Verso la regolamentazione della Intelligenza Artificiale. Dimensioni e governo

Andrea Simoncini*

1. Quale regolamentazione?

La regolamentazione dell'impiego della IA nel settore della giustizia si colloca all'interno del ben più ampio tema della disciplina dell'IA nel suo complesso. Proverò, quindi, a proporre alcune rapide considerazioni di natura generale entro cui collocare le riflessioni emerse sulle possibili applicazioni di tale tecnologia nell'ambito del potere giudiziario.

Trenta o quaranta anni fa, porre la domanda «Quale regolamentazione per l'Intelligenza Artificiale?» avrebbe voluto dire chiedersi: «Quale *legge* per l'Intelligenza Artificiale?».

Con riferimento a temi vasti ed articolati come quelli della tecnologia, la "regolazione", infatti, normalmente coincideva con la "legislazione"[1] e, a seguire, con tutta la catena normativa discendente delle fonti secondarie e regolamentari di natura esecutivo-integrativo-attuativa.

Oggi lo scenario è molto più complesso.

Al tempo in cui scriviamo, con il termine "regolamentazione" indichiamo più propriamente un fine – e non più solo un mezzo – quello di rendere una determinata attività oggetto di una norma giuridica; orbene, questo fine si può realizzare attraverso uno spettro di strumenti molto più ampio della sola legge parlamentare (ovvero dei suoi equivalenti).

La legge, oggi, è solo una delle possibili strategie attraverso le quali possiamo disciplinare l'uso della tecnologia. La ragione di questa maggior complessità sta, in primo luogo, nel fatto che nel frattempo si sono (ulteriormente) moltiplicati i livelli di autorità pubbliche dotate di potere normativo, ma, in secondo luogo, sono profondamente cambiate le forme attraverso cui questo potere normativo viene realizzato.

Solo per proporre un elenco non esaustivo di questa nuova "fenomenologia" della regolazione, si pensi alle forme della regolazione privata o della "self-regulation", ovvero alle categorie del "nudging power" elaborate da Sunstein e Thaler; si pensi, altresì, all'effetto normativo che comportano le scelte finanziarie ovvero di mercato, alla capacità regolativa delle scelte operate sulla responsabilità civile, fino alla nota teoria di Lawrence Lessig per il quale «Code is the law»; ma su questo variegato panorama torneremo in seguito.

Il livello di complessità del problema finisce, poi, per aumentare se consideriamo che l'IA è un fenomeno suscettibile di applicazioni trasversali in moltissimi settori (privati e pubblici), dal momento che essa rappresenta essenzialmente una tecnologia per le decisioni. Noi utilizziamo l'IA per agire: per calcolare, decidere, predire, conoscere, eseguire, profilare etc. Per questo la tecnologia algoritmica finisce per sollecitare un po' tutte le possibili dimensioni normative, soprattutto quando si pensi di utilizzarla

* *Professore ordinario di Diritto costituzionale, Dipartimento di Scienze Giuridiche (DSG) dell'Università degli Studi di Firenze. Mail: andrea.simoncini@unifi.it.*

[1] Intendendo per "legislazione" tutto lo spettro di norme primarie previste nel nostro sistema costituzionale sia di origine parlamentare che governativa.

ISSN 2284-4503

in un settore delicato per la tutela dei diritti di tutti, quale quello delle decisioni giudiziarie o attinenti il mondo della giustizia.

A testimonianza di questa complessità sta il fatto che moltissimi organismi internazionali ed europei stanno ponendo il problema di come orientarsi in materia di regolazione della IA: tra il 2020 e il 2021 sia il Consiglio d'Europa, che il Parlamento Europeo e la Commissione hanno adottato numerosi documenti[2]. Ed infine il 21 aprile 2021 la Commissione ha presentato, sulla base di questi documenti, una proposta di regolazione comprensiva dell'IA[3].

2. Le diverse dimensioni della regolazione

Provando a delineare sommariamente lo spettro delle possibili alternative, tre sono le dimensioni fondamentali da prendere in considerazione, a seconda che si consideri l'*efficacia* delle regole ovvero il *soggetto* che le emana ovvero infine, il *contenuto* delle stesse.

Innanzitutto, quanto *all'efficacia* delle regole, la distinzione ormai classica è quella tra le cosiddette norme di *hard law* e di *soft law*.

Le prime rappresentano un sistema di regole dotate di piena obbligatorietà in senso giuridico, ovverosia suscettibili, in caso di violazione, di applicazione giudiziaria e di esecuzione forzata. La *soft-law*, invece, è un sistema di regole non precettive, comunque caratterizzato da un diverso grado di persuasività, ovverosia in grado di svolgere effettivamente una funzione di orientamento e di indirizzo nei confronti dei loro destinatari, sebbene non suscettibili di attuazione giudiziaria.

Diverso scenario appare se invece assumiamo come punto di vista la prospettiva degli autori delle regole, ovverosia dei soggetti che le adottano. In questa dimensione si può proporre una tripartizione. *In primis,* vi sono le forme di «etero-regolazione pubblica», ovverosia regole poste da parte di autorità pubbliche *diverse* dai destinatari (soggetti internazionali, europei, nazionali, regionali, locali); all'opposto stanno le forme di «auto-regolazione privata», ovverosia le regole poste da parte degli stessi destinatari delle regole (la c.d. *self-regulation*: si pensi ai principi etici di Google sull'uso della AI, ovvero ai c.d. Community Standards di piattaforme social come Facebook o Instagram, si pensi infine alle norme di autoregolamentazione adottate da associazioni di categoria ovvero alle norme ISO e a quelle di standardizzazione prodotte da organizzazioni private intervazionali etc.); esiste, però, una terza strada da percorrere con riferimento agli autori delle norme, quella della «co-regolazione», ovverosia forme di uso misto tra auto – ed etero – normazione, spesso qualificate dall'attività "quasi" regolatoria

[2] EPRS | European Parliamentary Research Service, *Artificial intelligence: From ethics to policy*, PE 641.507, giugno 2020 (https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2020)641507); Id., *Artificial intelligence: From ethics to policy*, PE 641.507, giugno 2020 (https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_STU(2020)641547); S. SAMOILI, M.L. COBO, E. GOMEZ, G. DE PRATO, F. MARTINEZ-PLUMED, B. DELIPETREV, *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence,* 2020, (EUR 30117 EN), Luxembourg: Publications Office of the European Union; High-Level Expert Group on AI, *Ethics guidelines for trustworthy AI*, Brussels, European Commission. Oppure sul tema specifico della giustizia si veda CEPEJ, *Carta etica europea sull'utilizzo dell'intelligenza artificiale nei sistemi giudiziari e negli ambiti connessi*. Strasburgo, 2018

[3] European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM(2021) 206 final.

