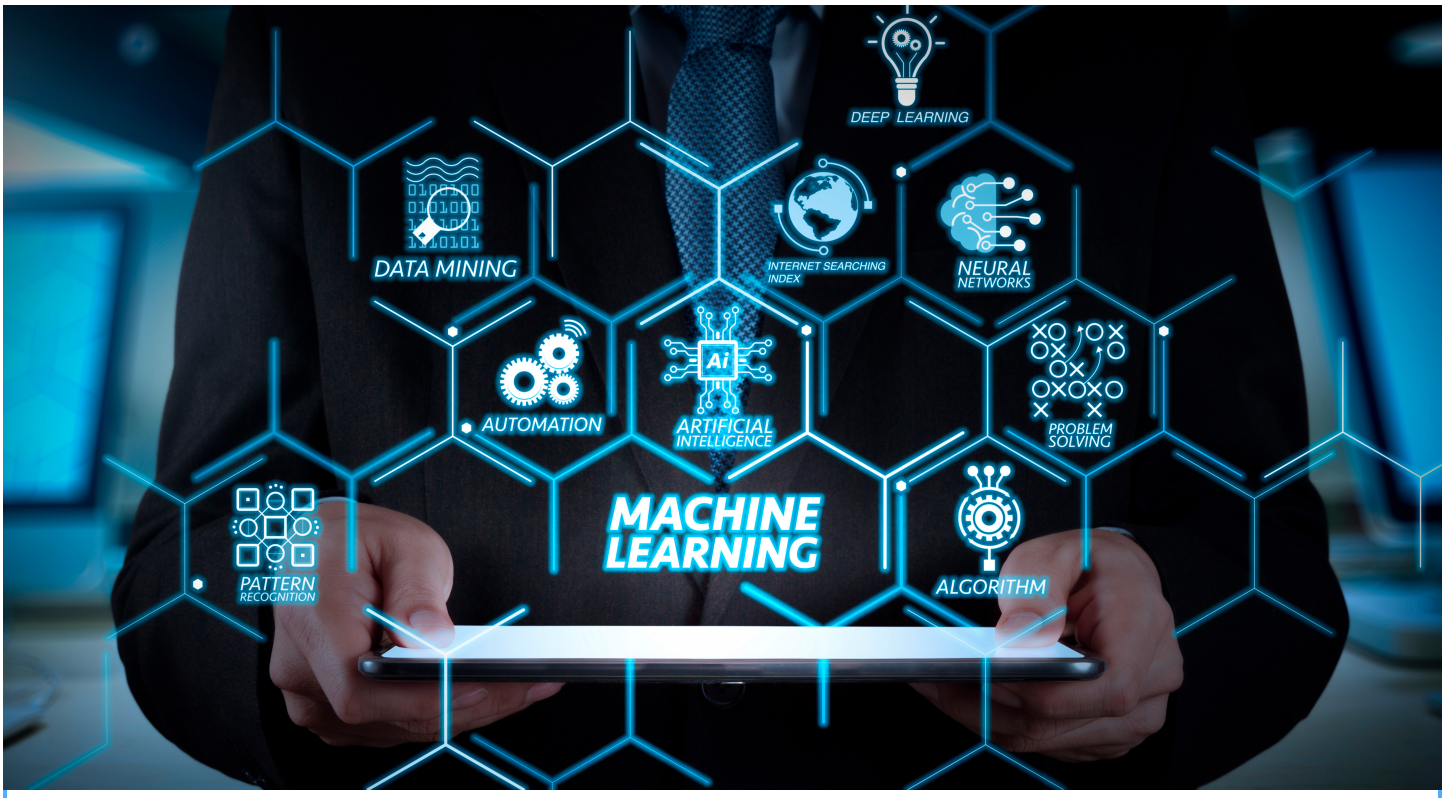
Oltre il codice: comprendere il Machine Learning con Google Teachable Machine (III)

DI THEFABLAB
19/03/2026

MACHINE LEARNING GOOGLE TEACHABLE MACHINE



Questo testo spiega come i dataset influenzano il comportamento dei modelli di machine learning e introduce il tema dei bias algoritmici attraverso Google Teachable Machine. Grazie ad attività pratiche, gli studenti possono scoprire come dati incompleti o poco vari possano portare l'Intelligenza Artificiale a commettere errori o sviluppare pregiudizi, imparando così l'importanza di creare dataset più equilibrati e accurati.

