



ALGORITMOS PÚBLICOS

UNIVERSIDAD ADOLFO IBÁÑEZ



Convocatoria pilotaje
Herramientas
Algoritmos Éticos

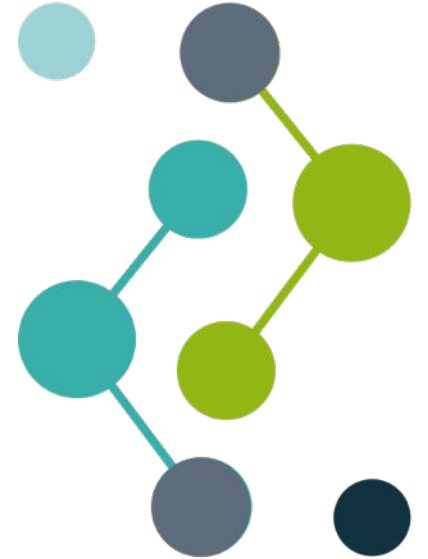


Presentación

La plataforma [Algoritmos Públicos](#) **centraliza iniciativas del GobLab UAI**, el laboratorio de innovación pública de la Escuela de Gobierno de la Universidad Adolfo Ibáñez.

Su objetivo es 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.

La plataforma sumó un nuevo espacio: **Herramientas de Algoritmos Éticos**.



Presentación

Ya están disponibles en su versión Beta **tres nuevas herramientas gratuitas, estandarizadas, de código abierto, en español y pensadas para la realidad chilena.**

Fueron desarrolladas por el equipo del GobLab UAI, gracias a fondos del **concurso ANID IDeA I+D 2023** (proyecto ID23I10357), en la línea temática “Transformaciones sociales y culturales para un nuevo modelo de desarrollo”.



		
Publicada	Publicada	Publicada
Ficha de Transparencia Algorítmica	Medición de Sesgos y Equidad	Evaluación de Impacto Algorítmico

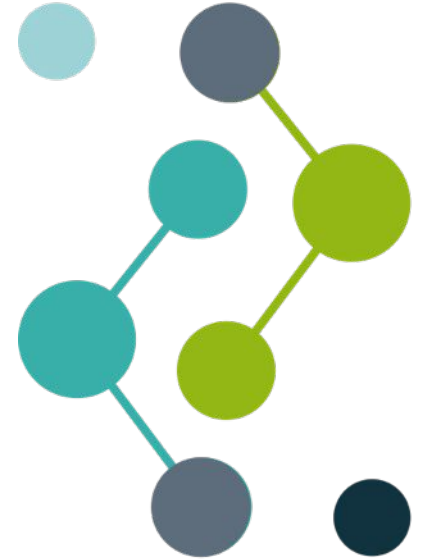
¿Para qué sirven las herramientas?

Para **aplicar los principios de la IA responsable** a través de un proceso único en el mundo.

Para una implementación ética y segura de estos proyectos en su Servicio.

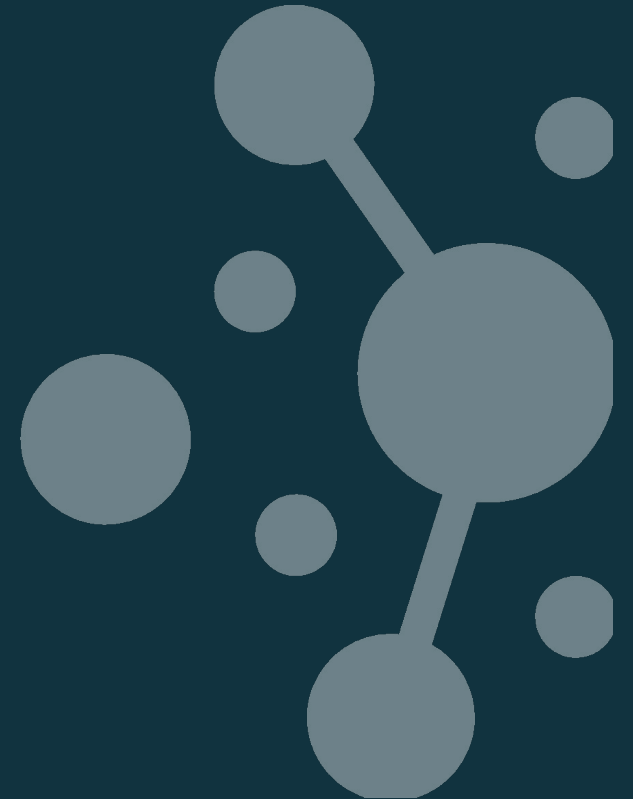
Para hacerse parte del [Plan de Acción](#) de la **Política Nacional de Inteligencia Artificial**.

