

# Género e inteligencia artificial en la educación en México: la realidad invisible

Urith Ramírez-Mera; Zulma Raquel Zeballos Pinto

## RESUMEN

La integración de la inteligencia artificial (IA) en la educación mexicana es prometedora, pero opera replicando desigualdades estructurales históricas. Este estudio analiza críticamente la manifestación y reproducción de sesgos de género en herramientas de IA aplicadas en contextos educativos nacionales. Mediante una revisión de literatura entre los años 2020 y 2025 y un análisis cualitativo con lente interseccional, se examinaron cinco estudios clave. Los resultados identifican sesgos sistémicos en: 1) representaciones visuales que estereotipan e invisibilizan a las mujeres; 2) modelos predictivos con errores de clasificación basados en género, y 3) diseños tecnológicos que excluyen a las mujeres como usuarias y creadoras. Estos hallazgos tienen raíces en la falta de datos interseccionales, la subrepresentación femenina en el desarrollo tecnológico y el acceso digital desigual. La discusión subraya que estos sesgos no son meros errores técnicos, sino la manifestación de exclusiones históricas, cuyo impacto en la educación desalienta vocaciones y perpetúa estereotipos. Se concluye que la mitigación de este fenómeno exige un enfoque multidimensional que combine las políticas de datos con la perspectiva de género, el desarrollo de una IA con enfoque feminista e interseccional y la integración de una alfabetización algorítmica en la práctica pedagógica.

**Palabras clave:** sesgos algorítmicos, género, educación en México, inteligencia artificial, revisión de literatura.

**Cómo citar:** Ramírez-Mera, U., Zeballos, Z. (2026). Género e inteligencia artificial en la educación en México: la realidad invisible. En Peña Guzmán, C. *Mujeres y su impacto en la ciencia y tecnología latinoamericana*. High Rate Consulting. <https://doi.org/10.38202/mujeresimpacto3>

# Gender and artificial intelligence in education in Mexico: the invisible reality

## ABSTRACT

The integration of artificial intelligence (AI) in Mexican education is promising, but it operates by replicating historical structural inequalities. This study critically analyzes the manifestation and reproduction of gender biases in AI tools applied in national educational contexts. Through a literature review between 2020 and 2025 and a qualitative analysis with an intersectional lens, five key studies were examined. The results identify systemic biases in: 1) visual representations that stereotype and render women invisible; 2) predictive models with classification errors based on gender; and 3) technological designs that exclude women as users and creators. These findings have roots in the lack of intersectional data, the underrepresentation of women in technological development, and unequal digital access. The discussion underscores that these biases are not mere technical errors, but the manifestation of historical exclusions, whose impact on education discourages vocations and perpetuates stereotypes. It is concluded that mitigating this phenomenon requires a multidimensional approach that combines data policies with a gender perspective, the development of AI with a feminist and intersectional approach, and the integration of algorithmic literacy into pedagogical practice.

**Keywords:** algorithmic biases; gender; education in Mexico; artificial intelligence; literature review.

## INTRODUCCIÓN

La expansión acelerada de la inteligencia artificial (IA) en los sistemas educativos ha inaugurado una etapa en la que las decisiones pedagógicas, administrativas e incluso afectivas circulan a través de arquitecturas algorítmicas que rara vez son transparentes y evaluadas. Bajo la apariencia de eficiencia y neutralidad, estos sistemas operan como engranajes que condensan historias de desigualdad, ausencias estructurales y formas de poder sedimentadas en los datos. Lejos de constituir simples herramientas técnicas, los algoritmos participan en la configuración cotidiana de oportunidades educativas, delimitando quién es escuchado, quién es visto y quién queda al margen. En este terreno movedizo, los sesgos algorítmicos emergen no como fallas excepcionales, sino como síntomas de un ecosistema sociotécnico que replica, amplifica o silencia identidades y trayectorias.

El ámbito educativo resulta especialmente vulnerable a estas dinámicas porque se entrelaza con procesos formativos que determinan futuros personales y colectivos. Desde sistemas de predicción del rendimiento académico hasta evaluadores automatizados de voz, imagen o texto, la IA es capaz de intervenir en decisiones de alto impacto sin ofrecer siempre mecanismos claros de auditoría o rendición de cuentas. Ello cobra especial relevancia cuando los algoritmos interactúan con marcadores sociales como el género, la etnia, la clase o la lengua, generando patrones de exclusión que pueden pasar inadvertidos precisamente por su carácter automatizado (Lozano-Mulet, 2025).

En América Latina, y particularmente en México, estas tensiones adquieren un matiz adicional. La región se encuentra atrapada entre la promesa transformadora de la IA y la persistencia de desigualdades estructurales que condicionan tanto su implementación como su evaluación (Ilbay, 2024); sin embargo, se manifiesta la necesidad de integrar esta tecnología al contexto educativo ante desafíos éticos y regulatorios. Al mismo tiempo, surge un impulso creciente por desarrollar perspectivas críticas, decoloniales y culturalmente situadas que cuestionan el rol del Sur Global como mero receptor de tecnologías foráneas. Desde esta perspectiva, tal como señalan Huesca y Fueyo Hernández (2025), analizar los sesgos algorítmicos en la educación mexicana implica no solo examinar el desempeño técnico de los modelos, sino también indagar cómo estos interactúan con estructuras históricas de desigualdad y con los imaginarios sociales que moldean qué cuerpos, voces y saberes son considerados legítimos.

