

Big data: orientarsi tra informazione, correlazione e opinioni

Scuola Internazionale Superiore per la Ricerca Interdisciplinare
Roma

Dove abita la verità?

Riflessioni sul vero e sul falso nell'epoca contemporanea

Michele Crudele

2020-02-08



INFORMAZIONE

Notizia, dato o elemento che consente di avere conoscenza più o meno esatta di fatti, situazioni, modi di essere, ecc.

Treccani

In ambito informatico si distingue tra dati (non interpretati) e informazioni (con un significato)

UN NOTO QUOTIDIANO ITALIANO

corriere.it

Coronavirus, chi è il primo morto fuori dalla Cina: 44 anni, era sano

di Redazione Esteri

I primi due contagiati dal coronavirus nelle Filippine, che da ieri è uno dei [ventiquattro Paesi colpiti dal virus](#) fuori dalla Cina, fanno particolarmente notizia: nella notte uno dei due è morto, ed è diventato la prima vittima del virus fuori dalla Cina, [che nel frattempo conta circa 304 vittime](#). L'uomo, le cui generalità non sono note, è un cinese di Wuhan, malato da tempo: i sintomi del virus avevano avuto tempo a degenerare in una «polmonite grave», hanno scritto i medici dell'ospedale di Manila dove il

**Notizia
replicata
da altre
pagine in
rete**

UNA FONTE
NOTORIAMENTE
AFFIDABILE

ANSA.it › Sport › Sci › **Mondiali di sci, Goggia argento in SuperG**

Mondiali di sci, Goggia argento in SuperG

Azzurra beffata per soli 2" dalla Shiffrin che si prende l'oro

ANSA.it › Ultima Ora › **Ergastolo a inventore Nautilus**

Ergastolo a inventore Nautilus

Per l'uccisione della giornalista svedese Kim Wall

Scriveri alla redazione

(ANSA) - COPENHAGEN, 25 APR - È stato condannato a morte per l'omicidio della giornalista svedese Kim Wall, l'inventore del sottomarino Nautilus, Peter Madsen. I resti della ragazza, che lui aveva invitato sul suo sommergibile per un reportage, furono trovati su una spiaggia vicino a Copenaghen alla fine di agosto, 11 giorni dopo la scomparsa, da un ciclista di passaggio.

Pubblicità 4w



L'illustrazione mostra l'arrivo di Rolando Petri

Ricordi, gioia, delusione, tensione. Gli attimi interminabili della gara, le emozioni e le immagini che rimangono, indelebili nella memoria di ciascuno. Momenti irripetibili, di gloria e tristezza che raccontano la storia delle Olimpiadi attraverso semplici gesti

Campioni del passato e del presente come Mark Spitz, Greg Louganis, Magic Johnson, Roger Federer, Pietro Mennea, Dorando Pietri, il maratoneta italiano che vinse alle Olimpiadi di Londra del 1908, ma fu poi squalificato perché sostenuto negli ultimi metri del percorso.

Il pluri medagliato; Michael Phelps, la mano tremolante del campione di box Cassius Clay con la torcia olimpica ad Atlanta.

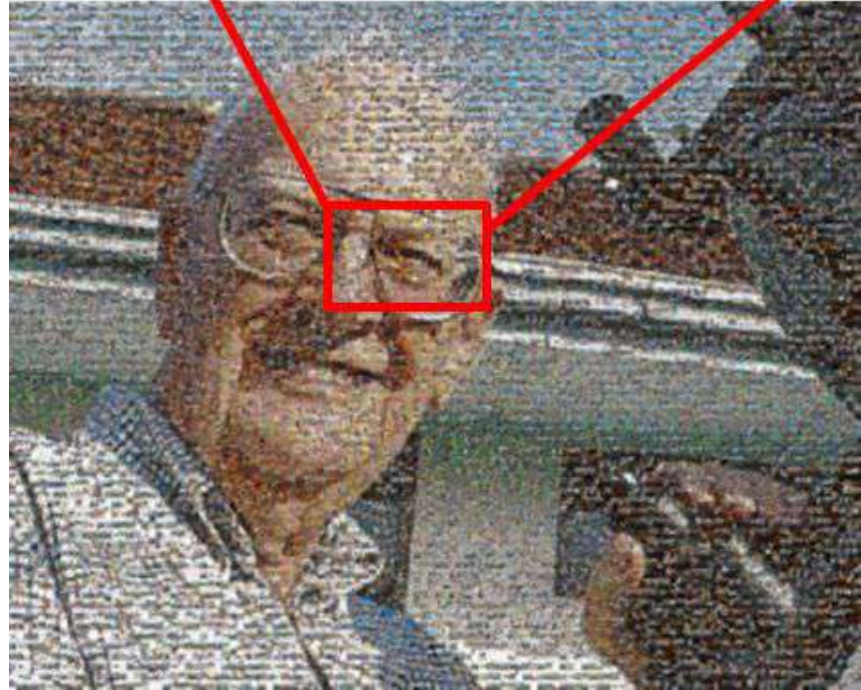
Usain Bolt, l'uomo più veloce del mondo con le scarpe d'orate slacciate; le lunghissime unghie delle velociste Ga

