

Ciencia de Datos

(¿Una alternativa viable para psicólogos?)

Juan C. Correa 

jcc@criticalcentrality.com



Al finalizar esta clase, usted va a

- 1 Comprender algunos desafíos de la ciencia de datos para la formación profesional del psicólogo.
- 2 Identificar las herramientas más usadas en la ciencia de datos y su ecosistema.
- 3 Observar algunos proyectos de "investigación psicológica" desde la óptica de la ciencia de datos.



1 Preliminares

- El dato "analógico" y "digital" de la investigación psicológica
- Ciencia de Datos: Una definición
- Programación Orientada a Objetos

2 Algunos Casos

- Puntualidad en pedidos de comida a domicilio
- El Efecto Sci-Hub

3 Desafíos para psicólogos



El dato "analógico" y "digital" de la investigación psicológica



Un ejemplo de dato "analógico" en la investigación psicológica aplicada (antes de existir Internet)

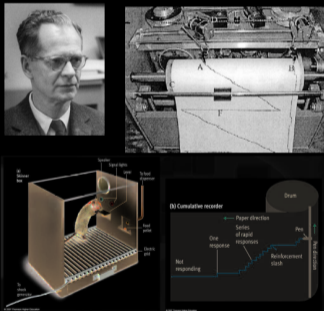


El psicólogo profesionalista que hace terapia o brinda consulta psicológica aplicando sus conocimientos teóricos para "comprender" los síntomas expresados por sus pacientes o clientes.

Históricamente, los datos en psicología provenían de estos dos contextos

LABORATORIO

Skinner y su registro acumulativo
Para medir la frecuencia
de la conducta operante de ratas
o palomas en laboratorio



FUERA DE LABORATORIO

Rensis Likert y su escala Likert
para medir opiniones, pensamientos,
y características de personalidad a
partir de cuestionarios



Un ejemplo de dato "digital" en la investigación aplicada (en la era de Internet)



Carlos Piña  @Piniisima · 10h

Noten la **desesperación e ignorancia** de los **propagandistas**, sobre la **Inteligencia Artificial**.

- No saben que es aprendizaje supervisado y no supervisado
- No distinguen entre test set y training set
- No entienden redes neuronales
- Nunca han programado un agente inteligente

32

162

397

6.1K



Carlos Piña  @Piniisima · 13h

El audio de **Batres**, con **95% de probabilidad** de ser **voz natural**, nos dice muchas cosas:

- ✓ Confirman sus campañas de ataque en redes
- ✓ Exhibe a sus columnistas propagandistas
- ✓ No tienen idea de qué es la Inteligencia Artificial
- ✓ Manipulan encuestas

El científico que investiga la probabilidad de que una nota de audio sea natural (verdadera) o fabricada por la inteligencia artificial (falsa).





Contents lists available at [ScienceDirect](#)

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid



Short Communication

Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes

Saifuddin Ahmed^{*}

Nanyang Technological University, Singapore



ARTICLE INFO

Keywords:

Deep fakes
Social media
Claim accuracy
Cognitive ability
Disinformation

ABSTRACT

We examine how individual differences influence perceived accuracy of deepfake claims and sharing intention. Rather than political deepfakes, we use a non-political deepfake of a social media influencer as the stimulus, with an educational and a deceptive condition. We find that individuals are more likely to perceive the deepfake claim to be true when informative cues are missing along with the deepfake (compared to when they are present). Also, individuals are more likely to share deepfakes when they consider the fabricated claim to be accurate. Moreover, we find that cognitive ability plays a moderating role such that when informative cues are present (educational condition), individuals with high cognitive ability are less trustful of deepfake claims. Unexpectedly, when the informative cues are missing (deceptive condition), these individuals are more likely to consider the claim to be true and share them. The findings suggest that adding corrective labels can help reduce inadvertent sharing of disinformation. Also, user biases should be considered in understanding public engagement with disinformation.

(Ahmed, 2021)



¿Y qué pasa si las explicaciones teóricas del psicólogo profesionalista son falsas?



¿Cómo estar seguro de tener los conocimientos teóricos correctos?



Es difícil estar 100 % seguros

(la psicología como cualquier otra ciencia solo alcanza el conocimiento probable)



¿Existen herramientas para saber hasta qué punto son válidos y ciertos los conocimientos teóricos?