Para aplicar la circular [Lineamientos para el uso de herramientas de inteligencia artificial en el sector público](#), enviada por el Ministerio de Ciencia y la Secretaría de Gobierno Digital.



Ficha de transparencia del modelo

Para modelos de ciencia de datos o de IA de cualquier naturaleza.



Transparencia

La transparencia fomenta la **confianza**, la rendición de cuentas y la responsabilidad en los procesos o la toma de decisiones.

Por el contrario, la falta de ella supone un problema profundo para la **calidad de la democracia** y la posibilidad de que la ciudadanía acceda a información valiosa sobre cómo se toman las decisiones que la afectan.



¿En qué consiste la herramienta?

- Documento detallado que ofrece información clave sobre un **modelo de ciencia de datos o IA**.
- 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.
- Fue creada basándose en Model Cards for Model Reporting (Mitchell, 2019), AI Factsheets de IBM (2019) e incorporando las preguntas del piloto del borrador de la Instrucción General de Transparencia Algorítmica*.
- Aborda **44 preguntas** de los distintos elementos, características y consideraciones del modelo.



*Disponible en el portal de Transparencia Activa.

Salida de la herramienta

Resultados sobre:

1. Detalles del modelo.
2. Visión general del modelo.
3. Clasificación.
4. Métricas de rendimiento.
5. Datos de entrenamiento.
6. Datos de evaluación.
7. Consideraciones éticas.
8. Advertencias y recomendaciones.
9. Reclamación.



Medición de sesgos y equidad estadística

Para sistemas predictivos de clasificación.



Sesgos y equidad

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.



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.

¿En qué consiste la herramienta?

- **Mide los sesgos de las variables**, 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”.
- El flujo de la herramienta está basado en Aequitas (2019, Ghani) y fue desarrollada en python para científicos de datos.