La gatta
frettolosa
fa i figli
ciechi

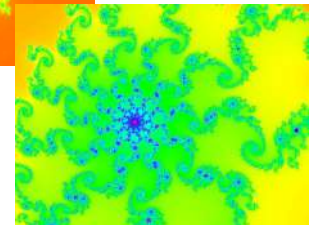
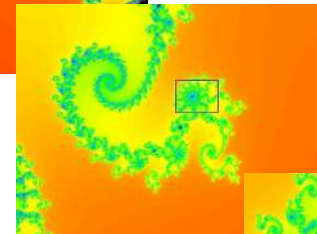
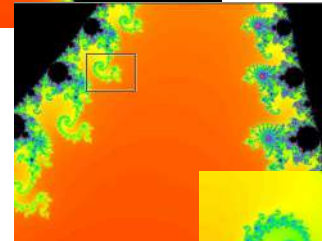
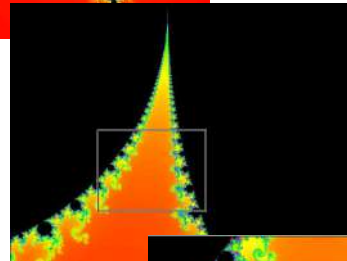
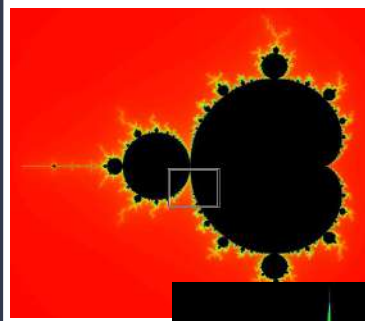
IL TROPPO STROPPIA

Necessariamente
tendiamo a perdere di
vista un'entità nel suo
complesso se
prestiamo attenzione
focalmente ai suoi
particolari

Michael Polanyi



FRATTALI



L'informazione
è generata
attraverso il
calcolo.

WIKIPEDIA ÜBER ALLES

È una tirannia contemporanea

Non possiamo farne a meno

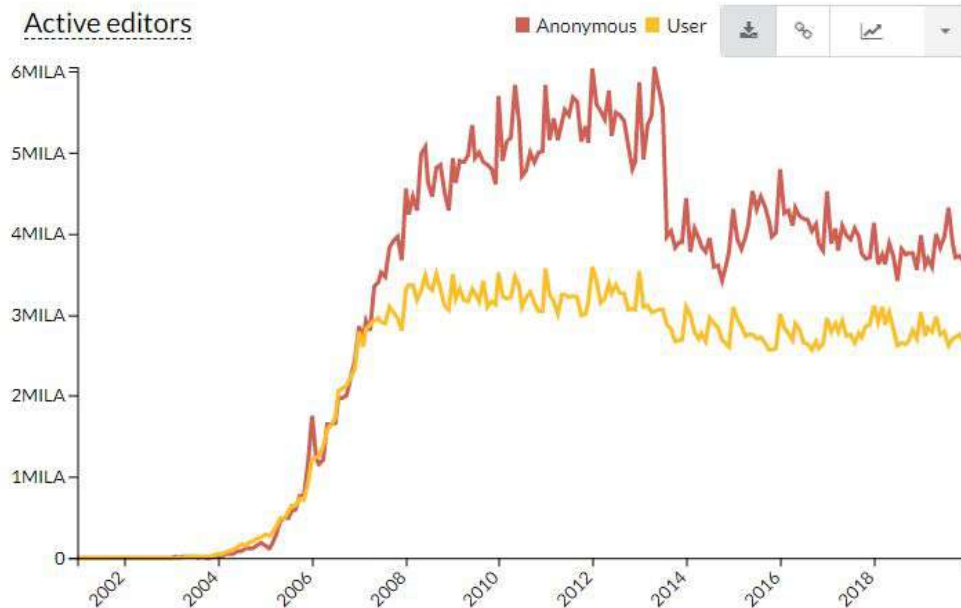
È somma di imperfezioni che dà
origine a (quasi) perfezione

La scrivono in pochi, la leggono tutti



CHI SCRIVE LA WIKIPEDIA ITALIANA

Autori con 5 o più modifiche, negli anni



in 1 mese accedono
50 milioni di dispositivi unici
su 40 milioni di utenti del web in Italia

CORRELAZIONE

Relazione **reciproca**, intima corrispondenza tra due termini, tra due (o anche tra più) elementi.