1 Con la ciencia de datos se pueden usar herramientas para acumular y analizar evidencia empírica que apoya o refuta una teoría (y ahora hay que aprender a distinguir evidencia natural de evidencia ficticia o fabricada por la IA).



- 1 Con la ciencia de datos se pueden usar herramientas para acumular y analizar evidencia empírica que apoya o refuta una teoría (y ahora hay que aprender a distinguir evidencia natural de evidencia ficticia o fabricada por la IA).
- 2 Con el diseño experimental se pueden establecer mecanismos que garantizan la calidad y validez de nuestras interpretaciones de la evidencia empírica (teorías).

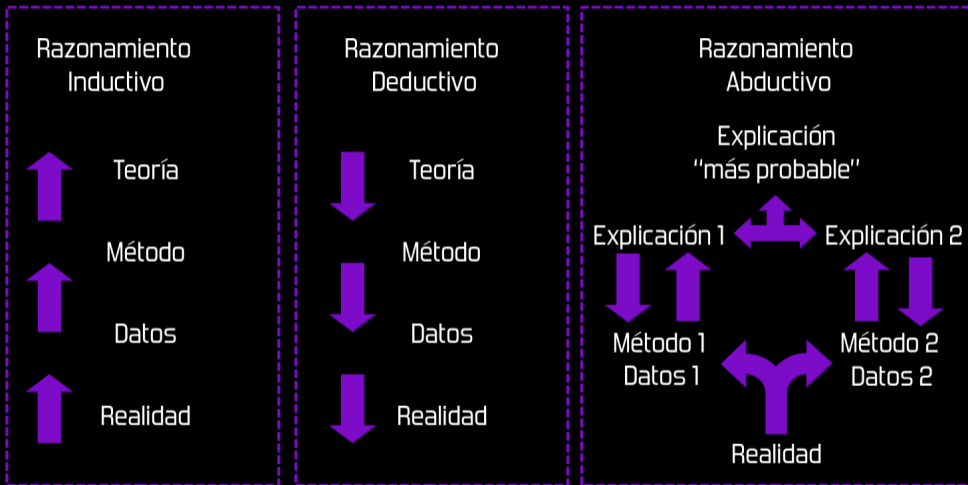


- 1 Con la ciencia de datos se pueden usar herramientas para acumular y analizar evidencia empírica que apoya o refuta una teoría (y ahora hay que aprender a distinguir evidencia natural de evidencia ficticia o fabricada por la IA).
- 2 Con el diseño experimental se pueden establecer mecanismos que garantizan la calidad y validez de nuestras interpretaciones de la evidencia empírica (teorías).
- 3 Con la filosofía de las ciencias se puede reflexionar sobre la vigencia de las teorías en psicología, considerando los cambios económicos, sociales y tecnológicos de nuestra sociedad.



- 1 Con la ciencia de datos se pueden usar herramientas para acumular y analizar evidencia empírica que apoya o refuta una teoría (y ahora hay que aprender a distinguir evidencia natural de evidencia ficticia o fabricada por la IA).
- 2 Con el diseño experimental se pueden establecer mecanismos que garantizan la calidad y validez de nuestras interpretaciones de la evidencia empírica (teorías).
- 3 Con la filosofía de las ciencias se puede reflexionar sobre la vigencia de las teorías en psicología, considerando los cambios económicos, sociales y tecnológicos de nuestra sociedad.
- 4 Con las tres (ciencia de datos, diseño experimental y filosofía de las ciencias) se puede incorporar la evidencia empírica a la reflexión sobre la producción y uso del conocimiento científico en psicología.






