Desafíos éticos para la empresa frente a la cuarta revolución industrial

María Paz Hermosilla
Directora GobLab UAI

Tipología de riesgos éticos

1. Privacidad
2. Discriminación
3. Opacidad

Privacidad y Seguridad de la Información

How to Protect Yourself After the Equifax Breach

By **RON LIEBER** UPDATED October 16, 2017

The credit reporting agency said the information of more than 145 million Americans had been compromised.

https://www.nytimes.com/interactive/2017/your-money/equifax-data-breach-credit.html

facebook

Correo electrónico o teléfono | Contraseña | Entrar

¿Has olvidado los datos de la cuenta?

Facebook te ayuda a comunicarte y compartir con las personas que forman parte de tu vida.

Registrarte

Es rápido y fácil.

Nombre | Apellidos

Número de móvil o correo electrónico

Contraseña nueva

Fecha de nacimiento

1 | oct | 1994

Género

Mujer | Hombre | Personalizado

Al hacer clic en Registrarte, aceptas las Condiciones, la Política de datos y la Política de cookies. Es posible que te enviemos notificaciones por SMS que podrás desactivar cuando quieras.

Registrarte

Crea una página para un personaje público, un grupo de música o un negocio.

Yale Journal of Law and Technology

Volume 16
Issue 1 *Yale Journal of Law and Technology*

Article 2

2014

A Theory of Creepy: Technology, Privacy, and Shifting Social Norms

Omer Tene
Rishon Le Zion, Israel

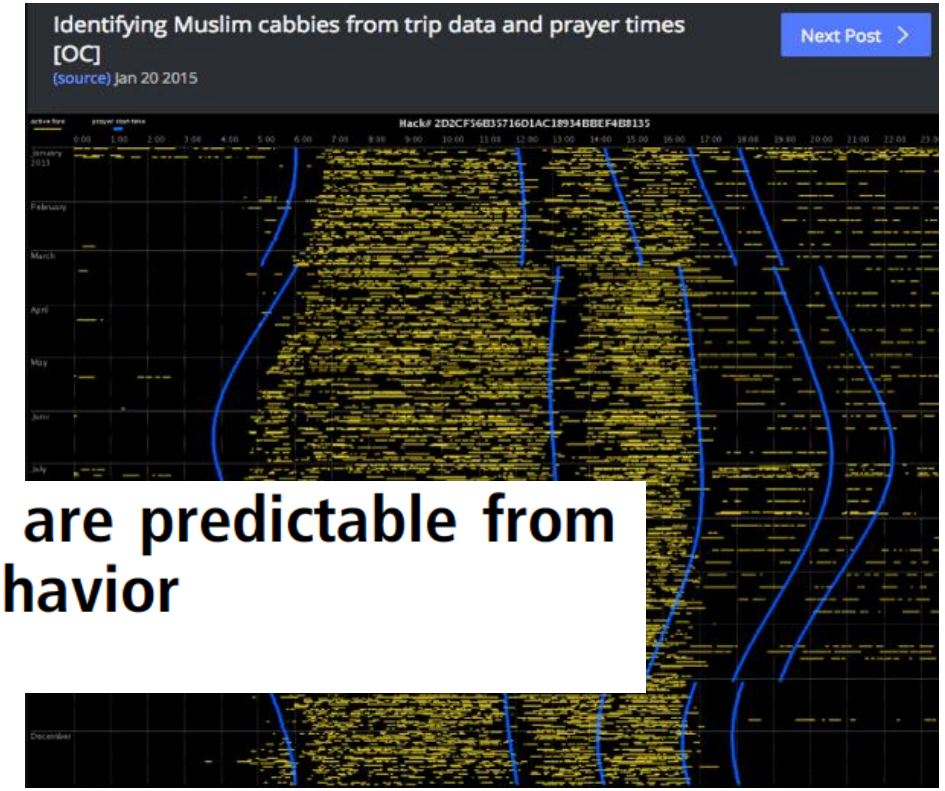
Jules Polonetsky


Girls Around Me
Girls Around Me scans your surroundings and helps you find out where girls or guys are hanging out. You can also see the ratio of girls to guys in different places around you.
Millions of people share their locations
Available on the App Store

PRIVACIDAD

Nuevas técnicas de análisis de datos + Nuevas fuentes de datos sobre comportamiento = Inferencias sobre información privada

Privacidad



Private traits and attributes are predictable from digital records of human behavior

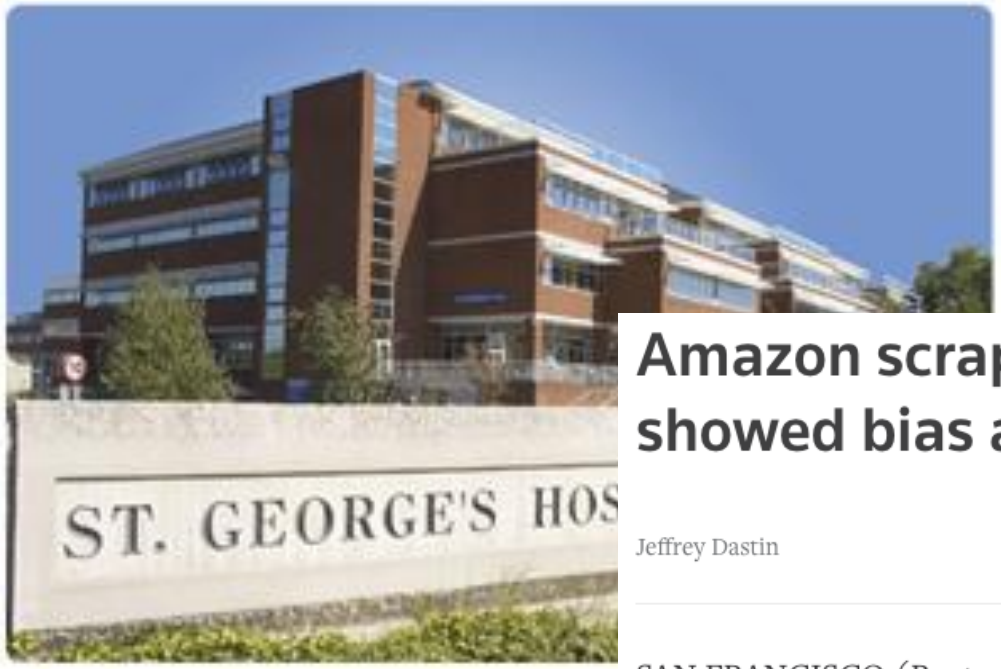
Michal Kosinski[a,1], David Stillwell[a], and Thore Graepel[b]

Discriminación

La discriminación es un trato diferente y perjudicial que se da a una persona debido a categorizaciones arbitrarias o irrelevantes.

Discriminación

Sesgos de decisiones reflejados en datos



Amazon scraps secret AI recruiting tool that showed bias against women

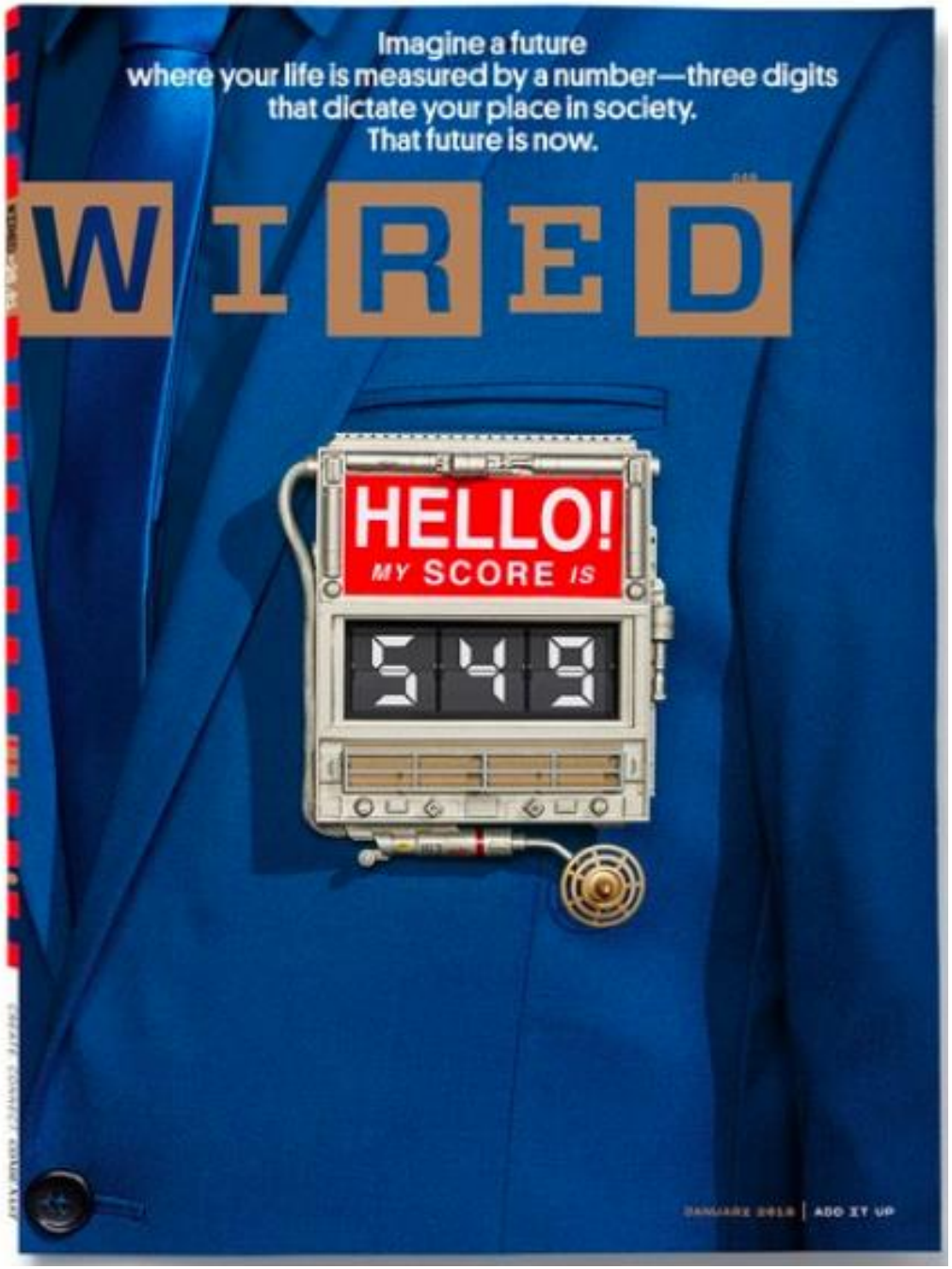
Jeffrey Dastin

8 MIN READ

SAN FRANCISCO (Reuters) - Amazon.com Inc's (AMZN.O) machine-learning specialists uncovered a big problem: their new recruiting engine did not like women.

Falsos positivos y falsos negativos

	Predice NO	**Predice SI**
Actual NO	Verdadero negativo	Falso positivo (Error Tipo I)
Actual SI	Falso negativo (Error tipo II)	Verdadero positivo