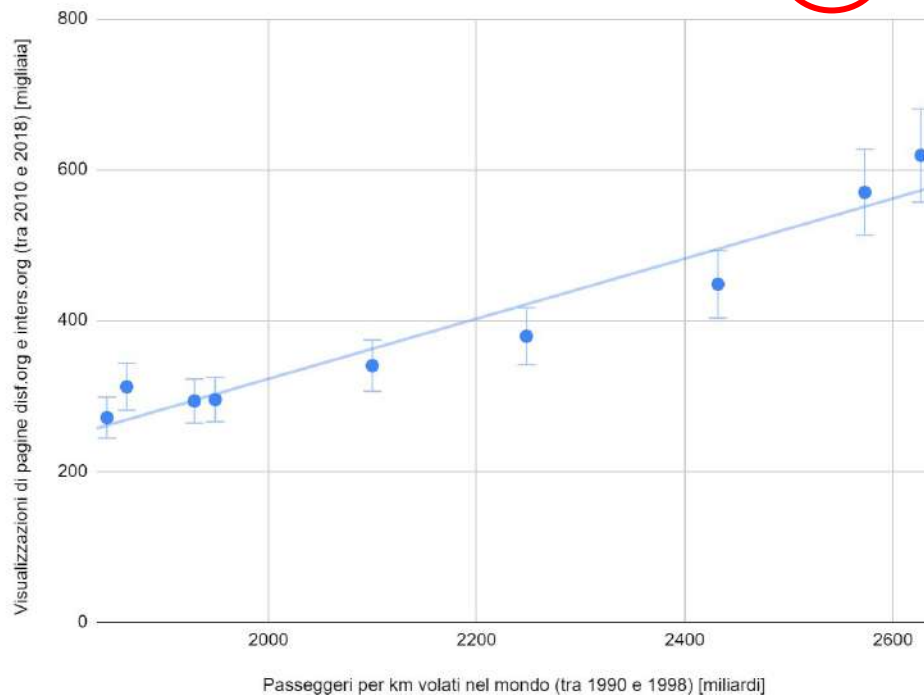
Tra due variabili empiriche esiste correlazione quando si constata la tendenza di una di esse a variare con un'approssimazione più o meno grande (con un grado di correlazione più o meno alto) in funzione dell'altra.

Treccani

UNA CORRELAZIONE SPECIALE QUASI PERFETTA

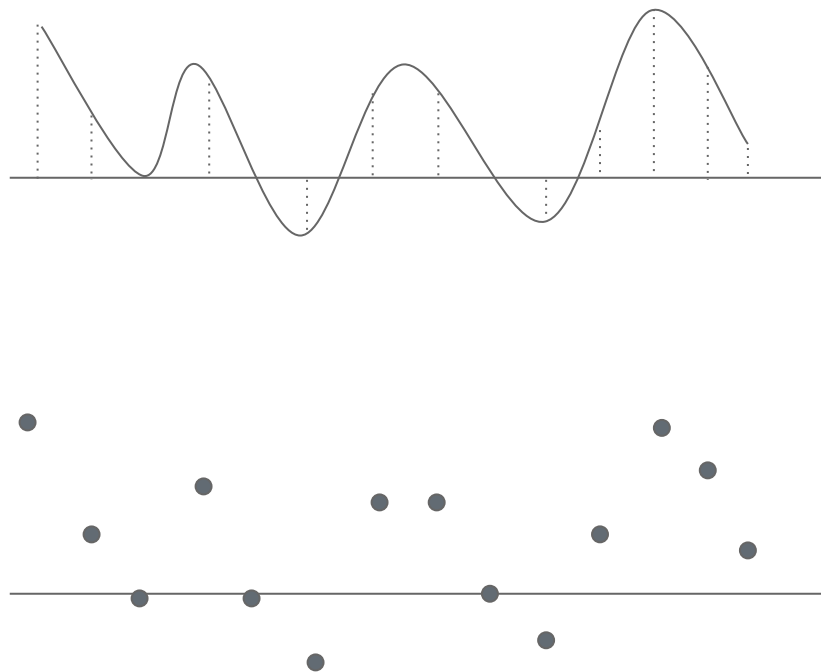
Visualizzazioni di pagine disf.org e inters.org (tra 2010 e 2018)
rispetto a Passeggeri per km volati (tra 1990 e 1998)

● Visualizzazioni di pagine disf.org e inters.org (tra 2010 e 2018)
— Linea di tendenza per Visualizzazioni di pagine disf.org e inters.org (tra 2010 e 2018) $R^2 = 0,93$



DIGITALIZZARE
VUOL DIRE
SCEGLIERE

Campionamento



PERCHÉ NGRAM VIEWER SOLO FINO AL 2008?