El presente estudio se inscribe en este cruce de caminos. Busca iluminar las formas en que los sesgos, particularmente los asociados al género, se manifiestan, se reproducen o se refuerzan en el uso de herramientas de IA dentro de entornos educativos en México, que, si bien ha generado debates internacionales, en el contexto mexicano es necesario profundizar (Salcido Ledezma, 2024). Al hacerlo, procura nutrir un debate urgente: el de cómo diseñar, regular y enseñar con tecnologías que no solo respondan a estándares

técnicos, sino que también honren la diversidad, la justicia social y la autonomía de las personas aprendientes (Lozano-Mulet, 2025). En un país donde las brechas educativas y tecnológicas no son abstractas, sino palpables (Huesca y Fueyo Hernández, 2025), comprender y enfrentar los sesgos algorítmicos no constituye un lujo intelectual, sino una necesidad ética impostergable.

### Sesgos algorítmicos en educación

Los sesgos algorítmicos, específicamente en el contexto educativo, se refieren a discriminaciones estructurales incorporadas en sistemas de IA que afectan desproporcionadamente a grupos específicos de estudiantes, pero también a otros actores educativos. Según Baker y Hawn (2022), estos sesgos se manifiestan cuando el rendimiento de un modelo es sustancialmente mejor o peor a través de grupos mutuamente excluyentes, reproduciendo y amplificando desigualdades estructurales existentes. De acuerdo con Suresh y Guttag (2020), se identifican seis dimensiones críticas:

1. **Sesgo histórico:** reproducción de decisiones discriminatorias pasadas en datos educativos.
2. **Sesgo de representación:** subrepresentación de grupos marginados en conjuntos de entrenamiento.
3. **Sesgo de medición:** variables que carecen de validez de constructo equitativa entre grupos.
4. **Sesgo de agregación:** modelos únicos inadecuados para poblaciones diversas.
5. **Sesgo de evaluación:** pruebas no representativas de poblaciones objetivo.
6. **Sesgo de despliegue:** uso de algoritmos en contextos no previstos.

Esta clasificación, que distingue entre los diferentes tipos de sesgos, permite reconocer el alcance de la IA dentro del contexto social y su caracterización como artefactos culturales que encarnan los valores y perspectivas de sus desarrolladores (Ferrero y Gewerc Barujel, 2019). debe destacarse que esta perspectiva rechaza la noción de neutralidad tecnológica y enfatiza cómo los sistemas educativos algorítmicos reflejan y refuerzan estructuras de poder existentes, lo cual es de relevancia crucial en esta investigación.

En concordancia con ello, autores como Kasun et al. (2024) introducen una crítica decolonial que cuestiona el epistemicidio del Sur Global en el desarrollo de IA educativa. Los mismos argumentan que la inclusión de conocimientos marginados no es solo cuestión de representación, sino de soberanía epistemológica (Huasca y Fueyo Hernández, 2025), donde el Sur Global debe transitar de consumidor a generador de IA.

**Tabla 1.**  
Sesgos algorítmicos en IA

| SESGO ALGORÍTMICO       | DESCRIPCIÓN  |
|-------------------------|--|
| Raza/etnia              | Modelos de predicción de graduación con tasas más altas de falsos positivos para estudiantes blancos y falsos negativos para latinos (Anderson et al., 2019) |
| Género                  | Algoritmos de predicción de abandono con menor precisión para estudiantes masculinos (Gardner et al., 2019)  |
| Nacionalidad            | Sistemas de corrección automática que califican desigualmente según país de origen (Bridgeman et al., 2012)  |
| Estatus socioeconómico  | Modelos entrenados en contextos privilegiados con pobre generalización a entornos de bajos recursos  |
| Discapacidades          | Sistemas de reconocimiento de voz con precisión reducida para personas con impedimentos del habla (Loukina y Buzick, 2017)                                   |
| Lengua Materna/Dialecto | Evaluaciones automáticas que penalizan variaciones dialectales   |

Por otro lado, los sesgos algorítmicos operan por medio de múltiples ejes de identidad simultáneos (raza, género, clase, capacidad, etc.). La investigación evidencia vacíos significativos en el análisis interseccional, particularmente para estudiantes no binarios, indígenas y otras identidades complejas (Tabla 1).

Si analizamos el impacto que tienen los algoritmos de la IA en el contexto educativo, es preciso reflexionar y colocar en tela de juicio los aspectos éticos y políticos que definen el uso de este tipo de tecnologías, reconocer a los beneficiarios y a los perjudicados por los algoritmos educativos e identificar en qué medida los sistemas algorítmicos limitan o expanden la autonomía de los aprendices.

Desde una perspectiva pedagógica, resulta necesario que exista una distinción entre tratamiento idéntico y tratamiento apropiado hacia el alumnado según necesidades específicas, es decir, evitar confundir equidad e igualdad, del mismo modo que es imperativo evitar suposiciones culturales incorporadas en lo que se considera “conocimiento válido”. Finalmente, la evaluación y los indicadores de desempeño también se enfrentan a tensión entre algoritmos para apoyo al aprendizaje y la toma de decisiones de alto impacto. Sin embargo, estos aspectos no son temas nuevos: los algoritmos educativos no crean sesgos nuevos, sino que automatizan y escalan discriminaciones históricas presentes en los datos y prácticas educativas tradicionales. La implementación acelerada de IA en educación, sin salvaguardas adecuadas, representa un riesgo existencial para la equidad educativa, particularmente para poblaciones históricamente marginadas.