Autori

THEFABLAB

TheFablab è attivo dal 2014 nell'ambito dell'educazione tecnologica e della fabbricazione digitale, progettando percorsi formativi per i giovani e le comunità. Collabora a progetti nazionali con aziende, enti pubblici, scuole e istituzioni culturali per promuovere le competenze STEM, l'alfabetizzazione digitale e la cultura scientifica. Attraverso programmi educativi *hands-on*, contribuisce a rafforzare il capitale scientifico delle nuove generazioni e l'approccio critico verso le tecnologie emergenti, sostenendo l'empowerment, l'inclusione e il benessere educativo.

Nel machine learning i **dataset** di addestramento sono elementi fondamentali per garantire il corretto funzionamento dei modelli. Google Teachable Machine, uno strumento *no code* per avvicinare il pubblico al machine learning, permette di rendere evidente la centralità del dato nell'addestramento delle IA e di definire le caratteristiche dei dataset ideali.

Nel secondo episodio della serie dedicata a Google Teachable Machine abbiamo approfondito i modi in cui i parametri avanzati, dai periodi al tasso di apprendimento, sono in grado di influenzare la prestazione del modello. In questo ultimo articolo ci concentreremo invece su come le caratteristiche dei dataset di addestramento possono avere un impatto sul funzionamento dei modelli di machine learning, con un'attenzione particolare al concetto di **bias**.

Nelle scienze cognitive, si definisce **bias** una forma di distorsione sistematica nel modo in cui interpretiamo informazioni e prendiamo decisioni, basato su pregiudizi e scorciatoie cognitive. In informatica, i **bias algoritmici** sono descritti come la tendenza di un sistema di intelligenza artificiale a manifestare errori ricorrenti che producono risultati non equi.

Spesso l'origine dei bias risiede proprio nei dataset con cui sono stati addestrati i modelli di machine learning: i pregiudizi umani si riflettono nei dataset e li influenzano. Imparare a riconoscere i meccanismi che generano bias è quindi una competenza fondamentale per adottare un **approccio human-centered all'intelligenza artificiale**.

