“Automaticamente illegali”: una proposta per i sistemi di intelligenza artificiale

Frank Pasquale e Gianclaudio Malgieri hanno recentemente proposto di considerare illegali i sistemi di intelligenza artificiale fino a prova contraria. Alle aziende l’onere di dimostrare che la loro tecnologia non è discriminatoria, non è manipolatoria, non è iniqua, non è inaccurata e non è illegittima nelle sue basi giuridiche e nei suoi scopi.

Daniela Tafani

I sistemi di intelligenza artificiale basati sull’apprendimento automatico (*machine learning*) sono utilizzati per ottenere classificazioni o produrre decisioni che hanno effetti rilevanti sulle vite delle persone, a una velocità e su una scala che non consentono un controllo umano significativo. Nei sistemi utilizzati per il riconoscimento facciale, ad esempio, o lo screening medico di immagini di tessuti umani o le decisioni di concedere o rifiutare un prestito, la quantità dei dati di partenza e la potenza di calcolo richiesta per la loro elaborazione fanno sì che la presenza di un essere umano nel processo (*human in the loop*) non sia in grado di fornire alcun controllo rilevante. Qualora un’attività sia automatizzata e si deleghi tuttavia a una persona il compito di intervenire, con prontezza fulminea, nei casi di emergenza, il ruolo dell’essere umano non può essere che quello di capro espiatorio, come pare sia stato previsto nei sistemi Tesla per le auto a guida autonoma.

Data l’impossibilità di un controllo umano *in itinere*, garanzie alternative potrebbero essere fornite dalla trasparenza o dall’assoluta affidabilità. Intrinseche ai sistemi di apprendimento automatico sono tuttavia le caratteristiche contrarie: questi sistemi sono infatti sono costitutivamente opachi (*black box*), soggetti a errori madornali – in quanto fondati su correlazioni statistiche di ogni genere, senza accesso al significato o al contesto – e ad attacchi avversari non rilevabili.

Nei primi anni del loro impiego, l'evidenza dei danni prodotti da simili sistemi è stata affrontata dalle *Big Tech* come un problema di discriminazioni, da eliminare con interventi di *design* tecnico. Le narrazioni sull'"etica dell'IA" e il "seducente diversivo del 'risolvere' i bias" hanno costituito un'operazione di cattura culturale, ossia di costruzione di una narrazione pubblicamente condivisa, attraverso il finanziamento e la direzione della ricerca e dei mezzi di intermediazione scientifica, fino a dettarne anche i toni. I giganti della tecnologia mirano a sottrarre i sistemi di IA alla regolazione giuridica, attraverso dichiarazioni di principi, linee guida, esperti, comitati e gruppi di lavoro sull'etica; a tale strategia di *ethic washing* si oppongono oggi la richiesta che le ricerche in questo ambito siano finanziate senza conflitti di interesse e la proposta di concettualizzare la questione nei termini della tutela dei diritti umani.

Paiono ormai superati, in virtù dello loro assurdità o malafede, anche i tentativi delle grandi aziende di sfuggire alle loro responsabilità appellandosi all'eccezionalità delle nuove tecnologie e proponendo, per gli effetti dannosi dei sistemi di apprendimento automatico, il riconoscimento di un vuoto di responsabilità, o una responsabilità distribuita anche tra gli utenti e le vittime (a differenza di quanto previsto per i profitti, secondo una consuetudine non eccezionale). Come osserva Andrea Bertolini, la dicotomia tra soggetti giuridici e oggetti non è superabile, *tertium non datur*, e l'unica classificazione ammissibile di tutte le tecnologie avanzate esistenti e ragionevolmente prevedibili – senza indulgere in tentazioni fantascientifiche – è quella di cose, oggetti e artefatti, prodotti dell'intelletto umano. Così concepite, esse rientrano chiaramente nella nozione di prodotto.

La fragilità della situazione è quella di una bolla giuridica, per usare un'espressione di Marco Giraudo: le grandi compagnie tecnologiche hanno fondato infatti il loro modello di business sull'appropriazione e la commercializzazione dei dati personali, in violazione di diritti giuridicamente tutelati, scommettendo su un successivo "salvataggio giuridico", in nome dell'inarrestabilità dell'innovazione tecnologica.

L'evidenza delle violazioni dei diritti individuali che hanno luogo quando si utilizzino sistemi di apprendimento automatico per attività che hanno effetti rilevanti sulle vite delle persone sta a fondamento della posizione di Frank Pasquale e Gianclaudio Malgieri. La loro proposta è di disciplinare i modelli di IA ad alto rischio incorporati oggi in prodotti e servizi attraverso una

presunzione di illegalità, ossia entro un sistema di "illegalità di default": fino a prova contraria, tali sistemi dovrebbero essere considerati illegali, e l'onere della prova contraria dovrebbe incombere alle aziende.

Prima di immettere sul mercato un prodotto o un servizio che incorpori sistemi di IA ad alto rischio, le aziende – a partire da quelle che esercitano ormai, per dimensioni e prerogative, una sovranità funzionale – avrebbero l'obbligo di dimostrare che la loro tecnologia non è discriminatoria, non è manipolatoria, non è iniqua, non è inaccurata e non è illegittima nelle sue basi giuridiche e nei suoi scopi. Potrebbero così trovare applicazione anche nei sistemi di IA i principi fondamentali del trattamento di dati personali fissati all'articolo 5 del Regolamento generale sulla protezione dei dati.

La proposta di Pasquale e Malgieri di inquadrare i sistemi di IA entro un regime di "illegalità di default" si fonda sulla priorità dei diritti individuali specificamente protetti dalla legge su un generico principio di innovazione, che è spesso la maschera dietro la quale i grandi soggetti economici rivendicano la tutela dei loro concreti interessi. Non è detto che i diritti individuali avranno la meglio, ma la proposta ha il pregio di chiamare le cose col loro nome.

**Questo articolo è stato pubblicato sul "Bollettino telematico di filosofia politica" dell'Università di Pisa.*