Sin embargo, no todo está perdido. Cuando se diseñan conscientemente, los algoritmos educativos pueden detectar y corregir sesgos humanos, ofreciendo potencial para

sistemas más equitativos que los gestionados exclusivamente por personas. Para ello, es necesario replantear la competencia digital no solo entre el profesorado y el estudiantado, sino en los profesionales encargados del desarrollo de dichas tecnologías. La reflexión crítica es necesaria para reconocer el impacto “casi invisible” de los algoritmos en la reproducción de discriminaciones a grupos vulnerables.

Por ello, abordar los sesgos algorítmicos en educación requiere un enfoque multidimensional que integre comprensión técnica, crítica sociopolítica y práctica pedagógica reflexiva. La equidad algorítmica no es un estado alcanzable mediante soluciones puramente técnicas, sino un proceso continuo de vigilancia crítica, adaptación cultural y compromiso ético. La promesa transformadora de la IA en educación solo se realizará plenamente cuando dejemos de tratar los sesgos algorítmicos como errores técnicos para comprenderlos como expresiones de tensiones sociales más amplias que demandan respuestas igualmente amplias y colaborativas.

### **La Inteligencia Artificial (IA) en la educación mexicana**

La IA ha emergido como una fuerza disruptiva a nivel global, prometiendo transformar sectores enteros, entre ellos, la educación. En el contexto mexicano, este potencial se vislumbra en medio de un panorama complejo, marcado por un crecimiento investigador reciente, profundas desigualdades estructurales y la urgente necesidad de desarrollar enfoques propios que respondan a las realidades locales.

El desarrollo de la IA en México es un campo joven, pero de crecimiento acelerado. Una revisión sistemática de la producción académica (Uc Castillo et al., 2025) revela que, si bien las publicaciones fueron escasas durante más de una década, han experimentado un auge significativo en los últimos cinco años, con picos notorios en 2021 y 2022.

Este interés creciente, sin embargo, presenta sesgos geográficos y temáticos. La investigación se concentra predominantemente en instituciones del centro del país, como la Universidad Nacional Autónoma de México (UNAM), y se ha enfocado principalmente en el área de las Ciencias Sociales, donde abundan los estudios sobre la percepción de docentes y estudiantes respecto al uso de plataformas educativas y tecnologías de la información (Uc Castillo et al., 2025).

Este cuerpo de trabajo inicial es valioso, pero también expone una paradoja fundamental: México, siendo la mayor nación de habla hispana, es tratado en el ámbito de la IA como un “recurso de baja densidad”, careciendo de modelos y conjuntos de datos específicamente diseñados para las particularidades del español mexicano (Chizhikova et al., 2024). Al mismo tiempo, se evidencia la falta de investigación de IA aplicada a otras áreas disciplinares como las ciencias ingenieriles, artes y salud.

Precisamente, la evaluación de tecnologías avanzadas como los Sistemas de Reconocimiento Automático de Voz (ASR) evidencia las consecuencias de esta omisión. Estudios

como el de Chizhikova et al. (2024) demuestran que modelos de vanguardia como Whisper y MMS, entrenados para un “español general”, presentan sesgos algorítmicos que se tornan críticos al adoptar una mirada interseccional.

Se ha identificado un sesgo consistente, aunque leve, a favor de las voces masculinas, el cual se potencia drásticamente al considerar atributos como el tono de piel, la edad y, de manera crucial, el timbre vocal. La interpretación de este resultado explica el hecho de que dado que los hablantes mexicanos suelen emplear un habla de tono más alto, las voces femeninas con timbres agudos son las más penalizadas, lo que explica que se presenten tasas de error significativamente mayores.

En consecuencia, la implementación acrítica de estas tecnologías en aulas mexicanas, por ejemplo, para transcribir clases o evaluar la pronunciación, corre el riesgo de profundizar las desigualdades existentes, perjudicando de manera desproporcionada a estudiantes mujeres, de piel morena o con particularidades en su habla.

Frente a este panorama de desafíos técnicos y éticos, surge una luz desde la práctica educativa ground-up. El trabajo de Kasun et al. (2024) en un taller con jóvenes marginados de Cuernavaca demuestra que el interés y la capacidad para entender y utilizar la IA son universales, trascendiendo las barreras socioeconómicas. El éxito de esta intervención radicó en una pedagogía culturalmente relevante que reemplazó ejemplos ajenos (como aspiradoras robóticas) por otros significativos para los jóvenes, como los riesgos de la clonación de voz para extorsión o la creación de imágenes de vaqueros locales mediante generadores de IA.

El análisis en profundidad de este caso subraya que el “Sur Global”, y México dentro de él, no debe concebirse como un mero consumidor de tecnología, sino como un potencial generador de IA, cuyos conocimientos y realidades pueden y deben enriquecer el desarrollo de estas herramientas (Kasun et al., 2024). Por ello que la colaboración transdisciplinaria entre tecnólogos y educadores con sensibilidad antropológica y de equidad fue un aspecto clave para diseñar una experiencia de aprendizaje exitosa y ética que minimizara el sesgo de género.

Sin duda, la IA en la educación mexicana se encuentra en una encrucijada, ya que, por un lado, existe un momentum investigador creciente y un potencial inmenso para personalizar el aprendizaje, automatizar tareas administrativas y ofrecer nuevas herramientas de creación. Por otro, este potencial está amenazado por sesgos algorítmicos intrínsecos, una distribución desigual de la capacidad investigadora y el riesgo de importar modelos pedagógicos y tecnológicos culturalmente incongruentes al contexto mexicano.