Una definición de Ciencia de Datos





REAL ACADEMIA ESPAÑOLA

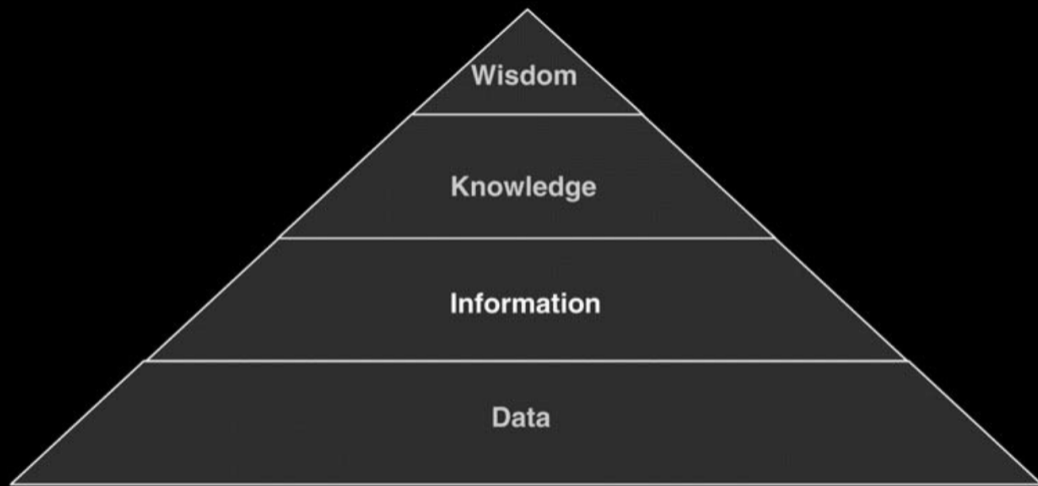
dato¹ 

Del lat. *datum* 'lo que se da'.

1. m. Información sobre algo concreto que permite su conocimiento exacto o sirve para deducir las consecuencias derivadas de un hecho. *A este problema le faltan datos numéricos.*
2. m. Documento, testimonio, fundamento.
3. m. *Inform.* Información dispuesta de manera adecuada para su tratamiento por una computadora.

La palabra "datos" se usa como sinónimo de "información" y "conocimiento".





Conceptual Approaches for Defining Data, Information, and Knowledge

Chaim Zins

Knowledge Mapping Research, 26 Hahaganah Street, Jerusalem 97852, Israel.
E-mail: chaim.zins@gmail.com

The field of Information Science is constantly changing. Therefore, information scientists are required to regularly review—and if necessary—redefine its fundamental building blocks. This article is one of a group of four articles, which resulted from a Critical Delphi study conducted in 2003–2005. The study, “Knowledge Map of Information Science,” was aimed at exploring the foundations of information science. The international panel was composed of 57 leading scholars from 16 countries, who represent (almost) all the major subfields and important aspects of the field. This particular article documents 130 definitions of *data*, *information*, and *knowledge* formulated by 45 scholars, and maps the major conceptual approaches for defining these three key concepts.

basic building blocks of the field. For this very reason, the formulation of systematic conceptions of data, information, and knowledge is crucial for the development of a systematic conception of Information Science, as well as for the construction of a systematic knowledge map of the field.

Data, Information, and Knowledge

The academic and professional IS literature supports diversified meanings for each concept. Evidently, the three key concepts are interrelated, but the nature of the relations among them is debatable, as well as their meanings.



The wisdom hierarchy: representations of the DIKW hierarchy

Jennifer Rowley

Bangor Business School, University of Wales, Bangor, UK

Received 30 January 2006

Revised 25 May 2006

Abstract.

This paper revisits the data–information–knowledge–wisdom (DIKW) hierarchy by examining the articulation of the hierarchy in a number of widely read textbooks, and analysing their statements about the nature of data, information, knowledge, and wisdom. The hierarchy referred to variously as the ‘Knowledge Hierarchy’, the ‘Information Hierarchy’ and the ‘Knowledge Pyramid’ is one of the fundamental, widely recognized and ‘taken-for-granted’ models in the information and knowledge literatures. It is often quoted, or used implicitly, in definitions of data, information and knowledge in the information management, information systems and knowledge management literatures, but there has been limited direct discussion of the hierarchy. After revisiting Ackoff’s original articulation of the hierarchy, definitions of data, information, knowledge and wisdom as articulated in recent textbooks in information systems and knowledge management are reviewed and assessed, in pursuit of a consensus on definitions and transformation processes. This process brings to the surface the extent of agreement and dissent in relation to these definitions, and provides a basis for a discussion as to whether these articulations present an adequate distinction between data, information, and knowledge. Typically information is defined in terms of data, knowledge in terms of information, and wisdom in terms of knowledge, but there is less consensus in the description of the processes that transform elements lower in the hierarchy into those above them, leading to a lack of definitional clarity. In addition, there is limited reference to wisdom in these texts.

Keywords: DIKW hierarchy; wisdom hierarchy; wisdom; knowledge management; wisdom management



La ciencia de datos puede entenderse como una disciplina orientada a "transformar" los datos en información y conocimiento (en cualquier formato almacenable en una computadora).

Información es el resumen (estadístico, computacional, o matemático) de los datos recolectados analógicamente (sin usar tecnología digital) o digitalmente (usando tecnología digital).

