Dianora Poletti, ordinario di Diritto privato e di Diritto dell'informatica, Direttrice del Centro di ricerca interdipartimentale in Diritto e tecnologie di frontiera dell'Università di Pisa (dianora.poletti@unipi.it)

Decisioni algoritmiche e diritto alla comprensione *(bozza della relazione presentata il 16/12/21)*

1. Di fronte alle decisioni che riguardano la vita di tutti noi, assunte sempre più di frequente con l'impiego degli algoritmi, appare decisivo analizzare quali strumenti siano in essere per comprendere il "pensiero" sconosciuto degli algoritmi e per verificare come "ragionano" gli stessi. Al riguardo si parla di trasparenza algoritmica e si usano, con una certa qual indifferenza, termini quali conoscibilità, spiegabilità, comprensibilità, leggibilità degli algoritmi, ma la materia va maneggiata con maggiore precisione.

Anzitutto, quando si affronta questo tema è opportuno chiarire a quali algoritmi ci si riferisce: un conto è parlare dei c.d. sistemi esperti, in grado di esibire i passaggi logici connessi alle decisioni assunte ("glass box"), un conto di algoritmi che autoapprendono, nei quali le quantità enormi di dati fornite per l'addestramento della macchina genera da parte del sistema correlazioni nascoste, che superano la mera inferenza logica da un insieme di dati e regole immessi dal programmatore ("black box"). Nell'ambito dell'insieme di sistemi di *Machine Learning* si staglia poi lo specifico settore applicativo del *Deep Learning*, che pone altri problemi, per la complessità dei modelli matematici utilizzati (cd. "reti neurali" artificiali).

Soccorre a mettere ordine al riguardo (quando l'algoritmo impieghi dati personali) il GDPR, che è stato tanto criticato per il suo essere nato già "vecchio" e per la sua incapacità se non di affrontare almeno di anticipare le grandi sfide poste dall'AI, ma che ci ha in realtà consegnato alcuni principi importanti: la non esclusività della decisione automatizzata; la comprensibilità della logica utilizzata per adottarla; il divieto di discriminazione derivante dalla profilazione impiegata per analizzare aspetti rilevanti della vita umana (ma anche per prevedere comportamenti, come recita la definizione dell'art. 4, che ben si adatta alla capacità predittiva degli algoritmi).

2. Questi principi valgono a declinare, nelle sue diverse articolazioni, il principio della trasparenza algoritmica, che in prima approssimazione può essere definito come l'obbligo, posto in capo ai soggetti che adottano decisioni con l'impiego di sistemi automatizzati di trattamento dei dati, di rilasciare ai destinatari una spiegazione comprensibile delle procedure utilizzate e di motivare le decisioni assunte sulla loro scorta.

Il primo profilo riguarda il diritto ad ottenere informazioni (recitano gli artt. 13 – informativa - e 15 – diritto di accesso del GDPR) circa "l'esistenza di un processo decisionale automatizzato". Questo aspetto corrisponde alla conoscibilità dell'impiego dell'algoritmo, cioè al diritto dell'interessato ad essere informato circa la presenza di un processo decisionale automatizzato a cui sono sottoposti i propri dati personali (una sorta, come da qualcuno è stata definito, di diritto alla "sincerità algoritmica").

Il secondo aspetto concerne il diritto di apprendere il grado di rilevanza assunto dall'algoritmo nel processo decisionale e corrisponde a ciò che potrebbe definirsi "trasparenza

sull'influenza" dell'algoritmo, rilevante in specie per valutare se il trattamento automatizzato rientri o meno nell'ambito applicativo dell'art. 22 del GDPR, il quale censura la creazione di una decisione realizzata tramite trattamenti completamente automatizzati dei dati personali. La norma impedisce soltanto l'automatizzazione integrale, postulando la necessità di un intervento umano nelle decisioni di questo tipo, ma nulla chiarisce sul grado di automatizzazione consentito nei singoli casi, oltre a soffrire di talune eccezioni (tra cui il consenso dell'interessato) e, per inciso, non sempre la presenza di una decisione interamente automatizzata è sufficiente a fare scattare il divieto posto dalla stessa, posto che occorre che la decisione algoritmica produca "effetti giuridici" sul soggetto o "incida in modo analogo significativamente sulla sua persona".

La terza articolazione si appunta sul diritto dell'interessato ad ottenere dal titolare del trattamento "informazioni significative sulla logica utilizzata" all'interno del processo decisionale automatizzato. Si tratta del c.d. diritto alla comprensibilità o alla spiegazione (algoritmica), del quale molto si è discusso, posto che la trasparenza garantita da *standards* pur elevati di informazione potrebbe essere insufficiente, se le informazioni non sono comprese dal destinatario.