Todo ello demuestra que el camino a seguir exige un compromiso con el desarrollo de una IA soberana y crítica. Esta deberá ser capaz de abordar proactivamente los sesgos desde una perspectiva interseccional que posibilite la formación de los docentes no solo como usuarios, sino como críticos y creadores. De igual modo el desarrollo de la misma deberá posibilitar y estimular que se centre el diseño

de políticas e intervenciones en la equidad y la relevancia cultural. Solo de este modo, la inteligencia artificial podrá cumplir su promesa de ser una verdadera herramienta de inclusión y transformación educativa para todos los mexicanos.

### Objetivo:

El objetivo de esta investigación es analizar e identificar la presencia y reproducción de sesgos de género en el uso de la IA aplicada en contextos educativos en México. Por ello, la pregunta central que guía esta investigación es la siguiente:

¿De qué manera se manifiestan y reproducen los sesgos de género en las herramientas de IA aplicadas en contextos educativos mexicanos, y qué estrategias de mitigación pueden implementarse desde la política pública, el desarrollo tecnológico y la práctica pedagógica?

### METODOLOGÍA

Este estudio se fundamenta en un diseño de investigación cualitativa. El mismo emplea una revisión integradora de la literatura como estrategia central (Regmi, 2023). Este enfoque es particularmente idóneo para sintetizar conocimientos provenientes de diversos tipos de fuentes como literatura teórica, empírica, informes de política pública y estudios de caso, lo cual permite una comprensión holística y crítica de un fenómeno complejo y emergente.

#### Estrategia de búsqueda y selección de fuentes

Para dar respuesta al objetivo planteado la recolección de información se realizó a través de una búsqueda amplia y exhaustiva en bases de datos académicas especializadas como: Scopus, Web of Science, RedALyC y ERIC, seleccionadas por su solidez en los campos de educación, estudios de género y ciencias de la computación. Para garantizar la relevancia contextual, esta búsqueda se complementó con una revisión en Google Scholar, con el fin de incorporar informes institucionales, políticas públicas y literatura gris de vital importancia.

La búsqueda se centró en documentos publicados entre 2020 y 2025, un período que coincide con la aceleración en la adopción de la IA en educación y la consolidación de la agenda de equidad global. La estrategia de búsqueda empleó una combinación de términos en español e inglés, estructurados en tres ejes conceptuales interconectados:

1. **Inteligencia Artificial y Tecnología:** “inteligencia artificial”, “sesgo algorítmico”, “reconocimiento de voz”, “automatic speech recognition”, “machine learning”, “equidad en IA”, “algorithmic bias”, “AI ethics”.
2. **Género y Educación:** “género”, “equidad”, “educación”, “mujeres”, “niñas”, “STEM”, “brecha digital de género”, “gender equity”, “education”, “women in STEM”, “digital gender gap”.

3. **Contexto Geopolítico:** “México”, “América Latina”, “Global Sur”, “política educativa”, “Plan de Estudios 2022”, “Latin America”, “Global South”, “educational policy”.

Para refinar los resultados, se establecieron criterios de selecciones que pueden ser visibles en la Tabla 2.

**Tabla 2.**  
Sesgos algorítmicos en IA

| CRITERIOS DE INCLUSIÓN  | CRITERIOS DE EXCLUSIÓN   |
|---|--|
| Artículos de investigación, revisiones sistemáticas y estudios de caso. | Libros completos, editoriales y artículos de opinión no sustentados empíricamente.                       |
| Literatura publicada entre 2020 y 2025.                                 | Literatura publicada antes de 2020.  |
| Documentos en español, inglés o portugués.                              | Enfoque exclusivo en contextos educativos de niveles superiores sin relevancia para la educación básica. |
| Enfoque explícito en la intersección IA, género/equidad y educación.    | Estudios puramente técnicos de IA que no aborden dimensiones sociales o educativas.                      |
| Documentos que enfaticen en el contexto mexicano.                       | Documentos internacionales que no consideren el contexto educativo mexicano.                             |

#### Proceso de análisis de datos

El análisis se llevó a cabo en dos fases principales, guiadas por el enfoque de Análisis Crítico del Discurso y la perspectiva interseccional:

1. **Fase 1.** Codificación y categorización. Los documentos seleccionados fueron analizados utilizando el software de análisis cualitativo Atlas.ti 9. Se desarrolló un libro de códigos mixto que permitió categorizar la información en dimensiones clave alineadas con el objetivo y pregunta de investigación (Tabla 3).
2. **Fase 2.** Análisis interseccional. Se examinaron los discursos presentes en los documentos analizados, identificando la presencia, ausencia o superficialidad del lenguaje sobre equidad de género y mitigación de sesgos. Simultáneamente, se analizó cómo las categorías de género se intersectan con otras, tales como la clase socioeconómica, la etnia y la ubicación geográfica en los estudios empíricos y experiencias pedagógicas revisadas.

Para garantizar transparencia y sistematicidad, se siguió un proceso de cinco pasos, adaptado de Regmi (2023) que especifica, de manera detallada, cómo se realizó la selección de los documentos que integraron el corpus de esta investigación. Es importante señalar que se revisaron bases de datos que no se consideraron al inicio de la investigación, pues se identificó que este tema ha sido poco estudiado.