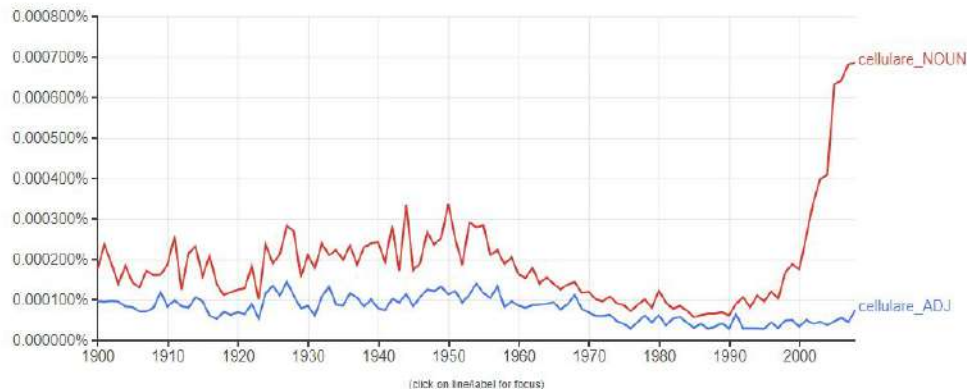
Google sottopone a scansione milioni di libri, continuamente

Ci sono molti libri scanditi successivamente al 2008 ma non in tutte le discipline

rischio di un risultato condizionato

Google Books Ngram Viewer

Graph these comma-separated phrases: case-insensitive
between and from the corpus with smoothing of [Search lots of books](#)



GARBAGE IN GARBAGE OUT

La disponibilità di maggiori dati da analizzare non è sempre positiva

Se i nuovi dati non sono pertinenti, aggiungono rumore

Se i nuovi dati sono parziali, aggiungono un bias

Come distinguere dati utili da inutili?

Machine learning

La collezione di immagini per l'allenamento

Ingannare il riconoscimento facciale

GOOGLE SEARCH ENGINE

Artificial intelligence would be the ultimate version of Google.

So we have the ultimate search engine that would understand everything on the Web. It would understand exactly what you wanted, and **it would give you the right thing.**

Larry Page, 2000

Quanti sanno come funziona la graduatoria dei risultati di una ricerca su Google?

La risposta unica alla ricerca vocale

GOOGLE NON È UN MODELLO

Google's founding philosophy is that we don't know why this page is better than that one: if the statistics of incoming links say it is, that's good enough. No semantic or causal analysis is required.

Chris Anderson, The End of Theory, 2008

All models are wrong, and increasingly you can succeed without them.

Peter Norvig, Google's research director

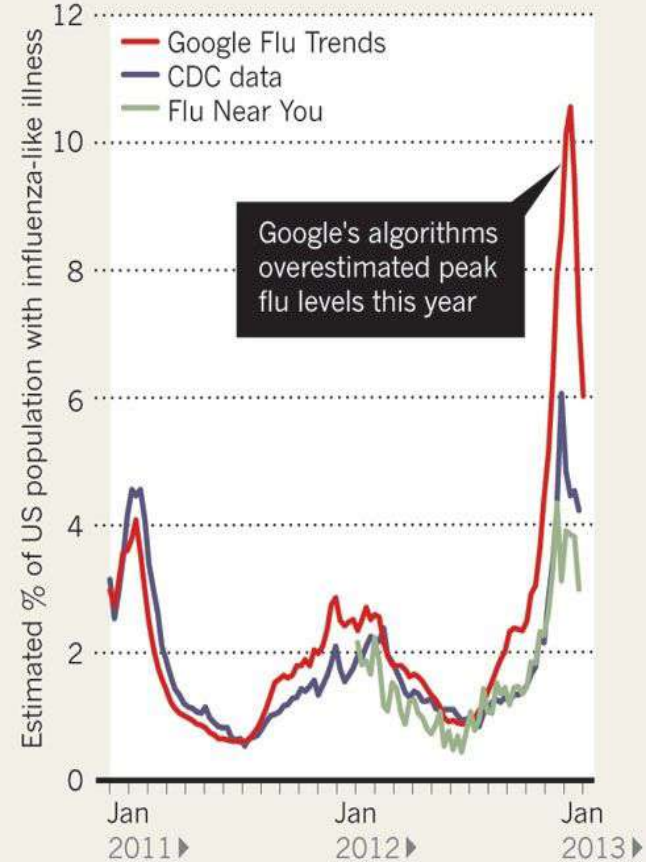
GOOGLE FLU TRENDS OR FLU FAKE?