Parte della dottrina ha cercato di ipotizzare una diversa conformazione della comprensibilità, proponendo il concetto di leggibilità, vale a dire "la capacità degli individui di comprendere autonomamente i dati e gli algoritmi di analisi, con una comprensione concreta dei metodi e dei dati utilizzati". La leggibilità, in altre parole, dovrebbe combinare la comprensibilità del funzionamento dell'algoritmo con la trasparenza rispetto alla sua utilizzazione. La casistica giurisprudenziale domestica è ancora abbastanza scarna, ma registra importante posizioni del Consiglio di Stato (8/0472019, n. 2270), che ha postulato un principio di "conoscibilità" dell'algoritmo inteso come "declinazione rafforzata del principio di trasparenza, che implica anche quello della piena conoscibilità di una regola espressa in un linguaggio differente da quello giuridico" e della Cassazione (25/05/2021, n. 14381), che ha legato la mancata trasparenza dell'algoritmo alla valida prestazione del consenso al trattamento di dati personali, sancendo che "il requisito di consapevolezza non può considerarsi soddisfatto ove lo schema esecutivo dell'algoritmo e gli elementi di cui si compone restino ignoti o non conoscibili da parte degli interessati".

L'ultima declinazione – le "conseguenze previste di tale trattamento per l'interessato" – si lega al rischio di produrre discriminazioni o distorsioni legate alla presenza di bias. I quali possono avere diverse fonti: può trattarsi di bias che si annidano nella programmazione algoritmica, di bias che si annidano nei dati di addestramento dell'algoritmo, di bias personali dei raccoglitori di dati (data gatherers) o anche di bias culturali o ambientali traslati non intenzionalmente all'interno del processo di raccolta dei dati. Tuttavia, di fronte al ricorso a sistemi adattivi e dinamici guidati dai dati, che sono in grado di mutare le loro risposte anche in base ai cambiamenti nell'ambiente, è difficile identificare dove si collochino i possibili bias in una regola di decisione che non è più predeterminata, potendo produrre risultati diversi per ogni istanza della sua esecuzione (ossia per ogni caso che gestisce).

3. Nella Rete, quasi più che in altri contesti, il problema dei diritti (e dei nuovi diritti, come quello che si va elaborando sul punto) è quello della loro effettività. Il diritto alla spiegazione

deve fare i conti con opacità che possono derivare anzitutto dalla presenza di diritti di privativa (dal diritto di autore ai segreti industriali e aziendali), con la conseguenza che legittimi interessi anche di natura economica possono imporre, rispetto all'accesso individuale alle informazioni rilevanti, una delicata operazione di bilanciamento.

Ma le opacità possono essere "intrinseche" al tipo di algoritmi impiegati. Se il GDPR parla di logica del processo automatizzato, non tutti gli algoritmi ne presentano una. Nella gran parte degli algoritmi "di nuova generazione" i criteri di inferenza non sono facilmente riproducibili e soprattutto comprensibili. Questo è particolarmente vero per le reti neurali, che basano le loro predizioni su pattern nascosti all'occhio del programmatore, inferiti direttamente dai dati.

Queste considerazioni impongono di interrogarsi se "esist[a] sempre una logica comprensibile negli algoritmi". La Risoluzione europea in materia di robotica e intelligenza artificiale approvata del febbraio 2019, nel sottolineare "l'importanza della spiegabilità dei risultati, dei processi e dei valori dei sistemi dell'IA, in modo da renderli comprensibili per un pubblico non tecnico e fornire a quest'ultimo informazioni significative, condizione necessaria per valutare l'equità e conquistare la fiducia", osserva in maniera significativa che la divulgazione del codice informatico "non risolverà di per sé la questione della trasparenza dell'Intelligenza artificiale, in quanto […] non spiegherebbe il processo di apprendimento automatico".

4. La sorte del principio della trasparenza algoritmica nella normativa europea *in fieri* appare altalenante. La soluzione proposta a livello europeo per le grandi piattaforme online, con il *Digital Services Act* (affiancato dal *Digital Markets Act*), appare in maniera promettente incentrata su una serie di principi, tra i quali la trasparenza, l'accessibilità ai dati e agli algoritmi, le informazioni complete agli utenti, l'autonomia nella scelta del grado di profilazione. E tuttavia, la proposta di regolazione dell'AI presentata il 21 aprile 2021 sembra incidere su tale questione in maniera recessiva, prevedendo, nella categoria della AI a "rischio limitato" (ma che include applicazioni di AI non così innocue), solo obblighi di trasparenza molto vaghi, soddisfatti da una semplice notifica ai consumatori/cittadini riguardo al fatto che un sistema AI sia operativo in quel contesto.

5. Per l'impiego degli algoritmi si invoca sempre più spesso un principio di trasparenza by design, che (sull'onda del percorso tracciato dal principio di privacy by design) assicuri che la loro progettazione ab origine sia strutturata tenendo presente l'interpretabilità. La trasparenza dovrà riguardare non solo il codice sorgente, ma anche i dati e il processo decisionale automatizzato.

Questo principio non sarà risolutivo dei complessi problemi che il loro uso produce ma aiuterà al loro superamento, anche se ha bisogno di importanti supporti: una attiva sorveglianza umana lungo tutto il percorso e soprattutto l'intreccio con il fondamentale principio di human right by design. Solo così si faciliterà il cammino verso un uso degli algoritmi che si potrebbe definire, con un termine oggi molto a la page, più sostenibile.