Ejemplo de ello es que, a pesar de que uno de los criterios de selección fue la fecha de publicación (2020), el primer documento encontrado fue de 2022, siendo este año el inicio de la discusión de la IA en materia de perspectiva de género en el contexto educativo mexicano. (Tabla 4)

**Tabla 3.**  
Libro de categorías y códigos

| CATEGORÍA                                    | CÓDIGO                                       | CONCEPTOS CLAVES   |
|--|--|--|
| Manifestaciones de sesgos de género en la IA | Sesgos en representaciones visuales          | Subrepresentación femenina, estereotipos raciales y etarios, cliché del "científico hombre blanco" |
|  | Sesgos en modelos predictivos                | Clasificación errónea por género, desbalance en datos, confusión hombre→mujer                      |
|  | Sesgos en diseño tecnológico                 | Tecnología percibida como "para hombres", exclusión en diseño de productos                         |
|  | Sesgos en datos de entrenamiento             | Datos no representativos, falta de diversidad en conjuntos como LAION                              |
| Orientaciones estructurales de los sesgos    | Brechas en participación femenina            | Baja representación en STEM, especialmente en IA (22 % global)                                     |
|  | Falta de datos interseccionales              | Datos no desglosados por género, raza, ubicación, etc.   |
|  | Exclusión en puestos de liderazgo            | Techos de cristal, comisiones mayoritariamente masculinas  |
|  | Acceso desigual a tecnología                 | Brecha digital, propiedad masculina de dispositivos  |
| Impacto en contextos educativos              | Perpetuación de estereotipos profesionales   | Imágenes IA refuerzan roles tradicionales, desalientan vocaciones                                  |
|  | Modelos predictivos con sesgo                | Clasificación de estudiantes por género basada en percepciones                                     |
|  | Falta de modelos educativos inclusivos       | Contenidos que no consideran necesidades femeninas   |
|  | Alfabetización digital crítica               | Necesidad de educar sobre sesgos algorítmicos  |
| Estrategias de mitigación                    | IA feminista y perspectiva de género         | Diseño interseccional, cocreación, auditoría ética   |
|  | Recolección de datos interseccionales        | Construcción de datasets representativos, políticas de datos abiertos                              |
|  | Educación crítica en IA                      | Alfabetización algorítmica, auditorías participativas en aula                                      |
|  | Políticas públicas con perspectiva de género | Cuotas en comisiones, incentivos para inclusión, educación STEM                                    |
|  | Técnicas técnicas de balanceo                | Oversampling, normalización, selección de características  |

**Tabla 4.**  
Sesgos algorítmicos en IA

| PASOS  | PROCESOS   |
|--|--|
| Paso 1. Formulación de propósito               | Definir el análisis crítico de la intersección IA-género-educación en México, identificando sesgos, políticas y prácticas pedagógicas.   |
| Paso 2. Búsqueda de literatura                 | Bases de datos: Scopus, WoS, RedAllyC, ERIC, Google Scholar y sitios de organismos. Términos de búsqueda basados en los tres ejes conceptuales. Periodo: 2020-2025.  |
| Paso 3. Selección y evaluación                 | Documentos iniciales: Obtuvimos un total de 520 documentos, de los cuales, 478 (bases de datos seleccionadas) y 42 (de otras fuentes). Después de un screening y considerando los criterios de inclusión y exclusión, se recuperaron 35 documentos cuyos ejes temáticos giraban en torno a las variables de nuestra investigación. Posteriormente, se realizó un análisis profundo de los documentos, seleccionando 5 documentos. El resto se excluyó ya que las temas de la investigación no formaban parte de los ejes centrales de la investigación, o estaban contextualizadas en otras latitudes latinoamericanas. La selección final fue de 5 documentos |
| Paso 4. Análisis temático                      | Codificación con Atlas.ti 9 y desarrollo de temas a través del Análisis crítico del discurso y la lente interseccional.  |
| Paso 5. Hallazgos y síntesis de la información | Resultados de la investigación con énfasis en la pregunta de investigación.  |

## RESULTADOS Y DISCUSIÓN

Esta investigación tomó como base cinco artículos científicos (Tabla 5).

Estos fueron los documentos que cumplieron con los criterios de selección definidos al inicio de la investigación. A partir de su análisis, se presentan los resultados y discusión.

### Las manifestaciones: cuando la IA nos devuelve una imagen incompleta

La evidencia es contundente y preocupante. Al pedirle a una IA que genere la imagen de un profesional forense, se observará que en el 90 % de los casos nos devolverá la figura de un hombre, predominantemente blanco, de edad mediana y con atributos estereotípicos como gafas y bata de laboratorio (Torres Zuñiga, 2024).

Esta no es una mera anécdota. Es la expresión sintomática de un problema estructural. Los modelos predictivos entrenados con datos de percepciones estudiantiles pueden alcanzar una precisión del 96 % en el entrenamiento, pero un análisis más profundo revela un patrón sistemático: el error más común es clasificar a los estudiantes hombres como mujeres (Ibarra-Vázquez et al., 2024). La IA, en su objetividad aparente, está aprendiendo y perpetuando nuestros prejuicios. Estos sesgos no se limitan a lo visual o predictivo. Se traducen en diseños tecnológicos que no consideran a las mujeres como usuarias o creadoras.

