

# HERRAMIENTAS DE ALGORITMOS ÉTICOS

El GobLab UAI, el laboratorio de innovación pública de la Escuela de Gobierno de la Universidad Adolfo Ibáñez ha trabajado desde 2020 en promover el uso responsable y transparente de algoritmos, inteligencia artificial y sistemas automatizados o semiautomatizados en el sector público chileno, buscando con ello apoyar a las instituciones públicas innovadoras.

En el sitio web Algoritmos Públicos están disponibles tres herramientas desarrolladas por el GobLab UAI que son gratuitas, estandarizadas, de código abierto, en español y pensadas para la realidad chilena. Permiten evaluar, prevenir y mitigar riesgos éticos en los modelos de ciencia datos e inteligencia artificial, es decir, apoyan la incorporación de estándares éticos y responsables durante todo el ciclo de vida de un proyecto de IA y ciencia de datos.

Socios:



Patrocina:

Financia:



## Herramientas

### Evaluación de Impacto Algorítmico

Identifica riesgos éticos desde una mirada técnica, legal y de gestión, y recibe recomendaciones de mitigación personalizadas al proyecto de IA.

1

### Medición de Sesgos y Equidad Estadística

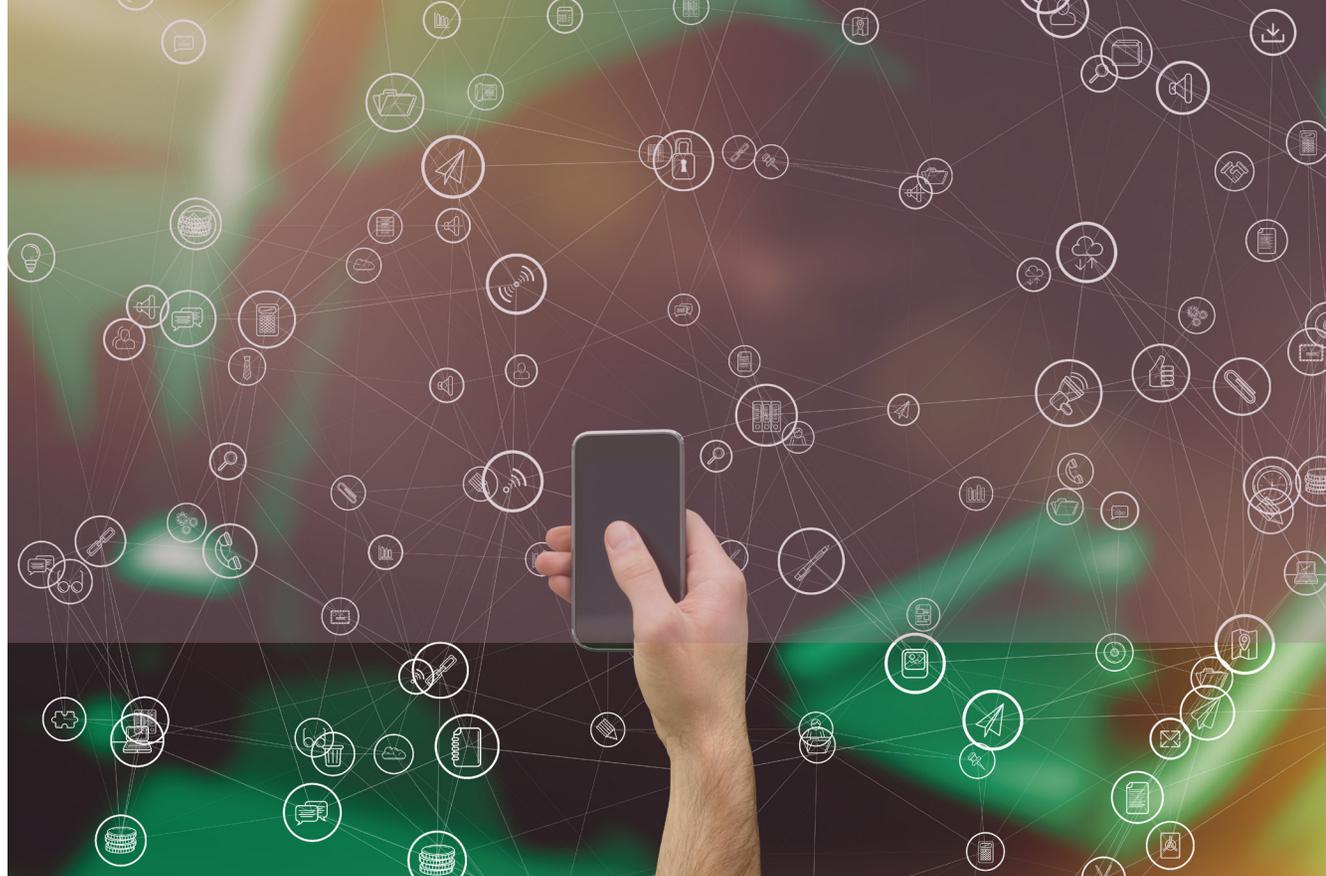
Mide e identifica sesgos en las predicciones del modelo de IA, para crear sistemas más justos y prevenir la discriminación.

