




# RECOMENDACIONES DE POLÍTICA PÚBLICA

Prevenir y mitigar el sesgo de género en los sistemas de alerta temprana basados en IA para la educación





Este documento de recomendaciones de política pública, elaborado por Itad, PIT Policy Lab, Women in Digital Transformation y Athena Infonomics, fue desarrollado en el marco de la iniciativa de Inteligencia Artificial Equitativa (Equitable AI) y es posible gracias al apoyo del pueblo estadounidense a través de la Agencia de los Estados Unidos para el Desarrollo Internacional (USAID). Los contenidos son responsabilidad de las organizaciones nombradas y no necesariamente reflejan los puntos de vista de USAID o del Gobierno de los Estados Unidos.

Fecha de publicación: Julio de 2023.

# Contenidos

Panorama del caso en Guanajuato, México	1
Introducción	2
Recomendaciones	3
Fase 1: Autoevaluación y reflexión	3
Fase 2: Preparando el camino	4
Fase 3: Involucramiento y transparencia	5
Fase 4: Fortalecimiento de los instrumentos existentes	6
Fase 5: Sensibilidad más amplia	7

## Panorama del caso en Guanajuato, México

Bajo la administración actual (2018-2024), la **Secretaría de Educación de Guanajuato (SEG)** ha estado trabajando para garantizar información válida, confiable y actualizada sobre el estado del sistema educativo local para informar la toma de decisiones. Por lo tanto, han estado desarrollando y armonizando bases de datos con información sobre estudiantes y rendimiento escolar general bajo la iniciativa **Trayectorias Educativas**. Además, para identificar a estudiantes en riesgo de deserción escolar, la Secretaría está trabajando con el Banco Mundial para crear un sistema de alerta temprana basado en Inteligencia Artificial (IA). El **Sistema de Acción Temprana para la Permanencia Escolar (SATPE)** tiene como objetivo brindar apoyo a estudiantes en riesgo de deserción para mejorar los índices de retención y egreso a través de acciones preventivas focalizadas.

Trayectorias Educativas reúne conjuntos de datos operativos internos y externos, con el objetivo de ser un instrumento para el seguimiento de los estudiantes, así como para orientar acciones personalizadas de la Secretaría. Este marco de datos consolidado permite la identificación de un conjunto de variables que tienen un impacto en la retención de estudiantes y en la calidad de la educación, lo que permite intervenciones basadas en evidencia y efectivas. También ha permitido desarrollar análisis predictivo sobre las trayectorias individuales de estudiantes en Guanajuato, para un seguimiento cercano junto con escuelas y padres de familia.

La iniciativa Trayectorias Educativas sirve de base para el SATPE. El desarrollo de un sistema de alerta temprana impulsado por IA para identificar a los estudiantes en riesgo de abandonar la escuela surgió de una colaboración anterior con el Banco Mundial, para aprovechar los datos existentes de la iniciativa de Trayectorias Educativas dentro de la visión del “Pacto por la Educación” del Estado.



## Introducción

Tanto Trayectorias Educativas como el Sistema de Acción Temprana para la Permanencia Escolar (SATPE) son iniciativas clave para informar la política educativa de Guanajuato. A principios de 2023, se identificó la primera lista de estudiantes en riesgo de deserción a través de análisis de datos y actualmente el equipo de la SEG está evaluando marcos de acción de intervenciones piloto. A nivel técnico, se están ajustando los últimos detalles para el despliegue del sistema de información, así como los mecanismos para alimentarlo con más registros en el marco de la iniciativa Trayectorias Educativas. Como se demostró a través del caso de uso de Inteligencia Artificial Equitativa (Equitable AI) tanto el diseño e implementación del SATPE, como las políticas públicas que éste informa, han presentado avances e iteraciones positivas. Sin embargo, aún existen desafíos relevantes a atender:

### ROL DE LA COMUNICACIÓN



hacia las personas impactadas por el sistema y su implementación (estudiantes, padres/madres de familia, directiva escolar) sobre el uso de la IA para la toma de decisiones en el ámbito educativo, previo a la implantación y despliegue del SATPE.

### CONCEPCIÓN E INTERNALIZACIÓN DE RIESGOS



asociados al uso de la IA, más allá de un posible mal uso de los datos personales (por ejemplo, sesgos en los datos y el potencial de estos sistemas para reproducirlos y expandirlos).

### CONSULTA PÚBLICA



hacia el diseño y despliegue de una política pública impulsada por la IA, incluyendo mecanismos de consulta y/o retroalimentación con distintos actores relevantes.

Estos retos son comunes a los gobiernos que comienzan a implementar proyectos de IA y se encuentran en proceso de desarrollo de marcos éticos y regulatorios internos para tales iniciativas. Además, reflejan muchos de los desafíos compartidos por los gobiernos subnacionales en México, la región de América Latina y el Sur Global.

Con la intención de promover prácticas que apunten hacia un uso más equitativo y responsable de la IA, este documento presenta una serie de recomendaciones de políticas accionables para mitigar sesgos e incorporar la perspectiva de género. Las recomendaciones están dirigidas a la Secretaría de Educación de Guanajuato para proyectos como el SATPE y Trayectorias Educativas, y como guía hacia cualquier proyecto futuro que involucre el uso de datos e IA. Sin embargo, estas recomendaciones son lo suficientemente generales como para ser adoptadas por otros gobiernos locales que también están comenzando a explorar políticas que usan datos e IA, incluso en sectores distintos al educativo.

# Recomendaciones

## Fase 1: Autoevaluación y reflexión

A través de la iniciativa de Inteligencia Artificial Equitativa (Equitable AI), surgieron numerosas preguntas desde la SEG en torno a cómo capturar aspectos que podrían tener implicaciones de género en los datos más allá de la desagregación por sexo, o si hay otras realidades no capturadas en la información. El alcance de las intervenciones planificadas adquiere nuevas implicaciones cuando se incluyen los Derechos Humanos, la ética y otros criterios de diseño. Esta discusión no es menor y permite identificar áreas de oportunidad, así como desafíos aún por resolver.

Se sugiere lo siguiente:



Documentar los procesos y aprendizajes del diseño e implementación de las iniciativas actuales, particularmente con relación a los riesgos que plantea la implementación de un sistema basado en IA, identificando las preguntas relevantes que se han planteado a lo largo del desarrollo de la iniciativa y cómo podrían resolverse o responderse en diferentes horizontes temporales.



Sopesar fortalezas y debilidades en un análisis de capacidades institucionales, especialmente con relación al uso ético y responsable de la tecnología, incluyendo los diferentes equipos y áreas involucradas en proyectos con datos e IA. Identificar necesidades específicas en temas como gobernanza de datos, ciberseguridad, género y ética de la IA.



Desarrollar una metodología de análisis de riesgos, o adaptar alguna existente [1], que incluya algunos de los aspectos abordados durante los talleres que tuvieron lugar en el marco de la iniciativa (Derechos Humanos, género, ética de la IA, etc.), involucrando a actores clave adicionales (estudiantes, padres/madres, sociedad civil, directiva escolar).




Desde la etapa de diseño de las intervenciones, los equipos deben preguntarse qué podría salir mal, mapeando a posibles actores involucrados para tener en cuenta sus preocupaciones. La **Lista de verificación para la implementación de IA** desarrollada y publicada en el marco de este proyecto puede ser un insumo útil para la identificación de riesgos y áreas de oportunidad; sin embargo, no reemplaza la consulta con las personas impactadas por la intervención.

[1] Por ejemplo, el [Marco de Gestión de Riesgos de IA](#) del Instituto Nacional de Estándares y Tecnología (NIST) de los Estados Unidos, o los módulos relacionados a la guía práctica [Gestión de proyectos de aprendizaje automático en el desarrollo internacional](#) de USAID.


## Fase 2: Preparando el camino

Para estandarizar las mejores prácticas en el trabajo con datos e IA, el área/organización del sector educativo involucrado en el desarrollo del caso de uso debe establecer criterios de decisión alineados con la ética y los Derechos Humanos. Con base en las preguntas planteadas y sistematizadas en el paso anterior, esta etapa sugiere establecer lineamientos claros para orientar la acción. Además, para abordar las preocupaciones sobre el consumo y la transferencia de información privada dentro y fuera del Secretaría, así como para mitigar los posibles riesgos de uso indebido, también se deben tener en cuenta los criterios de ciberseguridad y protección de datos personales.