**Tabla 5.**  
Artículos analizados

| AUTORES               | TITULO  | AÑO  | ADSCRIPCIÓN                             |
|-----------------------|---|------|---|
| Ibarra-Vazquez et al. | Gender prediction based on University students' complex thinking competency: An analysis from machine learning approaches                       | 2023 | Tecnológico de Monterrey                |
| Medina Vidal et al.   | Financial inclusion of vulnerable sectors with a gender perspective: Risk analysis model with artificial intelligence based on complex thinking | 2025 | Tecnológico de Monterrey                |
| Spitz et al.          | A gender perspective on Artificial Intelligence (AI) and the tech-sector Position paper for the multi-stakeholder dialogue in Mexico            | 2023 | No especifica                           |
| Torres Zúñiga         | Sesgos en la representación de profesionales de la ciencia forense en imágenes generadas por inteligencia artificial                            | 2025 | Universidad Nacional Autónoma de México |
| Vargas-Solar          | Intersectional study of the gender gap in STEM through the identification of missing datasets about women: A multisided problem                 | 2022 | Universidad de Lyon                     |

Como se señaló en el diálogo multisectorial en la Ciudad de México, la tecnología suele ser percibida como un espacio “de y para hombres”, lo que se traduce en productos, desde plataformas educativas hasta algoritmos de crédito, en donde se marginaliza a la mujer con educación básica (Medina Vidal et al., 2025), que no responden a las realidades, necesidades o cuerpos femeninos (Spitz et al., 2023).

### Las raíces:

#### Un problema de datos, poder y representación

¿Por qué la IA reproduce estos sesgos? La respuesta yace, en gran medida, en sus entrañas: los datos con los que se alimenta. Nos enfrentamos a una crisis de datos faltantes que no se han visibilizado en la construcción de las bases de datos de las que se alimentan las IA. No existen conjuntos de datos integrales e interseccionales que documenten de manera adecuada las contribuciones, la participación y las trayectorias de las mujeres mexicanas en Ciencia, Tecnología, Ingeniería y Matemáticas (STEM) (Vargas-Solar, 2022) o en solicitudes de crédito con base en la escolaridad (Medina Vidal et al., 2025).

Esta invisibilidad estadística tiene una correlación directa con la realidad, pues en México, aunque las mujeres representan más del 50 % de los estudiantes de posgrado en

STEM, solo conforman el 34 % del Sistema Nacional de Investigadores (SNI). La brecha se ensancha en los niveles más altos: únicamente el 5 % de las investigadoras alcanzan el nivel III, y en áreas como la computación, la cifra es aún más desoladora (Vargas-Solar, 2022). Sin datos, no hay evidencia; sin evidencia, no hay políticas públicas efectivas y, sin políticas, el ciclo de exclusión se perpetúa.

El origen del problema, por tanto, no es meramente técnico, sino profundamente estructural. Es un problema de poder y representación que se manifiesta en comisiones evaluadoras mayoritariamente masculinas, en la concentración del desarrollo tecnológico en ciertos grupos sociales y en un acceso desigual a la infraestructura digital, donde, en muchos hogares rurales, el teléfono inteligente —la llave al mundo digital— sigue estando en manos de los hombres (Spitz et al., 2023).

#### El impacto en la educación: formando futuros sesgados

Las consecuencias en el ámbito educativo son profundas. Cuando un estudiante busca referentes visuales de científicos en internet y se encuentra sistemáticamente con imágenes de hombres generadas por IA, internaliza un mensaje poderoso y dañino: la ciencia no es para mujeres. Esto desalienta las vocaciones científicas tempranas en las niñas y refuerza estereotipos que la sociedad lleva décadas intentando erradicar. Lo mismo sucede con la validación de propuestas de emprendimiento tomando como base el nivel educativo y el género, dando como resultado una interrelación entre los sistemas de pensamiento y de innovación (Medina Vidal et al., 2025).

Sin embargo, existe una luz de esperanza pedagógica. Cuando se involucra a los estudiantes en la auditoría crítica de estas herramientas, como en la actividad de dos horas documentada por Torres Zúñiga (2024), se produce un despertar de la conciencia crítica. Los estudiantes pasan de ser consumidores pasivos a usuarios críticos, capaces de cuestionar la supuesta neutralidad de la tecnología. Esta alfabetización algorítmica se revela como una competencia fundamental para la ciudadanía del siglo XXI.

Romper este ciclo requiere una intervención coordinada y multidimensional, pues no basta con ajustar los algoritmos; hay que transformar el ecosistema completo. Desde la política pública, es urgente crear un Observatorio Nacional de Sesgos Algorítmicos que audite las herramientas de IA utilizadas en el sector educativo público, ya que forma parte de las múltiples necesidades de las instituciones educativas (Huesca y Fueyo Hernández, 2025). Además, se deben implementar políticas de datos abiertos con perspectiva de género que obliguen a la transparencia y fomenten la creación de datasets representativos.

Desde el desarrollo tecnológico, la promoción de una IA con enfoque feminista es crucial. Esto implica integrar perspectivas interseccionales de género en todo el ciclo de desarrollo de la IA, desde el diseño hasta la implementación, priorizando la cocreación con comunidades diversas y estableciendo comités de ética robustos. Cuando se mira

desde una perspectiva pedagógica, la educación debe asumir un rol protagónico. Integrar la crítica de algoritmos en los currículos de todos los niveles educativos, formar a docentes en la identificación de sesgos tecnológicos y promover la creación de contenidos educativos con referentes femininos diversos son acciones inmediatas y de alto impacto.

La inteligencia artificial en la educación mexicana no es un ente ajeno; es un producto social que encapsula nuestras virtudes y nuestras taras. Los sesgos de género que exhibe no son un error técnico a corregir, sino el síntoma de exclusiones históricas y estructurales. Abordarlos exige mirar más allá del código y comprometernos con una transformación cultural, educativa y política que coloque la equidad en el centro del desarrollo tecnológico. El futuro de la educación en México no puede estar en manos de algoritmos que repiten los errores del pasado o que estén a merced de los intereses del mercado internacional; debe construirse con herramientas que nos ayuden a imaginar y crear un futuro más justo e inclusivo para todos.