2

### Ficha de Transparencia del Modelo

Documenta y comunica de manera clara la lógica de decisión detrás del sistema de IA y explica aspectos claves para promover la transparencia.

3



### ¿Para qué sirven estas herramientas?

- Para una implementación ética, responsable y segura de proyectos de IA y ciencia de datos en servicios públicos.
- Para dar cumplimiento a los objetivos de la Política Nacional de Inteligencia Artificial.
- Para aplicar la circular Lineamientos para el uso de herramientas de inteligencia artificial en el sector público, del Ministerio de Ciencia y la Secretaría de Gobierno Digital.

# 1 Evaluación de impacto algorítmico



- Busca que las decisiones automatizadas sean comprensibles y auditables por el público y las autoridades.
- Ayuda a identificar a los **responsables** de la implementación y operación de sistemas de IA, asegurando un uso ético y transparente.
- Esto permite fomentar la **confianza pública y evitar riesgos** como el sesgo o la discriminación, infracciones de protección de datos y opacidad.

## ¿En qué consiste la herramienta?

- Realiza un análisis detallado para **identificar riesgos éticos, sociales y técnicos y sus posibles daños**, para tomar medidas prácticas para evitarlos o mitigarlos.
- Recomienda pasos a seguir, guías y herramientas para tomar acciones que permitan mitigar los riesgos éticos del proyecto.

## Resultado o salida al aplicar la herramienta

**Recomendaciones personalizadas para mitigar riesgos éticos del proyecto:**

- Proporcionalidad
- Normativa
- Protección de datos
- Licencia social
- Gobernanza
- Rendición de cuentas
- Transparencia
- Ciberseguridad
- No discriminación - equidad

## 2 Medición de sesgos y equidad estadística



- El **error del sistema** es la diferencia entre el valor predicho y el valor real de la variable que se está estimando. Si este es **sistemático** en una dirección o subconjunto específico de datos, se llama **sesgo**.
- Por otra parte, la equidad es la cualidad de **tratar a todas las personas de manera justa e imparcial**, sin discriminación ni favoritismo, y asegurando que tengan acceso a las mismas oportunidades y recursos.

### ¿En qué consiste la herramienta?

**Mide los sesgos de las variables protegidas**, en las predicciones de los modelos de ciencias de datos de clasificación.

Fue diseñada para usarse durante la fase del proyecto de "evaluación del modelo".

Las variables protegidas son características personales, como el género, edad, situación socioeconómica o discapacidad, que están legal o éticamente resguardadas contra la discriminación.

### Resultado o salida al aplicar la herramienta

Identifica/Analiza sesgos en las variables protegidas del modelo de IA, usando nueve métricas disponibles:

- Tasa de verdaderos positivos
  - Falsos negativos/Tamaño del grupo
  - Tasa de falsa omisión
  - Tasa de falsos negativos
  - Paridad de selección igualitaria
  - Paridad demográfica
  - Falsos positivos/Tamaño del grupo
  - Tasa de falsos descubrimientos
  - Tasa de falsos positivos
- La legislación chilena actual, por medio de la Ley N° 20.609, establece la no discriminación arbitraria para variables protegidas como sexo, nacionalidad, situación socioeconómica, enfermedad o discapacidad, entre otras.

# 3 Ficha de transparencia algorítmica

- Transparenta información clave sobre un modelo de ciencia de datos o IA, lo que fomenta la **confianza**, la rendición de cuentas y la responsabilidad en los procesos o la toma de decisiones.
- Esto permite que la ciudadanía acceda a información valiosa sobre cómo se toman las decisiones que la afectan lo que mejora la **calidad de la democracia**.
- Esta herramienta permite cumplir con las Recomendaciones de Transparencia Algorítmica del Consejo para la Transparencia.

## ¿En qué consiste la herramienta?

- Resume la arquitectura del modelo, los datos de entrenamiento, las métricas de rendimiento, **posibles riesgos éticos** y otros aspectos cruciales para comprender y evaluar su funcionamiento.
- Aborda **44 preguntas** de los distintos elementos, características y consideraciones del modelo.

## Resultado o salida al aplicar la herramienta

### Documento claro y amigable con el usuario que transparenta:

- Detalles del modelo
- Visión general del modelo
- Clasificación
- Métricas de rendimiento
- Datos de entrenamiento
- Datos de evaluación
- Consideraciones éticas
- Advertencias y recomendaciones
- Reclamación



## Las herramientas de Algoritmos Éticos apoyan la incorporación de estándares éticos y responsables en el ciclo de vida de un proyecto de IA y ciencia de datos



Para más conocer sobre resultado o salida al aplicar las herramienta, acceso a ellas y más información general, ingresa a [www.algoritmospublicos.cl](http://www.algoritmospublicos.cl)

goblab@uai.cl



# HERRAMIENTAS ALGORITMOS ÉTICOS

Socios:



Patrocina:

Financia: