

LIMITACIONES EPISTEMOLÓGICAS DEL *BIG DATA*: UNA RESPUESTA ÉTICA Y POLÍTICA

**X JSPDCP - Jornada de Seguridad y
Protección de Datos de Carácter Personal**

**Escuela de Ingeniería UPV/EHU
[Vitoria-Gasteiz, 31 de octubre de 2018]**

**TXETXU AUSÍN (IFS-CSIC)
txetxu.ausin@cchs.csic.es**

EPISTEMOLOGÍA

La **epistemología** es la rama de la filosofía que trata del conocimiento, de su fundamentación y del método para alcanzarlo

<https://www.youtube.com/watch?v=p9ZdeARKTzE>

METODOLOGÍA 1

Hypothesis-driven model

Observación, articulación de hipótesis y su posterior contrastación con los datos disponibles a fin de establecer inferencias causa-efecto

METODOLOGÍA 2

Data-driven model

Extrae información no trivial, previamente desconocida y potencialmente útil de los datos, mediante el descubrimiento de patrones con significado (correlaciones): *data mining*

OJO AL DATA

No es lo mismo causalidad que correlación

La causalidad siempre implica correlación, pero la correlación no necesariamente implica causalidad

Falacia *cum hoc ergo propter hoc*: “con esto y, por tanto, a causa de esto”

El hecho de que dos eventos se den habitualmente a la vez o de manera consecutiva no implica que uno sea la causa del otro

CORRELACIONES ESPURIAS



tylervigen.com/spurious-correlations

SESGOS

Prejuicio en nuestro modo de procesar la información, una tendencia a percibir de modo distorsionado la realidad, una predisposición

Los sesgos no desaparecen nunca aumentando el tamaño de la muestra; por ello son sesgos y no confusores

ALGORITMOS SESGADOS

1.- Los sesgos implícitos en los datos

2.- El software (el algoritmo) hace suyo el prejuicio o la tendencia subyacente en la sociedad para poder acertar

Aunque los algoritmos se presentan bajo una apariencia de neutralidad, el caso es que no dejan de ser opiniones “encapsuladas”. Y los ejemplos de discriminaciones debidas a los algoritmos y el *big data* son cada vez más numerosos y conocidos

DATOS \neq CONOCIMIENTO

Disponer de ingentes cantidades de datos puede conducir a la confusión y a más falsos positivos

Se privilegia la toma de decisiones (el objetivo del algoritmo) como el principal sentido de la búsqueda de conocimiento (decisionismo), como si todo en la vida consistiese en elegir u optar entre dos alternativas (falacia del falso dilema)

Se pueden considerar valores importantes para un individuo o una sociedad que no cabe cuantificar o reducir a datos

Los algoritmos son conservadores: reflejan lo que hay, lo dado, el prejuicio subyacente en la sociedad, sin ninguna ambición transformadora

SOCIEDAD DE LA IGNORANCIA

Cultura de la distracción y de la baja reflexión: saturación, simplismo, fragmentación y desinformación por sobreinformación (o pseudoinformación-*fakes*: mentiras que queremos creer)

Disponemos de más datos que nunca en la historia la humanidad pero al mismo tiempo ese exceso provoca que los asuntos no se analicen tanto, que se prime la velocidad (aceleración) y la toma de decisiones, que no se delibere

ÉTICA Y POLÍTICA

- TRANSPARENCIA

El algoritmo, sobre todo en temas sensibles, ha de ser accesible, discutible y enmendable

- EVALUACIONES DE IMPACTO ÉTICO

Analizar el impacto del análisis *big data* en relación a la vulnerabilidad (menores, grupos marginalizados, enfermos, ancianos ...): *ethics by design*

- REGULADORES PÚBLICOS

Revisión por organismos públicos y competentes de los algoritmos, al menos de aquellos en instituciones públicas y en áreas especialmente sensibles relacionadas con los derechos fundamentales (justicia, salud, trabajo...)