Fallimento di un progetto innovativo di portata mondiale basato sulle ricerche della gente su Google

Flu Near You crowdsourcing funziona meglio

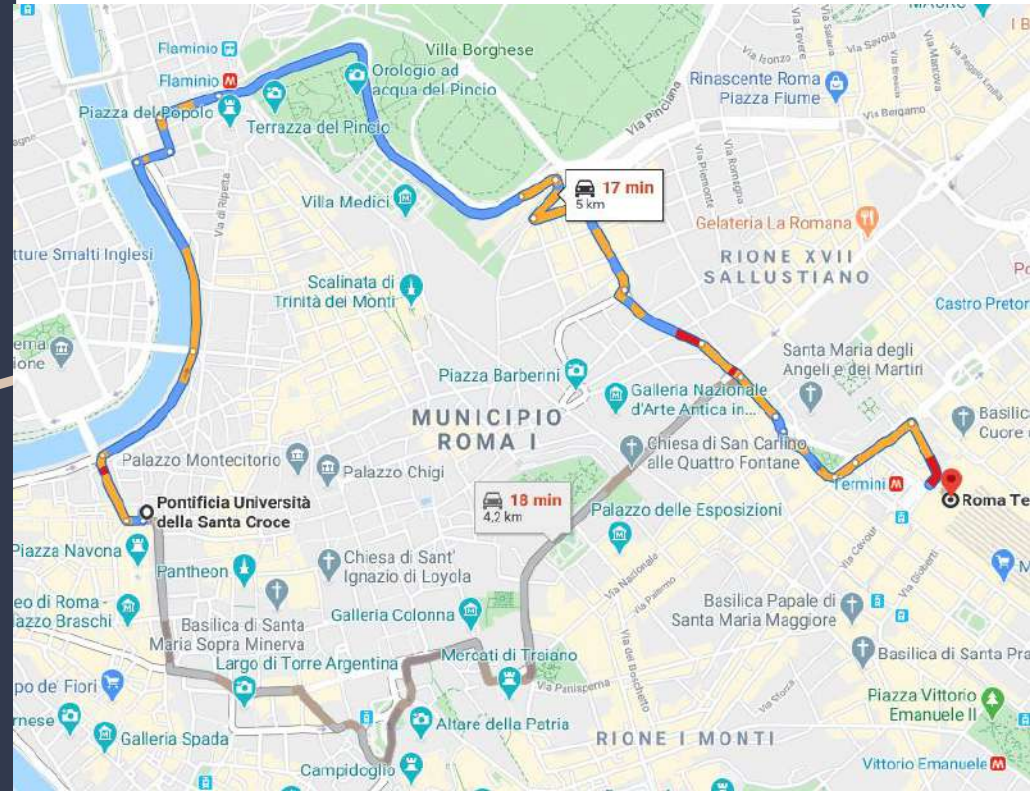
FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.



ACCOMPAGNAMI TU

Google Maps calcola i percorsi anche
in base al traffico registrato dai
dispositivi connessi in rete



ANALISI AUTOMATICA

Individuato come corriere di Al Qaida
era invece un giornalista che cercava di
intervistare gente di quel gruppo

L'analisi era basata sui suoi
spostamenti

Inferenza insufficiente

Un'analisi umana dei dati avrebbe
dedotto che non era un terrorista

GRAMELLINI 2013

La **dittatura dell'algoritmo** è l'ultimo rifugio di un certo tipo di persone, per lo più maschi intellettuali con il cuore a forma di granchio e gli occhi a forma di dollaro, che non riuscendo più a sentire niente si illudono di domare le loro insicurezze con una **serie di algide formulette** attinte dalla marea di dati personali che le nuove tecnologie mettono a disposizione.

Ormai esiste un algoritmo per tutto: il giornale perfetto, il pranzo perfetto, il delitto perfetto. Questi aridi manichini del sapere moderno **pensano di controllare la realtà**, racchiudendola in una previsione statistica che consenta di **anticipare i comportamenti umani** per offrirli in pasto ai pubblicitari.

INCINTA

Già nel 2012 Target riuscì a capire che un'adolescente era incinta prima che lo sapesse suo padre

sulla base degli acquisti di saponi,
integratori, fazzoletti

Il papà si lamentò con i dirigenti di Target perché le mandavano pubblicità di culle e pannolini

Are you trying to encourage her to get pregnant?