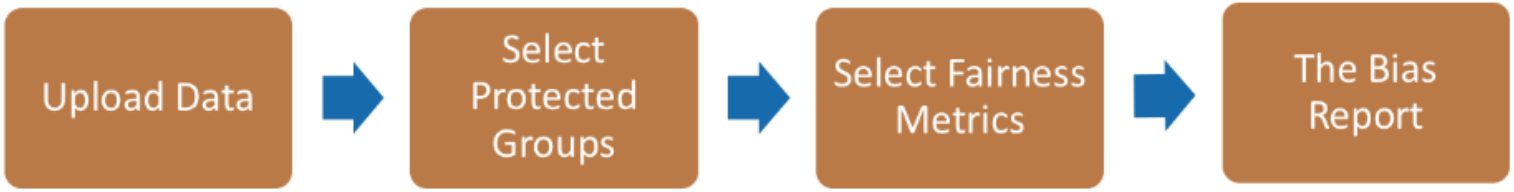
Gob_Lab UAI
UNIVERSIDAD ADOLFO IBÁÑEZ

https://www.wired.com/story/age-of-social-credit/

Análisis de Disparidades

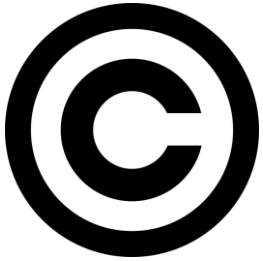
Bias and Fairness Audit Toolkit

This tool is powered by Aequitas, an open source bias audit toolkit for machine learning developers, analysts, and policymakers to audit machine learning models for discrimination and bias, and make informed and equitable decisions around developing and deploying predictive risk-assessment tools.



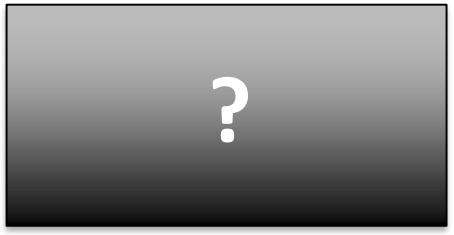
Opacidad

Opacidad intencional



Opacidad analfabeta

Opacidad intrínseca



OPACIDAD ALGORÍTMICA

Tipo de Opacidad	**Método de mitigación**
Intencional	Software libre, regulación "derecho a una explicación": por ej. explicaciones contrafactuales
Analfabeta	Fortalecer los programas educativos en computación, asesoría experta independiente a los afectados por los algoritmos
Intrínseca	Uso de modelos que sean más fáciles de interpretar, aunque tengan menor certeza. Explicaciones contrafactuales

Explicaciones contrafactuales

What-If Tool Home Introduction Features Demos About References AI Fairness Walkthrough PAIR

What If...

you could inspect a machine learning model, with minimal coding required?

Building effective machine learning models means asking a lot of questions. Look for answers using the What-if Tool, an interactive visual interface designed to probe your models better.
Compatible with TensorBoard, Jupyter and Colaboratory notebooks. Works on Tensorflow and Python-accessible models.

Check out this walkthrough to find out what you can learn about your models with the What-If Tool.

For more details, refer to the documentation.

Join the What-If Tool community on our Google Group.

Aumentan las dudas que generan estos mecanismos:

Los ciudadanos exigen saber cómo deciden los algoritmos que hoy afectan sus vidas

Pueden definir si un niño entra a la escuela que sus padres escogieron, si alguien es apto para recibir un bono o incluso si un tercero va o no a la cárcel. ¿Son estos sistemas infalibles o también cometen errores en su juicio?

ALEXIS IBARRA O.

El Estado, tanto en Chile como en el mundo, ya usa algoritmos para decidir quiénes reciben un bono, determinar aquellos que cometen fraude con sus licencias médicas o seleccionar el colegio público al que debe ir un niño. Incluso se han utilizado para enviar a prisión a personas que, al evaluarlas, tienen más posibilidades de reincidir en el delito por el que son juzgadas.

> *"Los algoritmos no solo se usan para entregar beneficios, sino también para sancionar a personas. Para el país, la transparencia algorítmica debe ser un tema de suma importancia".*
>
> **MARÍA PAZ HERMOSILLA,** DIRECTORA DEL LABORATORIO DE INNOVACIÓN PÚBLICA DE LA ESCUELA DE GOBIERNO DE LA UAI

En España, este tema ya está dando que hablar. La fundación Civio solicitó al Gobierno el código del *software* que decide quién es beneficiario de un bono, ante la evidencia de que personas que cumplían con los requisitos no lo recibían. La autoridad no entregó el código fuente, argumentando que estaba protegido por *copyright*.