Conocimiento es información que cobra significado en algún marco de referencia conceptual o teórico.

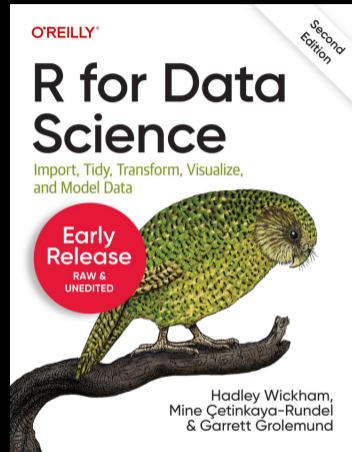
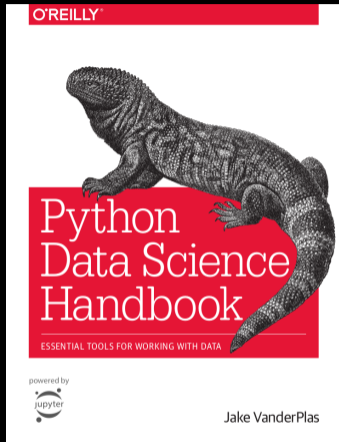




Aunque la distinción entre dato, información y conocimiento está bien establecida (Zins, 2007), sobre ella se distingue la importancia de introducir el concepto de "sabiduría" (Rowley, 2007).

Sabiduría implica el juicio ético o moral que le otorgamos al conocimiento que tenemos de una información dada.





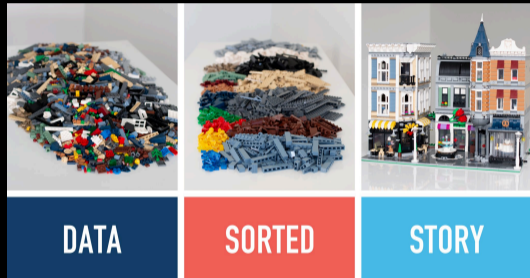
La ciencia de datos se hace con lenguajes de programación orientada a objetos.



Programación Orientada a Objetos



La P00 es un paradigma de la programación computacional que puede verse como piezas de lego. Cada pieza es un objeto con su propia forma y función. Estos objetos pueden interactuar entre sí, heredar características y ser bastante flexibles (Van Roy & Haridi, 2004).



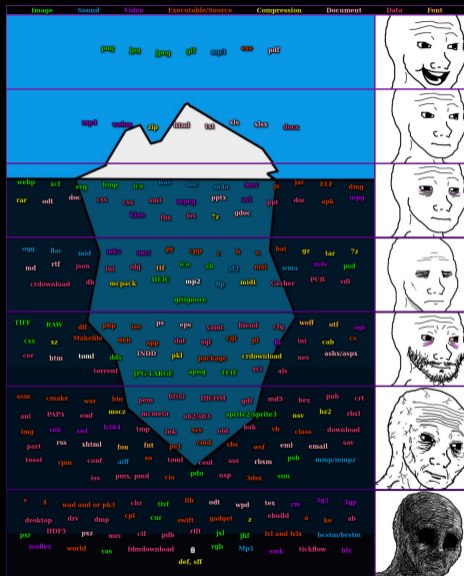
¿Por qué es necesaria la P00 para la ciencia de datos?



sorted by typical frequency of occurrence

when you are proposing an entry, please add a description

when you are proposing an entry, please add a description



La POO permite:



La POO permite:

- 1 Documentar procedimientos para analizar datos en formatos comunes (e.g., xlsx, csv, pdf) y no comunes (e.g., aiff, bfy), comprendiendo los formatos de datos como un iceberg (https://icebergcharts.com/i/File_Types)



La P00 permite:

- 1 Documentar procedimientos para analizar datos en formatos comunes (e.g., xlsx, csv, pdf) y no comunes (e.g., aiff, bfy), comprendiendo los formatos de datos como un iceberg (https://icebergcharts.com/i/File_Types)
- 2 Acelerar la transferencia tecnológica al favorecer la reproducción de hallazgos reportados en trabajos de investigación con repositorios de datos y códigos (apropiado para comunidades en línea como Kaggle o Hugging Face).



