Gli algoritmi e le decisioni inevitabili

Remo Siza, | 28 settembre 2021

In molte nazioni europee sono stati introdotti algoritmi e modelli matematici di previsione dei comportamenti dei beneficiari di servizi di welfare, modelli predittivi sui nuclei familiari a maggior rischio di abuso e maltrattamento sui minori, di controllo e sorveglianza delle aree economicamente più deprivate delle città. Esperti delle Nazioni Unite hanno espresso preoccupazione per l'emergere del "*digital welfare state*", e rilevano che troppo spesso i veri motivi alla base di tali programmi sono il taglio della spesa sociale, l'istituzione di sistemi di sorveglianza governativa intrusivi e la volontà di generare maggiori profitti per le imprese private (Aston, 2019).

Nel Regno Unito e negli Stati Uniti si utilizza il PredPol che consente di prevedere in quale parte della città è più probabile che si verifichi un crimine e utilizzando strumenti come "*stop, question and frisk*" (ferma, interroga e perquisisci) si interviene in modo "chirurgico" sulle persone più a rischio, senza conoscere realmente chi abita il quartiere, creando categorie di rischio basati su condizioni geografiche (il quartiere di provenienza), sulla base dei comportamenti passati e della regolarità osservate in alcuni comportamenti ritenuti antisociali. L'algoritmo Syri (System Risk Indication) è stato utilizzato in Olanda al fine di valutare l'attitudine a commettere frodi o abusi di coloro che percepiscono sussidi o altre forme di assistenza pubblica. In Danimarca, il progetto Gladsaxe attribuiva un sistema di punteggi comportamentali utilizzando i dati raccolti sul nucleo familiare relativi al reddito, all'istruzione, alla stabilità del nucleo familiare, al quartiere di residenza e a vari aspetti comportamentali. Un punteggio anomalo allertava i servizi sociali, che intervenivano per prevenire abusi e procedere all'affido.

Il Rapporto dell'Algorithm Watch (2020) rileva che molti di questi algoritmi hanno prodotto risultati molto negativi sulle condizioni di vita della popolazione. Nella costruzione di profili di rischio capaci di prevedere la probabilità di un individuo di truffare lo Stato, sulla base di dati precedentemente raccolti e analizzati, le discriminazioni sono spesso molto elevate. Il Rapporto rileva che se si amplia lo sguardo allo stato attuale dei sistemi di ADM (*automated decision-making*) in Europa, si scopre che gli esempi positivi, che arrecano chiari benefici, sono rari, e che la stragrande maggioranza degli usi tende ad aumentare i rischi per i cittadini, invece di essere loro d'aiuto (*Algorithm Watch*, 2020, p. 5). Molti dei sistemi indicati sono stati bloccati da movimenti, tribunali, perché ritenuti gravemente discriminatori nei confronti delle minoranze etniche e delle persone che abitano i quartieri più degradati.

Questi algoritmi si dicono predittivi perché si basano costantemente sull'ipotesi che il nostro futuro sarà una riproduzione del nostro passato. Non è più necessario conoscere gli individui e avviare ricerche empiriche, luoghi di coinvolgimento e di partecipazione. L'individuo viene svelato dalle sue stesse tracce, non ha interiorità, ha solo comportamenti, eventi, segnali. L'ombra che porta la traccia dei comportamenti delle persone è sufficiente a nutrire i calcoli e a riconoscere i comportamenti a rischio (Cardon, 2015). Per osservare gli individui non serve quantificare le categorie stabili (classi d'età, classi professionali…) ma le tracce, i segnali, i comportamenti ricorrenti sui quali costruire distinti profili che consentiranno di individuare i comportamenti statisticamente più probabili, i percorsi di cura e i premi assicurativi. In molti casi, però, i profili coincidono con i più tradizionali stereotipi: le famiglie povere e i giovani emigrati tendono a compiere maggiori reati, le madri sole tendono ad utilizzare in modo eccessivo le prestazioni di welfare, i reati e i crimini si concentrano in determinati quartieri.

I modelli algoritmi di previsione del rischio nonostante la loro reputazione di imparzialità sono il riflesso di obiettivi e ideologie (O'Neil, 2016). Non esiste mai un unico algoritmo per risolvere un determinato problema, ma una moltitudine di algoritmi possibili e alcuni discriminano meno di altri. Certamente si può lavorare in modo interattivo rendendo il modello e l'algoritmo più complesso per tener conto delle mancate regolarità che osserviamo. Con processi di "apprendimento supervisionato" si possono modificare il modello iniziale, eliminare discriminazioni e categorizzazioni riduttive.