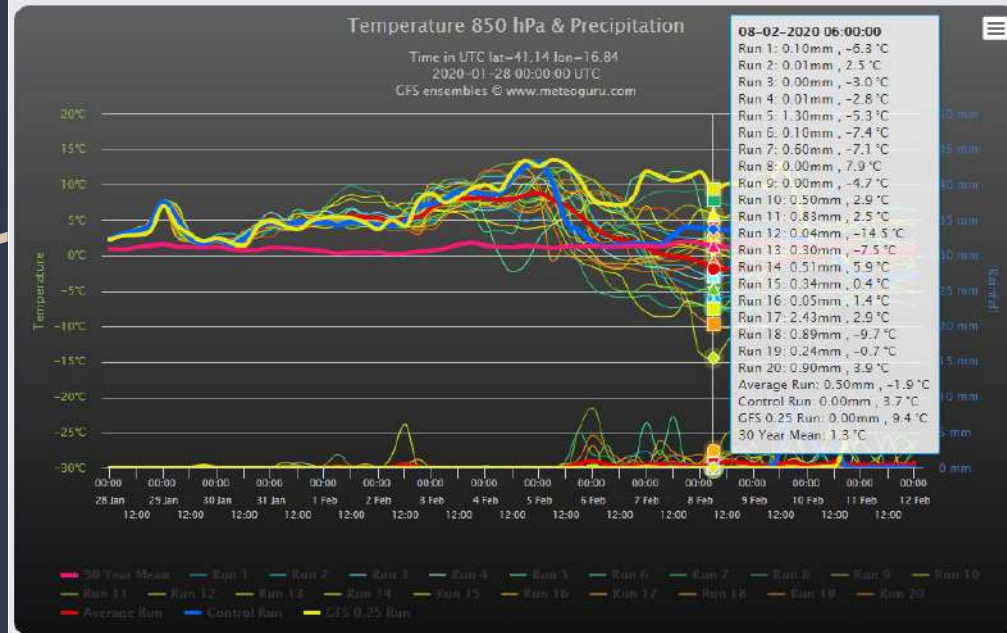
Dovette scusarsi successivamente

BIG DATA OR JUST THE NEEDED DATA FOR THE MODEL ?

Previsioni atmosferiche basate su modelli
matematici con variabili

ma non bisogna eccedere nel loro numero

Variazioni di parametri iniziali fanno
divergere le simulazioni nel tempo



BIG ERRORS

Quando ero universitario mi divertivo a fare previsioni dei miei voti sulla base della curva polinomiale precedente
era un gioco per imparare i metodi

Una grande azienda ha affidato a un laureando la previsione di consumi elettrici analizzando grandi moli di dati storici, senza associare alcun altro elemento di valutazione
è stato un esperimento quasi inutile

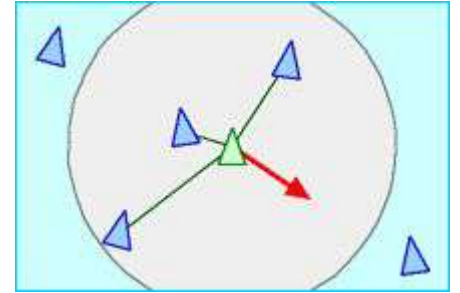
BOIDS

Gli stormi di
uccelli

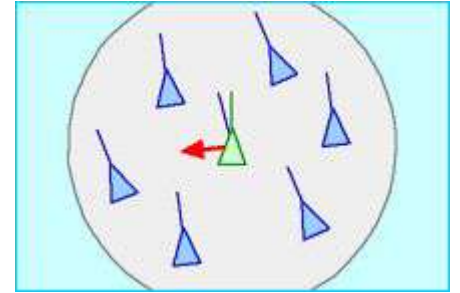
Tre semplici
regole

Un modello
elementare per
un risultato
complesso

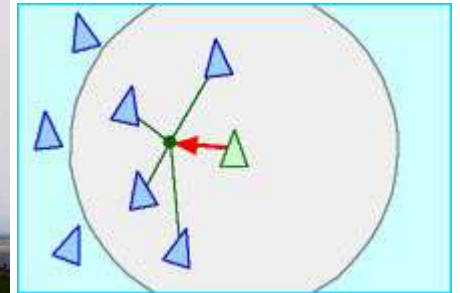
separazione



allineamento



coesione



LE MAREE

Ne conosciamo le cause

Periodicità, somma di componenti

Si prevedono tramite le serie storiche e
analisi di Fourier

Più dati storici, maggiore precisione



IL PREZZO DEL PETROLIO

Analisi storica inutile

Modelli predittivi non accurati

Machine learning riempie di dati inutili

Non si può predire il futuro di eventi
condizionati dal comportamento
umano

c'è sempre un "cigno nero"

Si può influenzare il comportamento
delle masse per ottenere un risultato
ma non con certezza

ASTRONOMIA SCIENZA ESATTA

I calcoli delle orbite dei pianeti erano accurati con il modello tolemaico, a base di deferenti, epicicli, equanti.

Il modello copernicano con orbite circolari non dava risultati accettabili.

A chi dovevano credere?

Ci accontentiamo dei risultati senza conoscere le cause?

BIG NUMBERS BIG PROBLEMS BIG ANSWERS

Il problema dei tre corpi non è risolvibile analiticamente

Nettuno è stato scoperto a tavolino

Wolf Cukier, 19 anni, stagista alla NASA, scopre un pianeta di un sistema stellare binario

These are the types of signals that algorithms really struggle with. The human eye is extremely good at finding patterns in data, especially non-periodic patterns like those we see in transits from these systems.