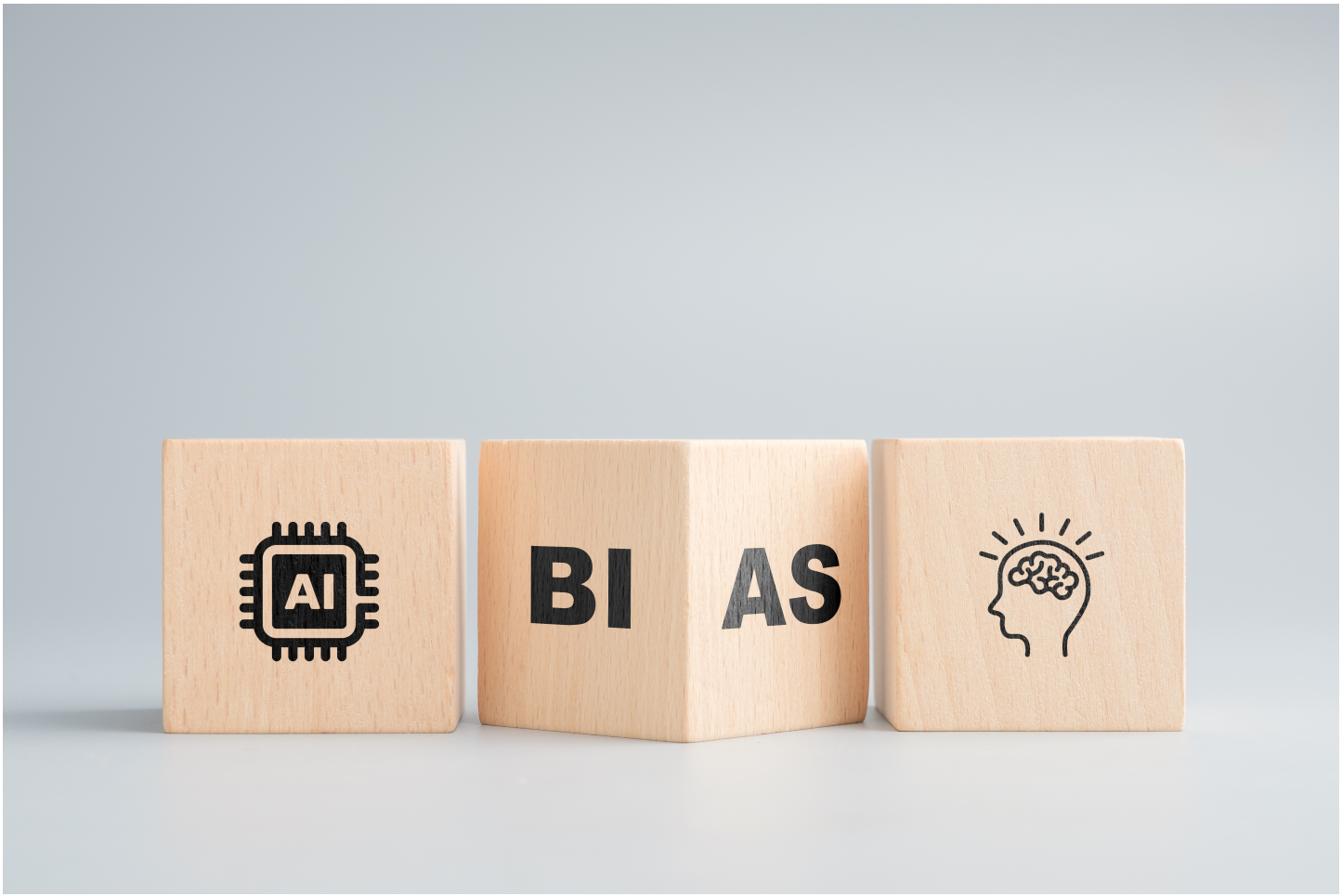
I BIAS ALGORITMICI

Esistono moltissime categorie di pregiudizi umani che potrebbero produrre bias algoritmici. Riportiamo di seguito alcuni esempi tra i più comuni e immediati da esplorare con uno strumento come Google Teachable Machine.

Bias di selezione. Questo bias compare quando i dati scelti per addestrare un algoritmo non rappresentano fedelmente la realtà statistica del mondo esterno. Una causa può essere la **sottorappresentazione**: se, per esempio, addestriamo un modello a riconoscere la categoria "Frutta", ma forniamo un dataset che contiene solo immagini di mele e pere, il modello non riconoscerà, per esempio, le fragole come frutti.
Un caso particolare di bias di selezione è il **bias di contesto**. Questo bias si verifica quando il modello associa un oggetto allo sfondo anziché al soggetto. Per esempio, se vogliamo addestrare un modello a riconoscere una torta e scattiamo diverse foto di torte in una cucina, il modello potrebbe finire per memorizzare l'aspetto della cucina, classificando come "Torta" anche una tazza o una pentola perché riconosce le piastrelle sullo sfondo.