## CONCLUSIONES

La integración de la inteligencia artificial en el sistema educativo mexicano se revela como un espejo tecnológico que refleja con inquietante precisión las desigualdades estructurales del país. Los hallazgos de esta investigación demuestran contundentemente que los sesgos algorítmicos no constituyen meras anomalías técnicas, sino la cristalización digital de exclusiones históricas de género, clase y origen étnico-racial que han caracterizado el desarrollo nacional. Además, es un tema que debe ser estudiado y del que se tiene poca información al respecto; prueba de ello es la escasa literatura encontrada para la realización de este análisis documental.

La evidencia documentada a través de cinco estudios independientes confirma que la IA educativa en México enfrenta una paradoja fundamental: mientras se promociona como herramienta de modernización e igualdad de oportunidades, en la práctica sistematiza y escala discriminaciones preexistentes. Desde las imágenes generativas que invisibilizan a las profesionales científicas (Torres Zuñiga, 2024) hasta los modelos predictivos que replican estereotipos de género (Ibarra-Vázquez et al., 2024), pasando por los sistemas de reconocimiento de voz que penalizan características fonéticas mexicanas (Chizhikova et al., 2024), se configura un ecosistema tecnológico que naturaliza la exclusión.

Este patrón responde a lo que Suresh y Guttag (2020) identifican como sesgos históricos, de representación y medición, pero con una particularidad mexicana: la intersección entre la herencia colonial en la producción de conocimiento y las dinámicas contemporáneas de exclusión digital. Como señalan acertadamente Kasun et al. (2024), México y el Sur Global transitan de consumidores pasivos a potenciales generadores de IA, pero este tránsito encuentra barreras estructurales en la concentración geográfica de la investi-

gación (Uc Castillo et al., 2025) y en la caracterización del español mexicano como “recurso de baja densidad” (Chizhikova et al., 2024).

Frente a este panorama, sería erróneo caer en el determinismo tecnológico. La investigación identifica espacios significativos de agencia y resistencia que apuntan hacia alternativas transformadoras. En primer lugar, urge profundizar en una alfabetización algorítmica crítica como herramienta de empoderamiento, demostrando que, cuando los estudiantes y los docentes desarrollan capacidad de auditoría tecnológica (Torres Zuñiga, 2024), pueden transitar de usuarios pasivos a críticos creadores. Además, prácticas pedagógicas culturalmente relevantes, como el taller documentado por Kasun et al. (2024) en Cuernavaca, evidencian que el interés y la capacidad para entender y utilizar la IA son universales cuando se reemplazan ejemplos ajenos por problemas significativos localmente. Asimismo, las iniciativas de datos abiertos y transparencia algorítmica sugieren caminos para construir soberanía tecnológica a partir del reconocimiento de los datos faltantes (Vargas-Solar, 2022) y la necesidad de conjuntos de datos representativos.

Los hallazgos realizados resaltan la importancia de repensar el contrato social que subyace a la integración tecnológica en educación. Se requieren mecanismos de gobernanza multiactor que superen el actual modelo de importación acrítica de soluciones estandarizadas. Específicamente, esta investigación sugiere el diseño de políticas de datos con perspectiva interseccional que tengan como prioridad el desarrollo de conjuntos de datos representativos de la diversidad mexicana, así como la generación de un marco regulatorio para auditorías algorítmicas independientes en el sector educativo público, con especial atención a sesgos de género y exclusión lingüística. Por otro lado, es indispensable la formación docente en materia de crítica algorítmica que trascienda la mera capacitación instrumental para desarrollar capacidad de evaluación ética y pedagógica de tecnologías educativas.

Sin embargo, se observa una inversión en investigación centralizada, lo cual limita fuertemente este campo, pues existe una concentración actual en instituciones del centro del país y deja fuera otros contextos que han sido históricamente marginalizados.

De esta manera, la IA en la educación mexicana se revela, en última instancia, como un proyecto político-pedagógico antes que meramente técnico. Su desarrollo futuro estará determinado no por capacidades computacionales, sino por decisiones fundamentales sobre qué tipo de sociedad queremos construir y qué valores deben guiar nuestra relación con la tecnología. La promesa transformadora de la IA educativa solo se realizará plenamente cuando dejemos de tratarla como solución técnica importada y la reconozcamos como espacio de disputa cultural donde se definen los futuros posibles para las próximas generaciones de mexicanos. El camino hacia una IA educativa verdaderamente equitativa exige confrontar no solo los sesgos en los algoritmos, sino las exclusiones en la estructura social que estos reflejan y amplifican.

En este contexto, la investigación futura deberá profundizar en el desarrollo de metodologías de evaluación contextualizadas y en el diseño de intervenciones pedagógicas críticas que permitan transitar de la identificación de problemas a la construcción colectiva de alternativas. Solo

mediante este compromiso con una IA educativa crítica, decolonial y éticamente fundamentada podrá cumplirse la promesa de una educación verdaderamente inclusiva para todos los mexicanos.