MICHAEL POLANYI

Nessuna conoscenza può essere interamente esplicita.

C'è sempre un aspetto sussidiario o implicito di un problema.

Il percorso conoscitivo si sviluppa attraverso l'interpretazione personale di una serie di indizi che orientano verso una consapevolezza che da implicita o tacita si fa esplicita.

OPINIONE

Concetto che una o più persone si formano riguardo a particolari fatti, fenomeni, manifestazioni, quando, mancando un criterio di certezza assoluta per giudicare della loro natura (o **delle loro cause**, delle loro qualità, ecc.), si propone un'interpretazione personale che si ritiene esatta e a cui si dà perciò il proprio assenso, **ammettendo tuttavia la possibilità di ingannarsi** nel giudicarla tale.

Treccani

TU CI CREDI ?

$$2 + 2 =$$

4

TU CI CREDI?

2691959

è un

numero primo

TU CI CREDI?

Il morso del mamba
nero è mortale



TU CI CREDI?

L'Italia è il Paese più
bello del mondo



GRAMELLINI

22 NOVEMBRE 2018

Ha ragione chi pensa, dice o scrive che la giovane cooperante milanese rapita in Kenya da una banda di somali avrebbe potuto soddisfare le sue smanie d'altruismo in qualche mensa nostrana della Caritas, invece di andare a rischiare la pelle in un villaggio sperduto nel cuore della foresta. [...]

Non riesco ad accettare gli attacchi feroci a qualcuno che si trova nelle grinfie dei banditi: **se tuo figlio è in pericolo di vita, il primo pensiero è di riportarlo a casa, ci sarà tempo dopo per fargli la ramanzina.**

GRAMELLINI IL GIORNO DOPO

Qualche furbacchione aveva preso l'incipit della rubrica - dove riconoscevo la logica di alcune argomentazioni contro la cooperante per arrivare nelle righe successive a rovesciarle - e me lo aveva attribuito. A quel punto è partito lo «shit storm». [...] I social hanno instaurato la dittatura dell'impulso, che porta a linciare prima di sapere e a sostituire la voglia di capire con quella di colpire. [...] E hanno talmente fretta di fartelo sapere da non accorgersi nemmeno che su Silvia tu la pensi come loro.

INTERVISTA AL DIRETTORE DE IL FOGLIO

Penso che i giornali hanno senso soltanto se rappresentano degli interessi, perché non raccontano la verità, ma un punto di vista. I giornali che dicono di voler raccontare la verità stanno imbrogliando i loro lettori, perché **la verità non esiste**. [...] Quel che conta è ammettere qual è il proprio punto di vista, la propria posizione, spiegare come si guarda il mondo e che idea se ne ha. Così, dicendo sinceramente da che parte stai, il tuo **lettore può fare la tara**.

CHI È CAUSA DEL SUO MAL

Anche se oggi è molto facile per chiunque recuperare l'originale delle parole del Papa, alla prova dei fatti ben pochi lo fanno, perché la maggioranza si fida ciecamente dei media, e soprattutto dei titoli strillati.

Lucetta Scaraffia, L'Osservatore Romano
2018-12-29

AFFIDABILITÀ

Autorevolezza

1. **Sono noti gli autori dei contenuti?**
2. Che reputazione hanno?
3. A chi sono affiliati?
4. Che importanza ha chi ha pubblicato il sito?

Rilevanza

1. Quanto è pertinente il contenuto rispetto al tema?
2. **Chi sono i destinatari previsti dagli autori?**
3. È appropriato il livello di approfondimento?
4. Esistono altre fonti sulla rete sullo stesso tema?

Accuratezza

1. Ci sono errori evidenti?
2. Ci sono prove verificabili delle affermazioni fatte?
3. Gli autori citano le fonti delle loro affermazioni?
4. **C'è la data di pubblicazione?**

Oggettività

1. Gli autori hanno lo scopo di convincere o informare?
2. Come influenza il contenuto questo scopo?
3. **Il testo contiene opinioni o fatti?**
4. Tentano di far passare per fatti quelle che sono opinioni?

CHARLES DARWIN

FAKE OR NOT FAKE?

La distinzione principale nei poteri mentali dei due sessi è costituita dal fatto che l'uomo giunge più avanti della donna, qualunque azione intraprenda, sia che essa richieda un pensiero profondo, o ragione, immaginazione, o semplicemente l'uso delle mani e dei sensi. Se vi fossero due elenchi di uomini e donne che eccelsero [...] non ci potrebbe essere confronto.