È necessario comunque tenere presente che un algoritmo o un modello matematico contiene inevitabilmente dei *bias*. Possiamo distinguere due tipi principali di *bias* (Jean, 2021). Alcuni sono cognitivi e dipendono dalla visione del mondo degli individui che lo sviluppano e sono legati a fattori sociali o culturali, alla loro percezione di ciò che accade. Altri *bias*, sono riferibili ai dati che utilizziamo e immettiamo nel nostro modello matematico che possono contenere errori (casuali, errori di misurazione, di campionamento, di concettualizzazione di un fenomeno…) che spesso non conosciamo. Spesso i dati che abbiamo a disposizione ci faranno sviluppare un algoritmo profondamente distorto che danneggerà la vita di molte persone. Comunque, non si possono eliminare tutti i *bias* e la loro persistenza deve costringerci a interrogarci sulle nostre certezze, per evitare conseguenze a volte disastrose, discriminazioni tecnologiche, quindi sociali o razziali (Jean, 2021).

Gli algoritmi non sono “neutrali”, non possono essere scambiati per tecnologie oggettive. Replicano, invece, gli assunti e le credenze di chi decide di adottarli e programmarli. È sempre un umano dunque, non “gli algoritmi” o i sistemi di ADM, a essere responsabile sia delle buone che delle cattive decisioni algoritmiche (Algorithm Watch, 2020). In alcune nazioni è stato istituito il diritto a ricevere una spiegazione da parte di un essere umano per l'output di un algoritmo e il diritto alla rettifica di una decisione algoritmica (Kaun, 2021). Per i profili di rischio, che riguardano le persone che abitano le aree più povere o determinati gruppi sociali sono spesso movimenti collettivi e associazioni che intervengono a loro favore. Insomma, i rischi di discriminazione devono essere prevenuti e mitigati con un'attenzione particolare per i gruppi che hanno un rischio maggiore (persone economicamente svantaggiate, membri della comunità LGBTI, persone con disabilità, minoranze etniche) (Council of Europe, 2019).

Rispetto ad una crescente differenziazione sociale, ad una irregolarità di molti comportamenti individuali e collettivi cresce l’esigenza di prevedibilità, di identificare una relazione causale tra la risposta che auspichiamo (regolarità dei comportamenti, inserimento lavorativo) e lo stimolo che attiviamo (beneficio di welfare, programmi preventivi, azioni comunicative) eliminando la contingenza delle reazioni umane. In realtà, noi non siamo soltanto la somma imprecisa e incompleta dei nostri comportamenti (Cardon, 2015) e con gli algoritmi non costruiamo certezze, ma solo approssimazioni talvolta utili, ma che comunque attendono una verifica empirica. L’essere umano funziona con una logica integrativa non è semplicemente un assemblaggio di organi così come non è riducibile ad una somma di informazioni. La distinzione essenziale da cui partire è quella tra "funzionare" ed "esistere" (Benasayag, 2019).

In molte parti d’Europa i sistemi di welfare cambiano per affrontare la crisi demografica, promuovere la conciliazione vita-lavoro, la sostenibilità del sistema sanitario. Cambiano forse ancora più profondamente per affrontare i comportamenti irregolari e imprevedibili delle famiglie a basso reddito, dei beneficiari di sussidi economici, dei disoccupati a lungo termine, degli abitanti delle periferie urbane. Queste tendenze modificano le relazioni di cura, le sue finalità, i valori e le relazioni che producono, gli strumenti e le risorse utilizzate nelle pratiche professionali quotidiane. Le due direzioni non si intrecciano più, non sono più convergenti e rischiano di creare due sistemi molto differenti nella stessa nazione. Il primo sistema, accogliente e reattivo, è rivolto alle persone che operano fattivamente nel sistema lavorativo. Il secondo, come ci ricordano molte ricerche, è tendenzialmente punitivo, condizionale, è governato da decisioni automatiche ed è rivolto quasi esclusivamente alle persone che gravano sui sistemi di welfare perché non hanno redditi da lavoro adeguati o hanno scarse possibilità di trovare una qualsiasi occupazione.

Bibliografia

- Algorithm Watch (2020) Automating Society, Berlin (edizione italiana)
- Alston, P. (2019) Report of the Special rapporteur on extreme poverty and human rights.
- Benasayag, M. (2019) La tirannia dell’algoritmo, Vita e Pensiero, Milano.
- Cardon, D. (2015), Che cosa sognano gli algoritmi, trad. it. Mondadori Università, Milano.
- Council of Europe-Commissioner for Human Rights (2019) Unboxing Artificial Intelligence: 10 steps to protect Human Rights, Bruxelles
- Jean, A. (2021) Nel paese degli algoritmi, trad. it. Neri Pozza Editore, Vicenza.
- Kaun A. (2021) Suing the algorithm: the mundanization of automated decision-making in public services through litigation, Information, Communication and Society (online).
- O’Neil (2016) Armi di distruzione matematica, trad. it. Bompiani, Milano.