



Donata Columbro

Quando i dati  
discriminano

Bias e pregiudizi in grafici,  
statistiche e algoritmi

IA

Il Margine



Al giorno d'oggi, quando si vuole dare l'idea di essere obiettivi, perché ci si prefigge di essere incontestabili, può scappare di bocca un'espressione diventata d'uso comune: «Lo dicono i dati». Ma siamo certi che quei dati così allettanti, che confermano il nostro giudizio, non nascondano un «pregiudizio»?

Quando leggiamo un articolo online o apprezziamo un'infografica colorata bisognerebbe chiedersi prima di tutto: chi trae beneficio dalla raccolta, dall'analisi e dalla rappresentazione di quel dato, e chi può esserne invece discriminato?

Donata Columbro suggerisce come, universalizzando e standardizzando concetti come quello di «normalità», si sia in passato più escluso che incluso, creando una rappresentazione del mondo che ha eliminato le anomalie. Stabilire però chi è dentro o fuori le statistiche non è un atto neutrale, ma una scelta, e come tale andrebbe insegnata e indagata. Solo una persona consapevole che i dati sono costrutti umani e sociali può impedire che siano usati per discriminare, invece che per lottare contro le disuguaglianze.

## Donata Columbro

1984

Giornalista, divulgatrice e scrittrice. Collabora con «Internazionale» e «La Stampa», per cui cura la rubrica *Data Storie*. È docente a contratto per l'università IULM di Milano e per l'Università della Svizzera italiana di Lugano. Insegna data journalism al Master di giornalismo della LUISS di Roma e ogni mercoledì pubblica una newsletter su dati, algoritmi e tecnologia. È autrice dei libri *Ti spiego il dato* (2021) e *Dentro l'algoritmo* (2022).

Pluralità. Eterogeneità. Apertura.

Il Margine, progetto editoriale di Erickson, dà voce a punti di vista dissonanti e originali per metterli al centro del pensiero e del dibattito.

IN COPERTINA Immagine generata da una intelligenza artificiale, prompt di Chris Hayes, 2022

PROGETTO GRAFICO Bunker

€ 10,00



**Donata Columbro**

**Quando i dati  
discriminano**

Bias e pregiudizi in grafici,  
statistiche e algoritmi

**IA**

Quando i dati  
discriminano

*I dati non sono un'entità tecnica,  
ma culturale. [...]*

*Il dato non è mai dato. Semmai  
è sempre datur: da dare.*

SALVATORE IACONESI, *DataPunk*



## Introduzione

*Authenticity is a lie*, dice oggi l'app che mi invia l'oroscopo e che si vanta di usare dati della NASA, linguaggio naturale e sofisticati algoritmi per portarmi la saggezza degli astri ogni mattina. L'autenticità è una bugia. Mi sembra un'ottima premessa per l'introduzione di un libro in cui si sostiene che i dati neutri e oggettivi non esistono. Ma i dati possono essere autentici? Nella filosofia esistenzialista, spiegava il linguista Tullio De Mauro nell'ultima edizione del dizionario da lui curata, l'autenticità è l'esistenza in cui «il singolo ritrova il proprio più profondo se stesso, diversamente dalla condizione superficiale e impersonale in cui vive abitualmente». <sup>1</sup> Una ricerca dello stato

<sup>1</sup> <https://dizionario.internazionale.it/parola/autenticita>, tutti i link citati erano raggiungibili al 2 gennaio 2024.

## QUANDO I DATI DISCRIMINANO

di purezza che inseguiamo costantemente, e a cui diamo moltissimo valore. Tra i sinonimi di autenticità troviamo anche «genuinità», «schiettezza». E quello che non è «genuino», di conseguenza, è sofisticato o adulterato. Non è *naturale*.

Ogni volta che si tenta di valutare la purezza di un'azione, una decisione o persino del comportamento degli esseri umani si fa risalire l'origine alla «natura» di quell'atto, o di quella scelta. È naturale? Allora è buono, anzi è puro. Non contaminato, quindi giusto. E se non è naturale, anzi, è «contro natura»? Allora dobbiamo allontanarcene, evitarlo, criticarlo persino.

Come se la nostra vita, urbanizzata, motorizzata, plastificata, avesse ancora qualche rapporto vero, concreto e dimostrabile con il mondo pre-umanizzato ed esistesse qualche modo per le persone di non contaminare alcunché.