Possiamo anche concludere [...] che se **gli uomini sono in molte discipline decisamente superiori alle donne, il potere mentale medio dell'uomo è superiore a quello di queste ultime.**

SU INTERNET MA NON DI INTERNET

Il vantaggio di avere il testo completo di libri stampati disponibile in rete

Una fonte affidabile e verificabile in biblioteca

The screenshot shows a Google search for the phrase "decisamente superiori alle donne". The search results page from Google Books is displayed. On the left, the book "L'origine dell'uomo e la selezione sessuale" by Charles Darwin is shown with a red circle around the title and author information. The main content area displays the search results for the phrase, showing a snippet of text from the book. The snippet reads: "superi l'uomo nell'intuizione, nel rapido apprendimento e forse nell'imitazione, ma almeno alcune di tali facoltà sono caratteristiche delle razze inferiori e quindi di un più basso e ormai tramontato stadio di civiltà. La distinzione principale nei poteri mentali dei due sessi è costituita dal fatto che l'uomo giunge più avanti della donna, qualunque azione intraprenda, sia che essa richieda un pensiero profondo, o ragione, immaginazione, o semplicemente l'uso delle mani e dei sensi. Se vi fossero due elenchi di uomini e donne che eccelsero maggiormente nella poesia, nella pittura, scultura, musica (sia si tratti della composizione che dell'esecuzione), storia, scienza e filosofia, con una mezza dozzina di nomi sotto ciascuna disciplina, non ci potrebbe essere confronto. Possiamo anche concludere, con la legge della deviazione dalla media così ben illustrata da Galton nel suo libro *Hereditary Genius*, che se gli uomini sono in molte discipline decisamente superiori alle donne, il potere mentale medio dell'uomo è superiore a quello di queste ultime."

ACQUISTA EBOOK - 1,99 €

Trova questo libro nella versione stampata ▼

0 Recensioni
Scrivi recensione

L'origine dell'uomo e la selezione sessuale
Di Charles Darwin

"decisamente superiori" Vai

Risultato 1 di 1 in questo libro per "decisamente superiori alle donne"

superi l'uomo nell'intuizione, nel rapido apprendimento e forse nell'imitazione, ma almeno alcune di tali facoltà sono caratteristiche delle razze inferiori e quindi di un più basso e ormai tramontato stadio di civiltà.

La distinzione principale nei poteri mentali dei due sessi è costituita dal fatto che l'uomo giunge più avanti della donna, qualunque azione intraprenda, sia che essa richieda un pensiero profondo, o ragione, immaginazione, o semplicemente l'uso delle mani e dei sensi. Se vi fossero due elenchi di uomini e donne che eccelsero maggiormente nella poesia, nella pittura, scultura, musica (sia si tratti della composizione che dell'esecuzione), storia, scienza e filosofia, con una mezza dozzina di nomi sotto ciascuna disciplina, non ci potrebbe essere confronto. Possiamo anche concludere, con la legge della deviazione dalla media così ben illustrata da Galton nel suo libro *Hereditary Genius*, che se gli uomini sono in molte discipline decisamente superiori alle donne, il potere mentale medio dell'uomo è superiore a quello di queste ultime.

DEEP FAKE BIG PROBLEM

Far dire cose che non ha mai detto,
sincronizzando le labbra con le parole
mai pronunciate

Mettere la testa di una donna sul corpo
di un'altra

Il 98% dei deep fake è pornografia

Basta un solo deep fake di un politico
importante per avere reazioni di
portata globale: borsa, alleanze,
proteste

Prepariamoci al peggio

INGLESE NON INGLESISMO

Fact checking

verifica dei dati
pratica giornalistica

Debunking

smontare le bufale
inutile a convincere i complottisti

Echo chamber

crea un ambiente isolato
autoconferma delle proprie teorie

CONSEGUENZE SOCIALI NEGATIVE

Dal momento che non posso conoscere niente in modo certo, neppure posso prendere decisioni definitive.

Tutto può cambiare, tutto è provvisorio.

Ci arrendiamo?

FAHRENHEIT 451 SCRITTO NEL 1953

Le opere dei classici ridotte così da poter essere contenute in quindici minuti di programma radiofonico, poi riassunte ancora in modo da stare in una colonna a stampa, con un **tempo di lettura non superiore ai due minuti** [...] Nessuna meraviglia che i libri non si vendessero più, dicevano i critici. **Non è stato il governo a decidere**; non ci sono stati in origine editti, manifesti, censure, no! ma **la tecnologia**, lo sfruttamento delle masse e la pressione delle minoranze hanno **raggiunto lo scopo**.

LA FORMAZIONE ALLO SPIRITO CRITICO

sisri.it
disf.org
inters.org

Necessaria una scuola di razionalità,
per saper valutare con spirito critico,
basandoci su dati e informazioni
verificabili o affidabili

Consigli di lettura

Cathy O'Neil

Armi di distruzione matematica, 2017

Walter Quattrociocchi e Antonella Vicini

Liberi di Crederci

Informazione, Internet e post-verità, 2018

Alessandro Vespignani

L'algoritmo e l'oracolo, 2019