largo del proceso investigativo, así como la disponibilidad de la plataforma PLOS ONE, que permitió el acceso a literatura científica en acceso abierto, insumo clave para el desarrollo de los análisis textuales y temáticos aplicados.

REFERENCES

- [1] C. Pradier, D. Kozłowski y N. S. Shokida, "Science for whom? The influence of the regional academic circuit on gender inequalities in Latin America," *arXiv preprint*, Jul. 2024. [Online]. Disponible: <https://arxiv.org/abs/2407.18783> frontiersin.org/link.springer.com/arxiv.org
- [2] PLOS ONE Editors, "Comparative analysis of gender disparity in academic positions based on U.S. region and STEM discipline," *PLOS ONE*, vol. 19, no. 3, e0298736, 2024. <https://doi.org/10.1371/journal.pone.0298736>

- [3] UNESCO Institute for Statistics (UIS), *Women in Science: UNESCO Institute for Statistics Fact Sheet No.60*, June 2020. [Online]. Available: <https://uis.unesco.org/sites/default/files/documents/fs60-women-in-science-2020-en.pdf> El País+15UNESCO UIS+15EveryONE+15
- [4] Hanna Landenmark y el equipo de PLOS ONE Guest Bloggers, “International Day of Women and Girls in STEM from a Latin American perspective,” *PLOS ONE* Blog, Feb. 10 2023. [Online]. Available: <https://everyone.PLOS-ONE.org/2023/02/10/international-day-of-women-and-girls-in-stem-from-a-latin-american-perspective/> EveryONE
- [5] UNESCO Institute for Statistics (UIS), “Other policy relevant R&D indicators: Female researchers as a percentage of total researchers,” *UNESCO UIS*, 2024. [Online]. Disponible: <https://swe.org/research/2025/global-stem-workplace/swe.org>
- [6] M. Barreiro-Gen y N. Bautista-Puig, “Women in sustainability research: Examining gender authorship differences in peer-reviewed publications,” *Frontiers in Sustainability*, vol. 3, p. 959438, 2022. <https://doi.org/10.3389/frsus.2022.959438>
- [7] D. R. Ejike Ewim y A. G. Dosunmu, “Bridging gaps: the intersection of women in engineering and gender studies—challenges, advances, and future directions,” *Discover Global Society*, vol. 3, art. 43, 2025. [Online]. Disponible: <https://link.springer.com/article/10.1007/s44282-025-00184-7> link.springer.com
- [8] M. Mohammadrezaei, J. C. Marques y A. M. Huq, “Use of text mining and natural language processing techniques in analyzing sustainability reports: A systematic literature review and assessment,” *Research in Sustainability*, Dec. 2024. doi:10.13140/RG.2.2.14129.29288
- [9] A. Cui, “Utilizing Web Scraping Technology to Monitor Public Opinion on Environmental Policies,” *ITM Web of Conferences*, vol. 73, art. 03004, 2025. doi: 10.1051/itmconf/20257303004. [researchgate.net+2nature.com+2actascientific.com+2itm-conferences.org](https://researchgate.net/publication/362252222)
- [10] J. Silge y D. Robinson, *Text Mining with R: A Tidy Approach*, 2nd ed., O'Reilly Media, 2022. Disponible en: <https://www.tidytextmining.com/> (consultado 2025)
- [11] R. Ulloa, F. Mangold, F. Schmidt, J. Gilsbach y S. Stier, “Beyond time delays: How web scraping distorts measures of online news consumption,” *arXiv preprint*, Nov. 2024. Disponible: <https://arxiv.org/abs/2412.00479>.
- [12] A. Braverman-Bronstein, D. Vidania-Pérez, A. F. Ortigoza, L. Baldovino-Chiquillo, F. Diez-Canseco, J. Maslowsky, B. Sánchez, T. Barrientos-Gutiérrez y A. Diez-Roux, “Adolescent birth rates and the urban social environment in 363 Latin American cities,” *BMJ Global Health*, vol. 7, no. 10, e009737, 2022. DOI: 10.1136/bmjgh-2022-009737 <https://www.sciencedirect.com/science/article/pii/S0277953622008723>
- [13] H. Boekhout, I. van der Weijden y L. Waltman, “Gender differences in scientific careers: a large-scale bibliometric analysis,” *arXiv*, abnu:2106.12624, 2021. Disponible: <https://arxiv.org/abs/2106.12624> arXivarXiv
- [14] A. B. Araujo, N. A. M. Araujo, A. A. Moreira, H. J. Herrmann y J. S. Andrade Jr, “Gender differences in scientific collaborations: women are more egalitarian than men,” *arXiv*, abnu:1610.05937, 2016. Disponible: <https://arxiv.org/abs/1610.05937> arXiv
- [15] J. Martopullo et al., “Understanding progress and challenges in women’s health and wellbeing in exemplar countries: a time-series study identifying positive outliers,” *The Lancet Global Health*, 2024. DOI: 10.1016/S2214-109X(24)00364-4 [The Lancet](https://www.thelancet.com)
- [16] OECD, “Gender differences in education, skills and STEM careers in Latin America and the Caribbean,” *OECD Education Working Paper No. 277*, OECD Publishing, mayo 2025. Disponible: https://www.oecd.org/en/publications/2025/05/gender-differences-in-education-skills-and-stem-careers-in-latin-america-and-the-caribbean_125d2445.html?
- [17] S. León y A. Carvajal, “Challenges and solutions for women’s health in Latin America,” *Promujer / Fundación Saber*, julio 2024. Disponible: <https://promujer.org/porta/2024/07/02/access-gender-equality-and-innovation-in-womens-health/>
- [18] “Women’s studies in the Latin American context: a bibliometric approach,” *F1000Research*, vol 14:234, 2024. Disponible: https://f1000research.com/articles/14-234/v1/pdf?article_uid=410f9501-7e77-4f47-9fab-0adc8a26ad4
- [19] A. G. Lewis, D. M. Hernández, I. C. Garcés-Palacio y A. S. Soliman, “Impact of universal health insurance benefits on cervical cancer mortality in Colombia: a time-series analysis from 2006 to 2020,” *BMC Health Services Research*, vol. 24, Art. 693, 31 mayo 2024. DOI: 10.1186/s12913-024-10979-0
- [20] P. C. Bermúdez, M. Arrivilla, K. J. Torres Poveda, D. M. Castrillón Liberos, L. E. Castillo Castillo y D. Neira Acevedo, “Barriers to adherence to cytology exam: a case study in low-income Colombian women,” *BMC Health Services Research*, vol. 23, Art. 796, 25 jul. 2023. DOI: 10.1186/s12913-023-09700-4
- [21] G. Widegren y J. Sand, *Gender and Sustainability: An International Research Review*, Swedish Secretariat for Gender Research, University of Gothenburg, 2021. Informe que revisa literatura académica internacional sobre los vínculos entre género e implementación de los ODS (2015–2021), posicionando la sostenibilidad ambiental dentro de una perspectiva de equidad de género. Disponible en el resumen PDF oficial: https://www.gu.se/sites/default/files/2021-12/Gender-and-Sustainability_An-International-Research-Review.pdf
- [22] H. Boekhout, I. van der Weijden y L. Waltman, “Gender differences in scientific careers: a large-scale bibliometric analysis,” *arXiv*, arXiv:2106.12624, 2021. Disponible en: <https://arxiv.org/abs/2106.12624>
- [23] OECD, *Gender and the Environment: Building Evidence and Policies to Achieve the SDGs*, OECD Publishing, Paris, 2021. DOI: 10.1787/3d32ca39-en