Il concetto di «naturale» è associato a un mondo in cui l'uomo (e la donna) non avevano ancora lasciato la loro impronta (distruttiva), sarebbe quindi preferibi-

## INTRODUZIONE

le a tutto ciò che è stato inquinato dalla nostra presenza, che, nel settore delle decisioni basate sulla quantificazione della realtà e dei comportamenti umani, presenta un problema di affidabilità e «imparzialità»: togliere l'elemento umano, volubile, pieno di pregiudizi, serve a identificare il dato come strumento neutro di conoscenza della realtà. Ma immaginare la purezza di un mondo naturale e pre-umano come l'unico davvero scevro dai problemi che derivano dall'impronta umana è inutile, oltre che dannoso.

Brigitte Vasallo, scrittrice, attivista e femminista spagnola, in un volume dedicato alle pratiche non monogame, sostiene che «[il] dibattito sull'ipotetica naturalità delle forme sociali viene sempre a consolidare lo stato delle cose: è sempre un argomento reazionario ed egemonico» (Vasallo, 2022, p. 49).

Con i dati tentiamo lo stesso approccio. Quando chiedo a una delle nuove intelligenze artificiali generative di definire l'espressione «dati grezzi», ChatGpT mi spiega che l'espressione, in informatica,

## QUANDO I DATI DISCRIMINANO

si riferisce a «dati che non hanno subito elaborazioni, manipolazioni o interpretazioni. Sono in un formato “puro”, direttamente dalla sorgente, senza essere stati modificati o trasformati».

Solo che «i dati grezzi sono un ossimoro», per citare il titolo di una raccolta di saggi a cura di Lisa Gitelman (2013), docente di cultura e media alla New York University. I dati grezzi, cioè ancora non elaborati dagli esseri umani, puri, naturali, non esistono. I dati sono il risultato di un'indagine, non la sua premessa. Non inciampiamo nei dati quando camminiamo per strada come faremmo con dei sassi o quando siamo in laboratorio a produrre ricerca scientifica, né quando entriamo in ufficio o svolgiamo qualsiasi altra attività. Siamo davanti a un costrutto umano, sociale, con tutte le conseguenze di tale assunto.

I dati sono cose che succedono, che esistono o si percepiscono e che qualcuno decide di osservare, contare e classificare, avevo scritto in *Ti spiego il dato*

## INTRODUZIONE

(2021, p. 14). Il fatto che io abbia usato il verbo «decidere» e poi inserito «osservare» prima di contare è il punto chiave per capire l'impossibilità di credere in una neutralità del dato. Non può esserci alcuna raccolta dati prima che avvenga una decisione che deriva dall'interesse nell'oggetto dell'indagine e che presuppone che si inizi a guardarlo con occhi diversi.

Eppure, la diffusione di espressioni come «data driven» (guidati dai dati) o «lo dicono i dati», che servono a portare autorevolezza alla tesi di chi le pronuncia, ci indicano che sui dati poggiamo un desiderio di oggettività che resta irraggiungibile. Vogliamo essere obiettivi perché cerchiamo di essere equi, imparziali, incorruttibili e incontestabili, e citando «i dati» siamo certi di andare in quella direzione. Invece qui dimostreremo che la strada verso l'equità non passa dalla neutralità, ma dalla consapevolezza della posizione che occupiamo nel mondo.

I dati non possono mai essere oggettivi o neutrali, perché sono influenzati

dal contesto in cui sono stati prodotti — non genericamente «raccolti» — e dalle scelte fatte fin da quando si è deciso di analizzare un certo fenomeno o una situazione. In più, la scelta di cosa includere dentro una categoria e cosa escludere può influenzare direttamente la loro rappresentazione e quindi l'intera narrazione che facciamo di quel fenomeno. Lo vedremo parlando di gap, di mancanze e buchi nelle statistiche, quali conseguenze hanno sulle decisioni politiche, sulla progettazione e sui prodotti delle intelligenze artificiali generative.

Indagheremo poi come si discrimina applicando regole matematiche alla società in modo coercitivo e repressivo, universalizzando e standardizzando concetti come quelli di «normalità» e «canone».

Quello che faremo con questo libro sarà approfondire, come se lo guardassimo al microscopio, cosa c'è dentro un dato, osservando ogni filamento, ogni granello, ogni dettaglio, per ricordarci che i dati restano sempre costrutti umani

## INTRODUZIONE

e sociali, situati e quindi calati nel contesto storico, geografico, economico in cui ci troviamo, momento per momento.