



Le conseguenze nefaste dell'autonomia degli algoritmi

La vita autonoma degli algoritmi tra opportunità straordinarie e rischi inattesi

La vita quotidiana ai tempi del XXI secolo è sicuramente dominata dall'uso delle tecnologie digitali. Le usiamo per lavorare (software professionali), per divertirci (piattaforme di intrattenimento come Netflix), per gestire le relazioni con altre persone (social media, da Facebook a Instagram). Da un punto di vista generale, questa situazione non è particolarmente diversa da altri periodi storici: sicuramente il boom è avvenuto con la rivoluzione industriale, quando le macchine sono diventate parte integrante del panorama storico del XVII secolo. Da lì in poi la tecnologia ha avuto un ruolo di primo piano ed è penetrata in ogni aspetto del mondo moderno.

Di cosa hanno bisogno gli algoritmi?

A ben vedere, in questi primi anni del duemila la tecnologia – che da analogica è diventata digitale – è un po' diversa da altri sistemi tecnologici che hanno caratterizzato le epoche passate e questa diversità sta negli algoritmi. Gli algoritmi sono procedure informatiche che hanno lo scopo di guidare il software nello svolgimento delle sue attività. La lavatrice che modifica il suo ciclo di lavaggio quando deve lavare capi delicati, il navigatore satellitare che ci dice qual è la via più breve del nostro viaggio, la piattaforma social

che ci mostra i contenuti prodotti e condivisi dalla nostra rete di contatti: sono tutti esempi di funzioni svolte da algoritmi, procedure digitali che consentono al software di funzionare nel miglior modo possibile. C'è una caratteristica che rende gli algoritmi così strategici nella vita contemporanea: gli algoritmi sono autonomi, ovvero non hanno bisogno di nient'altro che una serie di operazioni da compiere e un insieme di dati in base ai quali svolgere le funzioni per cui sono stati progettati. Detto in maniera molto semplificata: gli algoritmi decidono in autonomia come interagire a seconda del tipo di informazioni che raccolgono e che li caratterizzano. La lavatrice decide quale ciclo di lavaggio utilizzare a seconda del capo che deve lavare, il navigatore decide quale strada percorrere a seconda delle strade e del traffico, la piattaforma social decide quale contenuto mostrarci a seconda della nostra rete di contatti. In pratica gli algoritmi sono strumenti decisionali autonomi che svolgono il loro compito seguendo specifiche regole.

Il governo degli algoritmi

Quali sono queste regole? La risposta a questa domanda rappresenta il punto di forza e il punto di debolezza degli algoritmi. Le regole che seguono gli algoritmi alcune volte