En Europa existe un creciente movimiento ciudadano que aboga por el derecho a acceder al código de un algoritmo que toma decisiones que afectan sus vidas o, por lo menos, a conocer cómo funciona.

En Chile, la Universidad Adolfo Ibáñez (UAI) está estudiando el te-

FABIÁN RIVAS

lud, la Superintendencia de Seguridad Social, Fonasa y el Ministerio de Educación.

"Nos fue mal. En algunos no obtuvimos respuestas, otros se negaron a entregar la información o nos derivaron a otros estamentos, hasta

comité asesor que elaborará la Política Nacional de Inteligencia Artificial.

El presidente del Consejo para la Transparencia, Jorge Jaraquemada, dice que si el procedimiento o algoritmo asociado consta en un soporte instrumental, no debería negarse su entrega, a menos que exista una causal de secreto o reserva, pero ese es un análisis que se

principio constitucional en que se determina que son públicos los actos y resoluciones de los órganos del Estado, así como sus fundamentos y los procedimientos que utilicen.

Eventualmente, sin embargo, se puede argumentar que un algoritmo está bajo la protección de la propiedad intelectual. "No sería conveniente para una empresa que, por trabajar para el Estado, se vea obligada a entregar un código con pro-

cesos, agrega Romina Garrido, especialista en el área y fundadora de Privacy Consulting.

Aun así, si fuera posible acceder al código de un algoritmo, en la mayoría de los casos sería muy difícil entender la forma cómo funciona, según opina Marcelo Mendoza, investigador del Instituto Milenio de Fundamento de Datos.

"Existen dos tipos de algoritmo para esta toma de decisiones: los que operan a partir de reglas y los que aprenden a partir de datos para generar sus propias reglas. Los primeros son fáciles de entender, pero los segundos son cajas negras que hacen su trabajo, pero no se sabe cómo. El tema es que, hoy en día, el segundo tipo es el predominante", dice el académico.

Los sistemas de aprendizaje profundo o *deep learning* aprenden a partir de un insumo, que son los datos. "Muchas veces no se sabe cómo aprendió y podría existir sesgo. Por ejemplo, un algoritmo podría discriminar a personas de etnia mapuche dependiendo de los datos con los que fue alimentado", ejemplifica Mendoza.

De esta problemática surge un área de investigación llamada "Inteligencia artificial con explicación", donde es posible explicitar las razones de por qué un sistema algorítmico toma una determinada decisión.

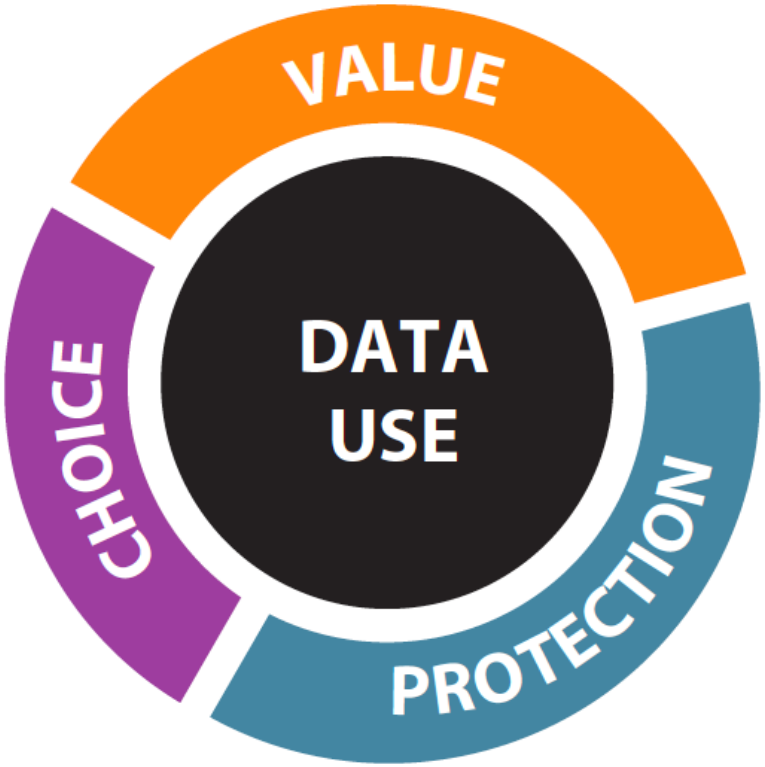
María Paz Hermosilla también menciona las técnicas contrafactuales, que son aquellas con que se