## REFERENCIAS

- Anderson, H., Boodhwani, A., & Baker, R.S. (2019). *Assessing the fairness of graduation predictions*. Proceedings of the 12th International Conference on Educational Data Mining, 488–491. [https://learninganalytics.upenn.edu/ryan-baker/EDM2019\\_paper56.pdf](https://learninganalytics.upenn.edu/ryan-baker/EDM2019_paper56.pdf)
- Baker, R.S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Bridgeman, B., Trapani, C., & Attali, Y. (2012). Comparison of human and machine scoring of essays: Differences by gender, ethnicity, and country. *Applied Measurement in Education*, 25(1), 27–40. <https://doi.org/10.1080/08957347.2012.635502>
- Chizhikova, A., Billingham, H., Elizabeth, M., Hossain, S., Kulkarni, A., Guibon, G., & Couceiro, M. (2024). *Factorizing gender bias in automatic speech recognition for mexican Spanish*. <https://univ-catholille.hal.science/LORIA-NLPKD/hal-04607587v2>
- Ferrero, F., & Gewerc Barujel, A. (2019). Algorithmic driven decision-making systems in education: Analyzing bias from the sociocultural perspective. *2019 XIV Latin American Conference on Learning Technologies (LACLO)* (pp. 166–173). <https://doi.org/10.1109/LACLO49268.2019.00038>
- Gardner, J., Brooks, C., & Baker, R. (2019). Evaluating the fairness of predictive student models through slicing analysis. Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 225–234). <https://doi.org/10.1145/3303772.3303791>
- Huesca, E. y Fueyo Hernández, E.M. (2025). Brechas en el uso de la IA en la educación superior En J.L. Cuevas Nava (Coord.). *Inteligencia Artificial en la educación superior [Gaps in the use of AI in higher education]*. In J.L. Cuevas Nava (Coord.). *Artificial Intelligence in higher education* (pp. 23–45). ANUIES. [https://publicacionesdigitales.anuiex.mx/wp-content/uploads/2025/10/Impresion\\_-Inteligencia-artificial-final.pdf](https://publicacionesdigitales.anuiex.mx/wp-content/uploads/2025/10/Impresion_-Inteligencia-artificial-final.pdf)
- Ilbay, E.L. (2024). Ética y regulación de la inteligencia artificial en tecnologías educativas aplicadas a contextos escolares latinoamericanos (Ethics and Regulation of Artificial Intelligence in Educational Technologies Applied to Latin American School Contexts). *Atlas Research Journal*, 2(1). <https://doi.org/10.65305/arj.v2n1.2024.18>
- Ibarra-Vazquez, G., Ramirez-Montoya, M. S., & Terashina, H. (2024). Gender prediction based on University students' complex thinking competency: An analysis from machine learning approaches. *Education and Information Technologies*, 29, 2721–2739. <https://doi.org/10.1007/s10639-023-11831-4>
- Kasun, G.S., Liao, Y.-C., Marguleux, L.E., & Woodall, M. (2024). Unexpected outcomes from an AI education course among education faculty: Toward making AI accessible with marginalized youth in urban Mexico. *Frontiers in Education*, 9, 1368604. <https://doi.org/10.3389/feduc.2024.1368604>
- Loukina, A., & Buzick, H. (2017). Use of automated scoring in spoken language assessments for test takers with speech impairments. *ETS Research Report Series*, 2017(1), 1–10. <https://doi.org/10.1002/ets2.12170>
- Lozano-Mulet, P. (2025). Sesgos algorítmicos desde una perspectiva interseccional. La necesidad de una alfabetización digital crítica en educación (Algorithmic Biases from an Intersectional Perspective: The Need for Critical Digital Literacy in Education). *Revista Izquierdas*, 54, 1–31. <http://dx.doi.org/10.4067/s0718-50492025000100240>
- Medina Vidal, A., Alonso-Galicia, P.E., González Mendoza, M., & Ramírez-Montoya, M.S. (2025). Financial inclusion of vulnerable sectors with a gender perspective: Risk analysis model with artificial intelligence based on complex thinking. *Journal of Innovation and Entrepreneurship*, 14(4). <https://doi.org/10.1186/s13731-025-00463-2>
- Regmi, K.D. (2023). Decolonising meritocratic higher education: Key challenges and directions for change. *Globalisation, Societies and Education*, 766–783. <https://doi.org/10.1080/14767724.2023.2210516>
- Salcido Ledezma, M.A. (2024). Reflexiones sobre la regulación en materia de inteligencia artificial: brechas entre Europa y México (Reflections on Artificial Intelligence Regulation: Gaps between Europe and Mexico). *Revista Iberoamericana*, 1(1), 31–40. <https://eritronio.org/index.php/revista/article/view/12>
- Spitz, S., Züger, T., & Mosene, K. (2023). *A Gender Perspective on Artificial Intelligence (AI) and the Tech-Sector. Position Paper for the multi-stakeholder dialogue in Mexico*. Zenodo. <https://doi.org/10.5281/zenodo.10159107>
- Suresh, H., & Gutttag, J.V. (2020). *A framework for understanding unintended consequences of machine learning*. arXiv preprint. <https://arxiv.org/abs/1901.10002>
- Torres Zúñiga, V. (2024). Sesgos en la representación de profesionales de la ciencia forense en imágenes generadas por inteligencia artificial (Biases in the Representation of Forensic Science Professionals in Artificial Intelligence-Generated Images). *TIES*, 12. <https://doi.org/10.22201/dgtic.26832968e.2025.12.57>
- Uc Castillo, J.L., Marín Celestino, A.E., Martínez Cruz, D.A., Tuxpan Vargas, J., Ramos Leal, J.A., & Morán Ramírez, J. (2025). A systematic review of Machine Learning and Deep Learning approaches in Mexico: challenges and opportunities. *Frontiers in Artificial Intelligence* 7(1479855). <https://doi.org/10.3389/frai.2024.1479855>
- Vargas-Solar, G. (2022). Intersectional study of the gender gap in STEM through the identification of missing datasets about women: A multisided problem. *Applied Sciences*, 12, 5813. <https://doi.org/10.3390/app12125813>