Davide Bennato

Professore di sociologia dei media digitali

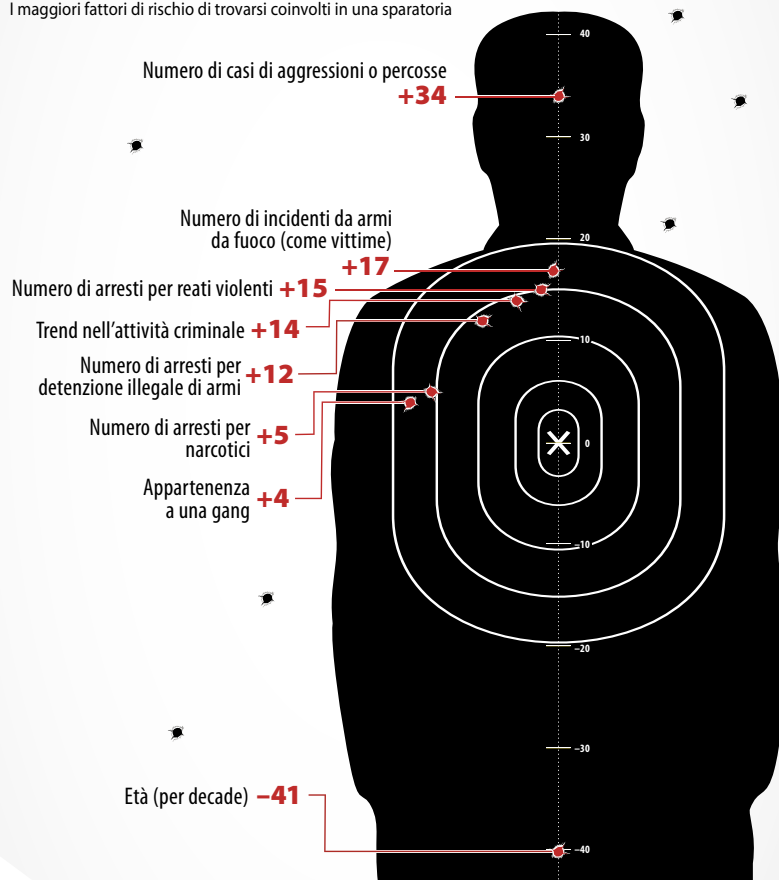
Università di Catania

sono regole matematiche se il settore in cui operano è stato modellizzato: è un principio matematico in base al quale il navigatore valuta qual è la strada più breve fra due punti geografici. Ma quando il settore non è stato modellizzato con strumenti matematici, spesso le regole seguite dagli algoritmi sono quelle del suo programmatore e, in quanto tali, non sono regole scientificamente fondate, ma sono principi arbitrari e – spesso – soggettivi¹. La piattaforma social che ci mostra i contenuti della nostra rete di contatti parte dal presupposto che noi vogliamo vedere i contenuti dei profili con cui interagiamo più spesso. Ma questa affermazione è vera fino a un certo punto: io di alcune persone voglio vedere i contenuti perché mi interessano le persone; di altre persone voglio vedere i contenuti perché mi interessano i contenuti. Ma questo l'algoritmo non lo sa. Per l'algoritmo vale solo l'interazione: se io non interagisco con un profilo, non importa che questo profilo condivida contenuti interessantissimi, a poco a poco questa persona smetterà di essere visibile nella mia rete di contatti e io lo dimenticherò. Chi ha deciso tutto questo? L'algoritmo di ottimizzazione dei contenuti del social: in una parola, l'ha deciso Facebook. In pratica Facebook non è uno strumento matematico di visualizzazione dei contenuti della nostra rete di contatti, bensì uno strumento ideologico di condivisione di contenuti dei profili in cui l'interazione è più frequente. Questo accade perché per Facebook "amico" vuol dire "profilo con

Spesso le regole che gli algoritmi seguono sono quelle del suo programmatore e in quanto tali non sono regole scientificamente fondate, ma sono principi arbitrari e – spesso – soggettivi.

L'algoritmo detective della polizia di Chicago

I maggiori fattori di rischio di trovarsi coinvolti in una sparatoria



Modificato da "Inside the algorithm that tries to predict gun violence in Chicago", *The New York Times*, 13 giugno 2017.

Ragionevoli dubbi.

Per individuare le aree a più alta probabilità di sparatorie la polizia di Chicago utilizza l'algoritmo Strategic Subject List che misura i fattori di rischio. Tuttavia le analisi sono discordanti rispetto alla percezione comune e ad alcuni commenti della polizia: infatti, essere stati già coinvolti in una sparatoria è un fattore di rischio più importante di essere vittima o autore di una sparatoria (+34) e dell'affiliazione ad una gang (+4). Per valutare correttamente l'affidabilità di Strategic Subject List, sarebbe necessario conoscere i dettagli con cui vengono calcolate le probabilità di rischio, informazioni però non note perché l'algoritmo si basa su una tecnologia proprietaria. Le valutazioni del rischio qui a fianco sono state elaborate da un lavoro del *New York Times* sulla base dei dati resi disponibili dalla polizia di Chicago: attraverso una semplice analisi statistica è deducibile la base logica dell'algoritmo. Il problema nasce dall'uso di uno strumento tecnologico che guida le decisioni su come dovrebbero agire le forze di polizia, senza poter valutare la bontà della strategia decisionale adottata.

cui interagiamo più spesso" e questa definizione di amicizia si applica solo in alcuni specifici casi alle persone con cui condividiamo un rapporto. Se questo vale per una cosa così semplice come i contenuti che vediamo su Facebook, questo può valere anche per altri casi? Assolutamente sì.

Le conseguenze (in)attese

Molti studiosi stanno cercando di capire quali sono le conseguenze reali di algoritmi progettati male, ideologici o semplicemente dalle conseguenze impreviste². Un libro recente che affronta questo tema è *Armi di distruzione matematica* di Cathy O'Neill, in cui l'autrice, da studiosa di big data e computer science, propone uno strumento di valutazione a quattro livelli per capire le conseguenze inattese, spesso nefaste degli algoritmi³.