Transparencia significativa

- Brauneis y Goodman (2018)
- "Lo que el público necesita saber"

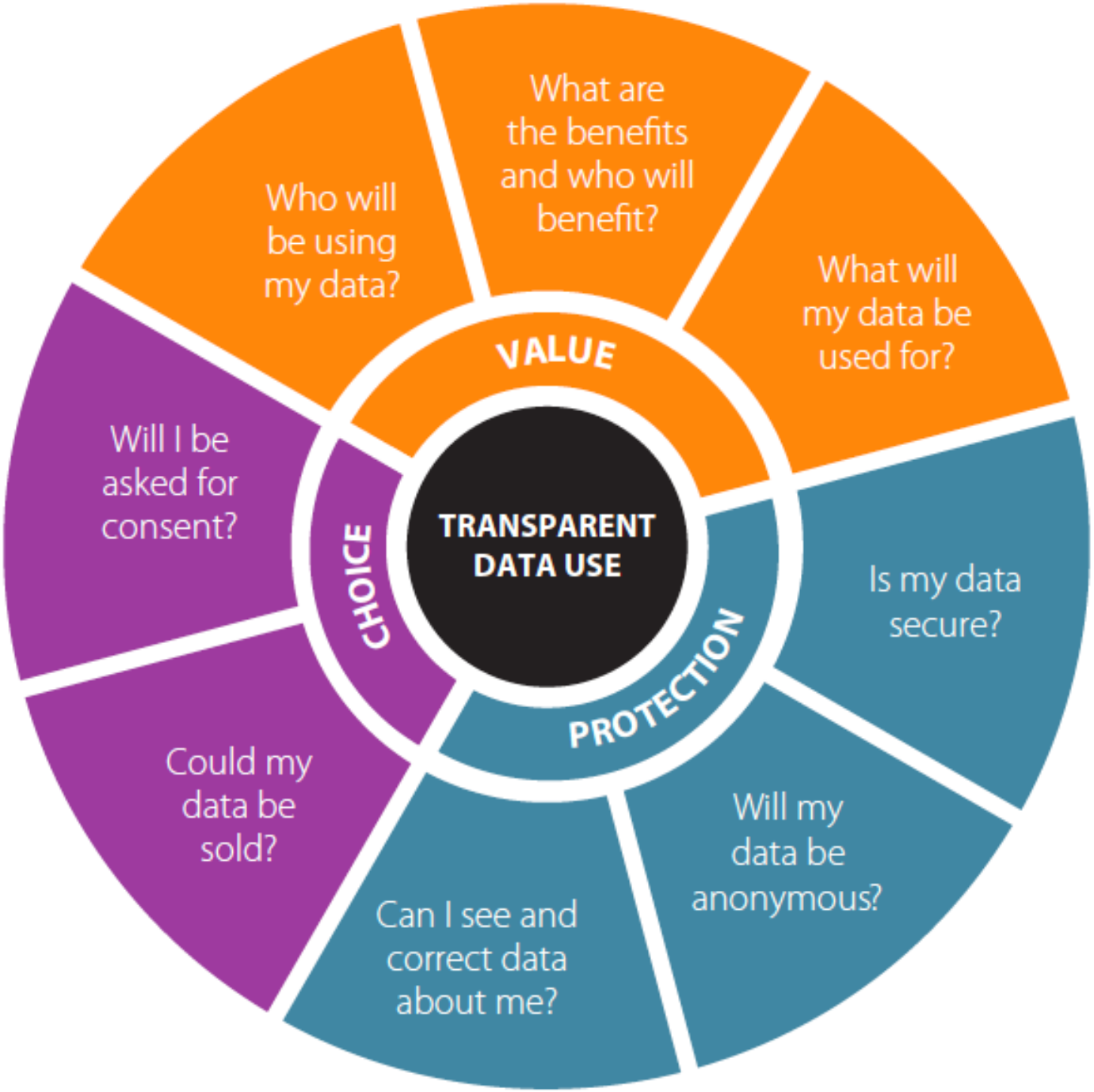
1. Política del algoritmo
2. Rendimiento del algoritmo
3. Equidad del algoritmo
4. Efectos del algoritmo en la capacidad del gobierno