La P00 permite:

- 1 Documentar procedimientos para analizar datos en formatos comunes (e.g., xlsx, csv, pdf) y no comunes (e.g., aiff, bfy), comprendiendo los formatos de datos como un iceberg (https://icebergcharts.com/i/File_Types)
- 2 Acelerar la transferencia tecnológica al favorecer la reproducción de hallazgos reportados en trabajos de investigación con repositorios de datos y códigos (apropiado para comunidades en línea como Kaggle o Hugging Face).
- 3 Aumentar la precisión del vocabulario usado para describir, analizar, explicar y predecir tendencias en los datos.



Algunos Casos



Puntualidad en pedidos de comida a domicilio





ELSEVIER

Contents lists available at ScienceDirect

Journal of Retailing and Consumer Services

journal homepage: www.elsevier.com/locate/jretconser



Evaluation of collaborative consumption of food delivery services through web mining techniques



Juan C. Correa^{a,*,1}, Wilmer Garzón^b, Phillip Brooker^c, Gopal Sakarkar^d, Steven A. Carranza^a, Leidy Yunado^a, Alejandro Rincón^a

^a Faculty of Psychology, Fundación Universitaria Konrad Lorenz, Bogotá, Colombia

^b Escuela Colombiana de Ingeniería Julio Garavito, Bogotá, Colombia

^c Department of Sociology, Social Policy, and Criminology at University of Liverpool, UK

^d Department of Computer Applications, Rasoni College of Engineering, Nagpur, India

(Correa et al., 2019)



Google Typical Traffic



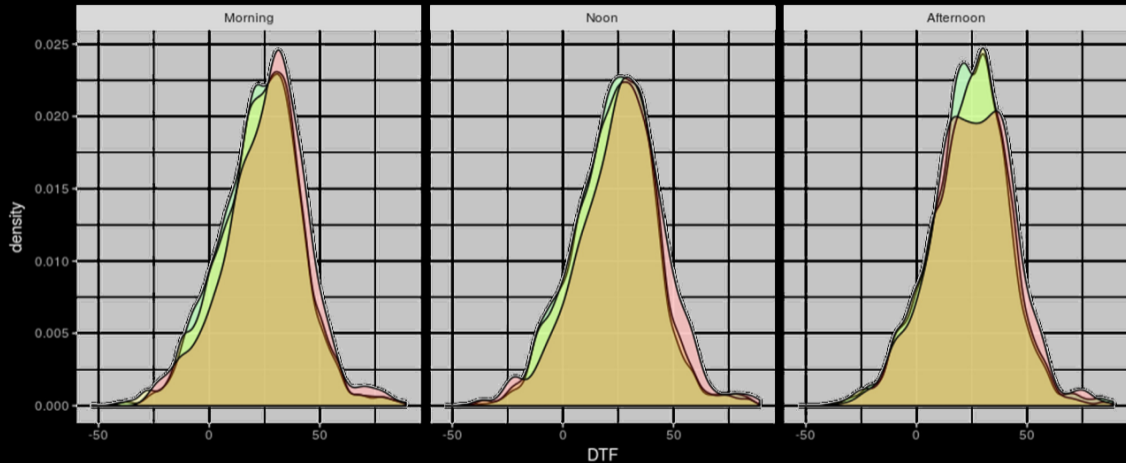
Green



Orange



Red



El Efecto Sci-Hub





The Sci-Hub effect on papers' citations

Juan C. Correa^{1,2} · Henry Laverde-Rojas³ · Julian Tejada⁴ ·
Fernando Marmolejo-Ramos⁵

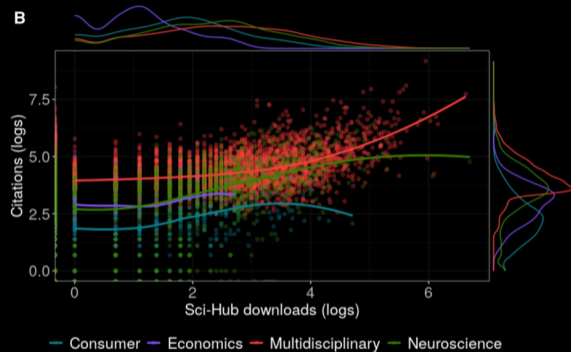
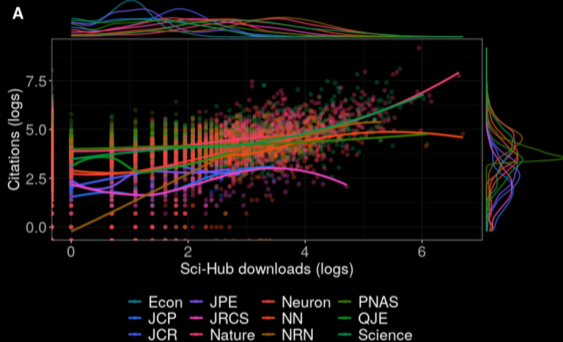
Received: 28 June 2020 / Published online: 25 January 2021
© Akadémiai Kiadó, Budapest, Hungary 2021