Il primo livello potremmo definirlo "conseguenze negative come proprietà emergenti", e avviene quando gli algoritmi manifestano errori in maniera non intenzionale a causa della loro programmazione. Caso emblematico quello di Google AdWords che suggeriva inserzioni pubblicitarie di consulenza legale se gli utenti che cercavano informazioni erano persone di colore, ipotizzando così che, se sei nero e cerchi informazioni legali, probabilmente hai problemi con la giustizia⁴. Gli algoritmi del secondo livello hanno conseguenze nefaste a causa di una programmazione superficiale. È questo il caso del sistema di autotagging della app di Google in cui per la foto di due persone di colore veniva suggerito

il tag "gorilla". Al terzo livello ci sono gli algoritmi sicuramente antipatici ma comunque legali: è il caso di PredPol, il software in dotazione della polizia di Los Angeles in grado di prevedere i reati sulla base di una analisi dei big data dei crimini precedenti, in puro stile *Minority Report*⁵. Infine, all'ultimo livello, ci sono gli algoritmi programmati con scopi intenzionalmente nefasti, spesso legati al controllo e alla sorveglianza, come il progetto Sesame Credit del governo cinese, un sistema in cui, a partire dalle piattaforme digitali come WeChat o Baidu, si raccolgono i dati sui cittadini che producono un punteggio su quanto le persone siano dei bravi cittadini. Esiste poi tutta una famiglia di conseguenze nefaste legate al fatto che gli algoritmi basati sul machine learning – termine contemporaneo che definisce l'intelligenza artificiale – sono in grado di apprendere dal contesto in cui si trovano.

Fuori controllo

Ma cosa succede quando il loro contesto di riferimento è palesemente negativo? La risposta a questa domanda l'hanno data rapidamente i programmatori della Microsoft quando in tutta fretta hanno dovuto terminare il loro progetto di chatbot perché la situazione era diventata piuttosto ingestibile. In pratica, per testare il proprio sistema di intelligenza artificiale la Microsoft ha messo online su Twitter Tay, un chatbot – ovvero un software automatico in grado di interagire – che imparava dalle domande che venivano poste dagli

altri utenti. Forse per scherzare, forse per semplice trolling, gli utenti cominciarono a fare domande sempre più politicamente scorrette, fino a far diventare Tay la prima intelligenza artificiale razzista, negazionista dell'Olocausto, sostenitrice dell'uso politico del genocidio e in genere aggressiva con ogni forma di minoranza⁶.

Questa è una delle palesi conseguenze negative delle tecnologie autonome: spesso evolvono verso forme di autonomia che nessuno vorrebbe. In sintesi, bisogna prendere sempre più sul serio l'idea che gli algoritmi non sono solo delle tecnologie, ma sono dei soggetti autonomi che svolgono un ruolo attivo nella società in cui viviamo e questo ruolo ha delle conseguenze sulla società e sugli individui che non sempre sono positive o auspicabili. Poco importa se per errori di programmazione, proprietà emergenti, decisioni eticamente discutibili o interazione in ambienti socialmente e ideologicamente violenti, gli algoritmi sono ormai elementi alla base della vita quotidiana del XXI secolo e bisogna cominciare a studiare le loro conseguenze, soprattutto se inattese.

Tay ha impiegato solo 16 ore per diventare un software dotato di un linguaggio violento e aggressivo. Cosa sarebbe accaduto se non fosse stato semplicemente un sistema per chiacchierare ma se – poniamo – fosse stato il sistema operativo di un robot come si stanno sviluppando negli ultimi tempi? Chi vorrebbe interagire con un tecnologia autonoma, antropomorfa e razzista? ❏

1. Bennato D. Il computer come macroscopio. Big data e approccio computazionale per comprendere i cambiamenti sociali e culturali. Milano: Franco Angeli, 2015.
2. Cardon D. Che cosa sognano gli algoritmi. Le nostre vite al tempo dei big data. Milano: Mondadori, 2016.
3. O'Neill C. Armi di distruzione matematica. Milano: Bompiani, 2017.
4. Racism is poisoning online Ad delivery, says Harvard professor. MIT Technology Review, 4 febbraio 2013.
5. Berg N. Predicting crime, LAPD-style. *The Guardian*, 25 giugno 2014.
6. Bright P. Tay, the neo-Nazi millennial chatbot, gets autopsied. *Ars Technica*, 26 marzo 2016.