Trusted Data Use Guidelines



Licencia social:
"La aceptación social por parte de la comunidad de la forma en que se usan los datos"

http://www.trusteddata.co.nz/


What are the benefits and who will benefit?
What will my data be used for?
Is my data secure?
Will my data be anonymous?
Can I see and correct data about me?
Could my data be sold?
Will I be asked for consent?
Who will be using my data?
VALUE
PROTECTION
CHOICE
TRANSPARENT DATA USE

POLITICA / CONTROVERSIAL IMPLEMENTACIÓN

El método que aplica Urtubey para predecir el embarazo adolescente

En pleno debate por la despenalización del aborto, el gobernador salteño propone usar inteligencia artificial.

por *Bàrbara Defoix*


ALLEGHENY COUNTY
LIVE
WORK
EXPLORE
GOVERNMENT
NEWS AND EVENTS
ABOUT
ALLEGHENY FAMILY SCREENING TOOL
Type what you're looking for
» Emergency Contacts
» Detailed DHS How Do I?
» Careers
» Doing Business
HOW DO I...
SEE MORE
› Government › Health and Human Services › Human Services › News and Events › Accomplishments › Allegheny Family Screening Tool
The Allegheny Family Screening Tool
Predictive-risk Modeling in Child Welfare in Allegheny County
The Allegheny County Department of Human Services (DHS) is committed to finding new
ACCOMPLISHMENTS AND INNOVATIONS
Advancing Integration
Allegheny Family Screening Tool

Google - Principles of Artificial Intelligence "AI should..."

1. Be socially beneficial.
2. Avoid creating or reinforcing unfair bias.
3. Be built and tested for safety.
4. Be accountable to people.
5. Incorporate privacy design principles.
6. Uphold high standards of scientific excellence.
7. Be made available for uses that accord with these principles.

https://www.blog.google/topics/ai/ai-principles/

Google – Lo que no haremos

1. Technologies that cause or are likely to cause overall harm. Where there is a material risk of harm, we will proceed only where we believe that the benefits substantially outweigh the risks, and will incorporate appropriate safety constraints.

2. Weapons or other technologies whose principal purpose or implementation is to cause or directly facilitate injury to people.

3. Technologies that gather or use information for surveillance violating internationally accepted norms.

4. Technologies whose purpose contravenes widely accepted principles of international law and human rights.

Facial recognition technology: The need for public regulation and corporate responsibility

Jul 13, 2018 | Brad Smith - President

LA GESTIÓN ÉTICA DE LOS DATOS

Por qué importa y cómo hacer un uso justo de los datos en un mundo digital

César Buenadicha
Gemma Galdon
María Paz Hermosilla
Daniel Loewe
Cristina Pombo

BID

Gob_Lab UAI

UNIVERSIDAD ADOLFO IBÁÑEZ

María Paz Hermosilla
Directora GobLab UAI
paz.hermosilla@uai.cl
@mphermosilla
@GoblabUAI