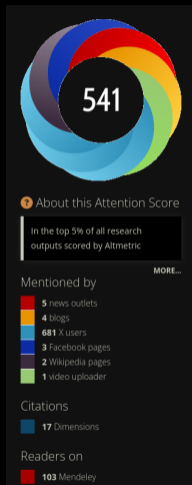
Abstract

Citations are often used as a metric of the impact of scientific publications. Here, we examine how the number of downloads from Sci-Hub as well as various characteristics of publications and their authors predicts future citations. Using data from 12 leading journals in economics, consumer research, neuroscience, and multidisciplinary research, we found that articles downloaded from Sci-Hub were cited 1.72 times more than papers not downloaded from Sci-Hub and that the number of downloads from Sci-Hub was a robust predictor of future citations. Among other characteristics of publications, the number of figures in a manuscript consistently predicts its future citations. The results suggest that limited access to publications may limit some scientific research from achieving its full impact.

Keywords Citations · Knowledge dissemination · Scholar consumption · Scientific impact · Sci-Hub







Geographical breakdown

Country	Count	As %
France	47	7%
United States	37	5%
Spain	36	5%
United Kingdom	32	5%
Germany	28	4%
Brazil	24	4%
India	14	2%
Switzerland	14	2%
Poland	11	2%
Other	149	22%
Unknown	289	43%

Demographic breakdown

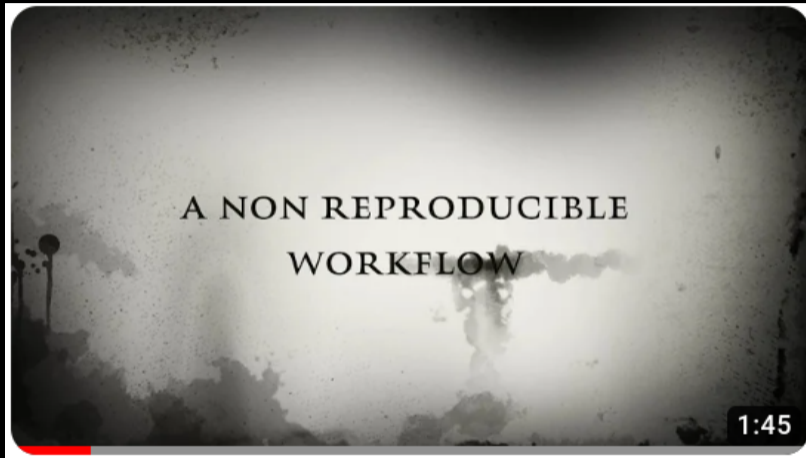
Type	Count	As %
Members of the public	490	72%
Scientists	144	21%
Science communicators (journalists, bloggers, editors)	34	5%
Practitioners (doctors, other healthcare professionals)	13	2%



Desafíos para psicólogos



¡Menos clics, más código!



(<https://youtu.be/s3JldKoA0zw?si=xDVesUsw0-tz9m5B>)



Mayor alfabetización científica



Software
"sin código"



Software
"basado
en código"



lógica, computación, matemática y estadística



Referencias I

- Ahmed, S. (2021). Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes. *Personality and Individual Differences*, 182, 111074. <https://doi.org/10.1016/j.paid.2021.111074>
- Correa, J. C., Garzón, W., Brooker, P., Sakarkar, G., Carranza, S. A., Yunado, L., & Rincón, A. (2019). Evaluation of collaborative consumption of food delivery services through web mining techniques. *Journal of Retailing and Consumer Services*, 46, 45-50. <https://doi.org/10.1016/j.jretconser.2018.05.002>
- Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science*, 33(2), 163-180. <https://doi.org/10.1177/01655551506070706>
- Van Roy, P., & Haridi, S. (2004). *Concepts, Techniques, and Models of Computer Programming*. Massachusetts Institute of Technology.
- Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *Journal of the American society for information science and technology*, 58(4), 479-493. <https://doi.org/10.1002/asi.20508>



<https://doi.org/10.1002/asi.20508>