Bias di segnalazione. Gli esseri umani tendono a documentare le circostanze particolarmente insolite e memorabili, mentre è più raro tenere traccia di quello che è ordinario. Consideriamo per esempio il caso delle recensioni ai prodotti: i clienti che torneranno sul sito di acquisto per lasciare una recensione saranno in media o molto soddisfatti o molto insoddisfatti, ci sarà quindi una maggiore quantità di recensioni estreme con toni molto accesi. Supponiamo di voler addestrare un algoritmo perché rilevi il livello generale di soddisfazione verso un prodotto a partire dalle recensioni: l'accuratezza dell'output sarà influenzata dalla polarizzazione dei dati causata dal bias di segnalazione, rendendo quindi il sistema meno affidabile.

Bias di conferma. È la tendenza umana a cercare e favorire informazioni che confermino le proprie convinzioni preesistenti. Questo bias può essere replicato anche dagli algoritmi: supponiamo per esempio che un'azienda addestri un algoritmo di machine learning per fare uno screening automatico dei curriculum, selezionando i più promettenti sulla base di dati storici associati ai migliori dipendenti. Se, storicamente, l'azienda ha messo nelle condizioni molti più uomini che donne di ricoprire posizioni di potere e raggiungere performance elevate, l'algoritmo tenderà a



ESPLORARE I BIAS ALGORITMICI CON GOOGLE TEACHABLE MACHINE: UN ESEMPIO DI ATTIVITÀ PRATICA

I bias algoritmici sono un argomento cruciale per trasmettere agli studenti l'importanza dei dataset e della loro accuratezza nella generazione degli output di ogni strumento di IA. Google Teachable Machine permette di strutturare attività immediate e accessibili per cogliere questi aspetti. Vi proponiamo di seguito un esempio di attività per introdurre alla vostra classe i **bias di contesto**.

Per prima cosa, create un progetto del tipo **Pose** su Google Teachable Machine. Questi progetti si basano su **PoseNet**, un modello pre-addestrato in grado di stimare la posa di una persona in tempo reale mappando punti critici del corpo umano, come le articolazioni. Suddividiamo l'attività in **quattro fasi**.

Creazione dei dataset. Create due classi all'interno del progetto: "**Lavoro**" e "**Riposo**". Nella classe "Lavoro", inserite immagini di studenti in piedi davanti alla lavagna, intenti a scrivere o a leggere. Nella classe "Riposo", addestrate il modello con immagini di studenti seduti su una sedia in pose rilassate.

Addestramento e stress test. Addestrate il modello e testate il suo funzionamento. Se riconosce correttamente le pose "Lavoro" e "Riposo" quando somigliano alle immagini che avete fornito nel dataset, provate a inscenare uno ***stress test* usando condizioni diverse**: cosa succede, per esempio, se mostrate al modello l'immagine di una persona seduta a terra o su un banco? Se il modello non riconosce queste pose come "Riposo", potreste essere di fronte a un bias di contesto: il modello non associa la classe alla posizione della persona seduta, ma al contesto (la sedia, lo sfondo, la situazione in cui sono state scattate le foto).

Discussione e correzione del bias dei dataset. Per correggere il modello è necessario condurre un nuovo addestramento, con **nuovi set di immagini**. Questa volta, prima di cominciare l'attività, discutete insieme per stabilire le caratteristiche che le immagini del tipo "Riposo" dovrebbero avere in modo da evitare i bias di contesto: per esempio, potreste fornire immagini di persone di diverso aspetto sedute su diversi tipi di supporto, usando sfondi diversi e diverse condizioni di illumina

Nuovo addestramento e test. Addestrate il modello con il nuovo dataset e testate la sua risposta. Ripetete la proce ino a quando, secondo voi, il bias di contesto sarà stato minimizzato.

Questo semplice esperimento mostra come la qualità e la varietà dei dataset influenza direttamente il comportamento di un modello di machine learning. Anche strumenti accessibili come Google Teachable Machine permettono quindi di comprendere

in modo concreto come nascono i bias algoritmici e perché **progettare con attenzione i dateset** sia una parte fondamentale dello sviluppo dell'intelligenza artificiale.