ISSN 2284-4503

o giurisdizionale di autorità indipendenti (ad esempio, pareri o autorizzazioni generali delle *Autorities* di protezione dei Dati Personali, ovvero codici deontologici approvati dalle stesse *Autorities*).
Infine, la terza dimensione riguarda il contenuto della regolazione. In questa dimensione il fattore discriminante è la diversa modulazione della intensità normativa. Agli estremi dello spettro troviamo norme di divieto assoluto e norme di liceità assoluta; ovviamente ipotesi estremamente improbabili, ma che servono per definire il quadro teorico di riferimento in cui poi calare la regolazione concreta. Nel mezzo, per dir così, troviamo norme di divieto o di liceità relative ovverosia, prescrizioni in cui certe tecnologie sono consentite, purché rispettino determinate condizioni.
In quest'area, ovviamente, si colloca la maggior parte delle regole oggi indirizzate alla tecnologia in generale e alla IA in particolare. Sempre in questo segmento dobbiamo annoverare le forme più recenti di regolazione ispirate al cosiddetto "risk-based approach". Questo approccio è utilizzabile nel caso in cui un sistema tecnologico non sia oggetto né di un divieto assoluto né di liceità assoluta; consideriamo, dunque, strumenti o prodotti non vietati in sé, ma che possono essere utilizzati a patto che si rispettino determinate condizioni.
Nel risk-based model, l'oggetto della disciplina non è in primo luogo l'evento di danno e le sue conseguenze, ma, appunto, il "rischio" del danno, ovverosia la probabilità del suo verificarsi[4]. In tal modo, piuttosto che un sistema di tutele successive, che si limitino a stabilire quali conseguenze derivino dalla violazione di certe norme, si intende realizzare una forma di protezione preventiva, che riduca o azzeri la probabilità stessa delle violazioni[5]. In questa prospettiva, l'obbligazione principale per il destinatario della norma è quella, innanzitutto, di analizzare la propria attività e valutare i rischi connessi alla adozione di certe tecnologie; conseguentemente, deve adottare schemi organizzativi interni ed esterni vòlti a ridurre ovvero ad annullare tali rischi. Normalmente questi schemi organizzativi sono validati da organismi indipendenti di valutazione e controllo.
Ritengo che questa forma di regolazione meriti un'attenzione particolare dal momento che essa sta diventando quella di riferimento nelle scelte normative in questo settore; essa, infatti, ha un indubbio vantaggio che la rende preferibile ad altri modelli normativi: quello di "garantire" la liceità di una condotta in maniera preventiva, ovverosia, puntando su adempimenti richiesti *ex ante* ai destinatari, piuttosto che renderli responsabili *ex post* delle effettive lesioni o dei danni. Questo modello rispetto ad altri, soprattutto se applicato alle tecnologie pericolose, è decisamente il più favorevolmente orientato al mercato, in quanto consente agli imprenditori che producono e commercializzano tali tecnologie di considerare la cosiddetta "legal compliance" – ovvero il rispetto della normativa – come un costo di produzione e, così, di internalizzare tale costo nel calcolo economico della propria attività, piuttosto che rimanere esposti alla incertezza del dover rispondere comunque delle violazioni della legge. E proprio per questi motivi il "risk-based approach" è particolarmente funzionale agli scopi dell' Unione Europea, che da sempre è alla ricerca del difficile equilibrio tra l'elevata protezione dei diritti e l'obiettivo primario della crescita economica e della creazione del mercato unico; dimostrazione di questo è il fatto che il modello basato sulla gestione del rischio ha ispirato in parte la redazione del GDPR (il

[4] R. BALDWIN, J. BLACK, *Really responsive regulation*, in *The Modern Law Review*, 1, 2008, 65 ss.
[5] J. BLACK, *Risk-based regulation: choices, practices and lessons being learnt. Risk and Regulatory Policy,* in *Improving the Governance of Risk*, OECD, Paris, 2010.

ISSN 2284-4503

regolamento di disciplina dei dati personali)[6], ma è il modello fondamentale di riferimento della recente proposta di disciplina dell'*Artificial Intelligence Act*.

Va sottolineato, però, che il punto più sensibile sul piano costituzionale nella scelta di questa tecnica di regolazione, rimane la tutela di quelle sfere di libertà dei soggetti che qualifichiamo come "diritti fondamentali"[7]. Se, infatti, ci troviamo dinanzi a tecnologie che possono ledere un diritto fondamentale – oppure no, dipendendo da come esse sono realizzate – allora ben si potrà utilizzare il modello basato sulla gestione del rischio, a patto che gli elementi caratterizzanti i diritti fondamentali in gioco siano stati tutti correttamente presi in considerazione nella valutazione ex ante dei rischi.

Ma dinanzi a tecnologie che ledono comunque i diritti fondamentali (si pensi all'impiego di algoritmi strutturalmente non esplicabili e, dunque, non motivabili, nel settore della giustizia o della amministrazione pubblica, ovvero a sistemi che si basano sulla elaborazione di dati che non possono essere trattati in alcun modo perché discriminatori o sensibili) in questi casi non si potrà invocare l'adozione di misure preventive per impedire l'esercizio della tutela contro la violazione del diritto sia essa in via amministrativa che giudiziaria.