Se sugiere lo siguiente:



Fijar o adoptar principios internos que orienten la actuación de la Secretaría de Educación en consonancia con los principios con los que el país ya se ha comprometido. Crear un Grupo de Trabajo que se encargue de sistematizar y diseñar los criterios éticos para las intervenciones con datos e IA dentro del área/organización. La **Guía de Inteligencia Artificial (IA) Ética** desarrollada y publicada en el marco de este proyecto puede ser un insumo útil en este sentido, ya que proporciona ejemplos de tales prácticas.



Crear un protocolo de gobierno de datos que permita establecer reglas y criterios sobre cómo se almacena, consume y utiliza la información dentro de la Secretaría. Dicho protocolo deberá establecer distintos niveles de seguridad de los datos y reglas para su transferencia dentro y fuera de la organización. También debe contener criterios para la calidad de los datos, asegurando que exista información completa, precisa y representativa de la población objetivo del caso de uso. Este debe considerar criterios de interoperabilidad entre otras secretarías del Gobierno de Guanajuato.



### Fase 3: Involucramiento y transparencia

Las partes interesadas, tales como padres/madres de familia de estudiantes, líderes comunitarios, autoridades educativas y demás personal operativo, deben ser incluidos desde las fases de diseño e implementación dentro del Ciclo de Tecnología Responsable. Su rol va más allá de una discusión inicial de los posibles riesgos, para ser considerados en una etapa piloto y poder retroalimentar a lo largo del ciclo de vida del proyecto. Además, el derecho a auditar los resultados del modelo debe ser considerado en el marco de las atribuciones de cada parte involucrada, así como un mecanismo de explicabilidad de los resultados.

Se sugiere lo siguiente:



Crear mecanismos de consulta abiertos a la sociedad civil, grupos de interés específicos y población impactada (incluyendo tanto grandes empresas como pymes, ciudadanía y profesionistas del sector de educación). Desde la etapa de diseño, deben asegurarse de que las partes involucradas entiendan cómo les impacta una decisión y en qué medida se basa en las predicciones de un sistema de IA. Esta discusión debe fomentarse siempre que haya nuevos usos de los datos con implicaciones potenciales para el bienestar de ciertos grupos o de la población en general.




Diseñar un flujo de explicabilidad, que permita a las personas usuarias finales y grupos impactados entender el por qué de la toma de decisiones de IA, desde la identificación de estudiantes en riesgo que requieren apoyo, hasta la implementación de intervenciones. Delinear claramente en qué medida el modelo sugiere cada componente de la acción específica y cuál es el razonamiento para la intervención. Esto es importante para identificar de dónde parte un determinado curso de acción, cuándo es o no una recomendación automatizada y quién ha implementado una medida de control o verificación sobre tal decisión.




Diseñar una estrategia de comunicación clara para las partes impactadas por el SATPE (estudiantes, docentes, directiva escolar, padres/madres de familia, líderes comunitarios, entre otras partes interesadas). Esto debería priorizar la comprensión de la estrategia en el contexto de la audiencia objetivo; por ejemplo, niños y niñas, estudiantes de diferentes niveles socioeconómicos, población rural y urbana, etc. En el diseño de la estrategia de comunicación, la Secretaría debe cuestionar qué discursos pueden ser estigmatizantes (por ejemplo, “riesgo de deserción escolar” vs. “riesgo de interrumpir la trayectoria educativa”) y cómo complementa la estrategia de inclusión de las intervenciones.





Ofrecer campañas de sensibilización e incentivos para el uso ético de datos e IA para pymes, especialistas en ciencia de datos y empresas emergentes involucradas en la prestación de servicios; así como capacitación para considerar dichos aspectos en las adquisiciones, tanto para proveedores potenciales/actuales como para personal de gobierno.