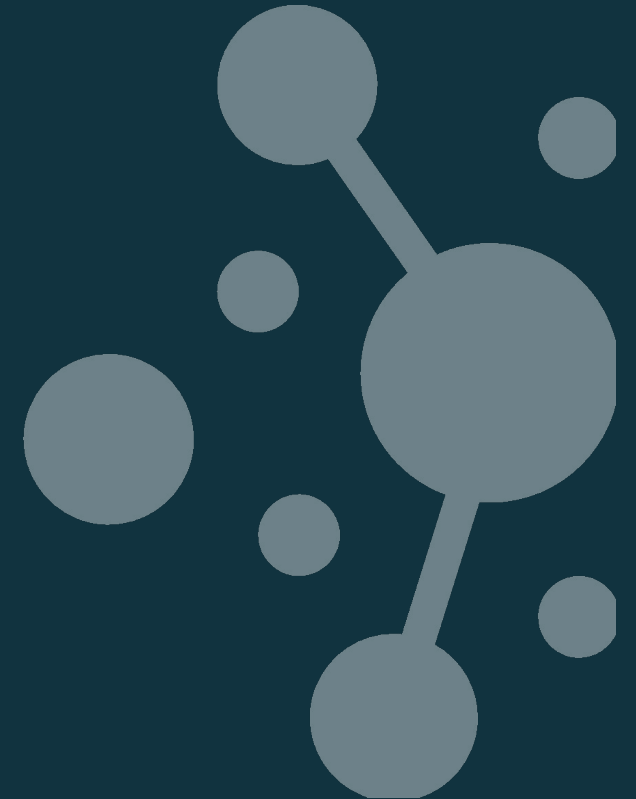
Salida de la herramienta

- Separa la variable protegida seleccionada por “categorías” y entrega **cuatro métricas disponibles**:
 1. Tasa de falsos positivos.
 2. Tasa de falsos negativos.
 3. Tasa de falsa omisión.
 4. Tasa de falso descubrimiento.
- Qué métricas se deben utilizar dependerá del tipo de proyecto (punitivo o asistencial).



Evaluación de Impacto Algorítmico

Para modelos de ciencia de datos o de IA de cualquier naturaleza.



Rendición de cuentas y responsabilidad

Estos principios buscan garantizar **que las decisiones automatizadas sean comprensibles** y auditables por el público y las autoridades.

Ayudan a identificar a los **responsables** de la implementación y operación de sistemas de IA, asegurando un uso ético y transparente.

Finalmente el objetivo es fomentar la **confianza pública** y **evitar riesgos** como el sesgo o la discriminación, infracciones de protección de datos, opacidad.



¿En qué consiste la herramienta?

- Es una **metodología que se aplica antes** del desarrollo o implementación de un sistema que utilizará algoritmos, como por ejemplo una IA.
- Realiza un análisis detallado de un sistema algorítmico o inteligencia artificial para **identificar riesgos éticos, sociales y técnicos y sus posibles daños**, para tomar medidas prácticas para evitarlos o mitigarlos.
- 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.



Salida de la herramienta

Resultados basados en dimensiones:

1. Proporcionalidad
2. Normativa
3. Protección de datos
4. Licencia social
5. Gobernanza
6. Rendición de cuentas
7. Transparencia
8. Ciberseguridad
9. No discriminación-Equidad



Convocatoria a pilotaje

Para probarlas y aplicarlas con asesoría de nuestro equipo experto.



¿Quiénes pueden participar?

Buscamos **proyectos de ciencia de datos o inteligencia artificial en el sector público**, que estén en alguna de las siguientes fases:

- Desarrollo.
- Implementación.
- Producción.

Pueden ser utilizadas en proyectos desarrollados por un equipo del organismo público o por un proveedor externo.



ETAPA 1

Sesión informativa online

Se realizarán mensualmente para interesados.

Público objetivo: jefaturas intermedias y profesionales que están implementando IA y ciencia de datos en el Estado.



ETAPA 2

Sesiones de capacitación online

Se realizarán mensualmente, sobre cada una de las herramientas, para quienes se inscriban en el pilotaje.

Público objetivo: integrantes del equipo que está implementando IA y ciencia de datos (área de negocio, técnicos, jefatura y empresa proveedora si hubiera).



ETAPA 3

Ejecución del piloto

3.1 Aplicación de prueba de la herramienta.

3.2 Servicio envía retroalimentación a la UAI, por plataforma y encuesta.

La ejecución puede ser de dos tipos:

Piloto tipo A - Servicio público ejecuta la herramienta.

Aplica para ambas herramientas.

Piloto tipo B - UAI ejecuta la herramienta.

Solo para la Medición de sesgos y equidad.



ETAPA 4

Retroalimentación UAI a resultados pilotos

Una vez recibidos los resultados de la herramienta y la retroalimentación completa del Servicio, la UAI entregará retroalimentación por escrito y a través de una reunión.



ETAPA 5

Casos de éxito

Cuando la UAI considere que un piloto podría ser un caso de éxito, solicitará el permiso del servicio para **difundirlo** y elaborar documentos de divulgación y otros materiales comunicacionales para **publicar en la plataforma Algoritmos Públicos** y otros medios.



¡Hazte parte del pilotaje!

Inscripciones en:

<https://algoritmospublicos.cl/herramientas>

Contacto:

goblab@uai.cl