Otras acciones incluyen: Estos proyectos pueden beneficiarse de una revisión e iteración continua de sus procesos de diseño e implementación. La capacidad de participar en debates nacionales e internacionales, compartir aprendizajes y llevar a cabo sesiones de intercambio de conocimientos entre pares a nivel regional, nacional, estatal y/o municipal puede garantizar un enfoque más ágil de los programas, que pueden cambiar y adaptarse a medida que se dispone de nueva evidencia.


#### **Fase 4: Fortalecimiento de los instrumentos existentes**

Existen algunos precedentes que guían la actuación de las prácticas de la Secretaría con relación a aspectos específicos, como la ética en la función pública o la protección de datos personales (por ejemplo, en México existe la Ley Federal de Datos Personales en Posesión de los Particulares). Ser capaces de coordinar los esfuerzos existentes e incorporar una perspectiva ética de la IA y la protección de datos personales como criterio permite construir una base sólida para políticas éticas, responsables y confiables impulsadas por IA.


Se sugiere lo siguiente:



Incluir lineamientos básicos para la protección de datos personales, abordaje ético de IA y priorización de Derechos Humanos en los códigos de conducta existentes para personas funcionarias públicas.



Consolidar un Comité de Ética para establecer y monitorear el cumplimiento de los mecanismos para asegurar el despliegue de una IA ética y responsable en todas las intervenciones realizadas por la Secretaría. Este comité puede estar vinculado a políticas o equipos de innovación internos (por ejemplo, el Laboratorio de Innovación Educativa de la SEG en conjunto con UNESCO).



Promover la alfabetización transversal en IA dentro de la Secretaría y otras agencias del Gobierno del Estado para generar capacidades específicas en los tomadores de decisiones que interactuarán con los sistemas de IA; por ejemplo, personas interactuando directamente con los resultados del SATPE para asegurarse de que sus predicciones se interpreten correctamente.

## Fase 5: Sensibilidad más amplia

El SATPE es un primer ejercicio que está creando conciencia sobre las implicaciones éticas y los riesgos asociados con el uso de la IA para la toma de decisiones en políticas públicas educativas. Sólo se puede aumentar la precisión del diagnóstico y la eficacia de las intervenciones midiendo y evaluando los resultados. Contemplar mecanismos de verificación más allá de los datos y el algoritmo es fundamental para asegurar el componente humano como máximo responsable de la toma de decisiones.

Se sugiere lo siguiente:



Contrastar las alertas de riesgo de abandono con indicadores alternativos e información cualitativa sobre el impacto del proyecto. Un ejercicio de análisis de casos sobre una muestra aleatoria representativa puede arrojar luz sobre qué otros elementos del contexto complementan o contradicen las predicciones del SATPE.



Considerar las métricas de equidad de datos, como las proporcionadas por el Kit de herramientas AI Fairness 360, para identificar cuándo pueden existir sesgos en los datos y corregirlos, antes de la clasificación/predicción.[2] Revisar estas métricas con frecuencia durante el uso de los sistemas de IA, ya que pueden cambiar con el tiempo a medida que se agregan nuevos datos al sistema.



Desarrollar una amplia sensibilidad para calibrar los procesos de toma de decisiones habilitados por IA. También dependerá de qué tan sensibles sean los resultados y las intervenciones a los comentarios de las partes interesadas. Por ejemplo, si incorporar o no mecanismos para incluir información adicional de estudiantes, directiva escolar o padres/madres identificadas.



Generar métricas adecuadas. Si bien la medición y evaluación de resultados es parte del ciclo de vida del proyecto en el Secretaría, es prioritario el uso de esquemas adecuados a la realidad del fenómeno. La efectividad de los resultados deberá ponderarse con criterios de equidad e inclusión, y deberán poder separar los resultados de la predicción (precisión) y las intervenciones (efectividad) en el análisis.

[2] Otras herramientas y recursos disponibles son: People + AI Research (PAIR) de Google, Aequitas kit de herramientas de auditoría y sesgo de Carnegie Mellon University, Fairlearn, Datasheets for Datasets de Microsoft, etiquetas de datos de Data Nutrition Project, y Model Card de Tensorflow.

# RECOMENDACIONES DE POLÍTICA PÚBLICA

Prevenir y mitigar el sesgo de género en los sistemas de alerta temprana basados en IA para la educación