In altri termini, per poter adottare il *risk-based approach*, rimane cruciale aver correttamente e chiaramente tracciato la linea distintiva tra tecnologie *vietate*, in quanto *comunque* lesive di diritti fondamentali, e quelle *consentite,* a patto che vengano adottate le misure di gestione del rischio.

Rimane, infine, un'ultima considerazione generale che riguarda le diverse dimensioni della regolamentazione dell'IA: occorre andare verso una regolamentazione generale dell'IA come *genus* tecnologico, oppure è preferibile porre in essere discipline diversificate a seconda delle *species* (dei settori) in cui essa verrà applicata?

Direi che questa alternativa – che sicuramente permane come interrogativo teorico – in realtà, sia sostanzialmente risolta se si guarda alla prassi.

È infatti evidente che le applicazioni della IA nel settore della medicina dovranno essere regolate diversamente dalle applicazioni alle previsioni metereologiche, così come il settore della giustizia di cui ci occupiamo, richiede cautele ben diverse da quelle che debbono essere adottate nel settore della automazione dei processi industriali.

La regolamentazione della IA dovrà tener conto di alcuni principi generali che debbono avere una applicazione "cross-cutting", ma poi prevedere sistemi di disciplina adeguati all'oggetto preso in considerazione.

In conclusione, ritengo che la disciplina della IA, sia nella sua parte generale, che in quella diretta al settore della giustizia, debba utilizzare in maniera coordinata tutte queste dimensioni normative.

Quando parliamo di tecnologia, la decisione normativa rilevante non riguarda tanto la scelta della singola fonte che si intende utilizzare – *hard* o *soft law*, regolamento o direttiva europea, divieto assoluto o modello *risk-based*, etc.) – quanto la costruzione di un sistema coordinato di governo delle diverse tipologie di fonti, che converga verso lo scopo di regolamentare la AI, cioè di renderla compatibile

[6] C. COLAPIETRO, A. IANNUZZI, *I principi generali del trattamento dei dati personali e i diritti dell'interessato,* in *Innovazione tecnologica e valore della persona: il diritto alla protezione dei dati personali nel Regolamento UE 2016/679*, L. CALIFANO, C. COLAPIETRO (a cura di), Napoli, 85 ss.

[7] Per una visione ispirata alla contrapposizione netta tra "risk-based approach"e "right-based approach" cfr. F. HIDVEGI, D LEUFER, E. MASSÈ, *The EU should regulate AI on the basis of rights, not risks* https://www.accessnow.org/eu-regulation-ai-risk-based-approach/

ISSN 2284-4503

con quel sistema costituzionale di tutela dei diritti che consideriamo generalmente con il termine *rule of law*.
In concreto, la regolamentazione dell'IA può avvenire a seconda dei settori utilizzando tutto il vasto strumentario che nasce dalla matrice tra tutte le diverse variabili che abbiamo ricordato: efficacia, destinatario, intensità della regola.
È necessaria una strategia complessiva nella scelta del tipo di fonte, che si articoli a seconda del livello e del settore, capace di dare razionalità ed effettività ai singoli strumenti normativi.
D'altronde, a ben vedere, oggi il quadro è già composto da un variopinto *patchwork* di *pieces of legislation* diverse, che utilizzano più o meno tutte le dimensioni normative descritte.

3. Le ragioni di un governo coordinato delle diverse opzioni normative

Ci si può chiedere da dove derivi la ragione di questa scelta di un sistema di governo coordinato di fonti diverse anziché l'opzione per una o più di esse.
La ragione principale risiede nel grave deficit di effettività che ad oggi sconta il sistema normativo in questi ambiti.
Le forme più classiche di normazione, quelle indicate dalla letteratura aziendalistica come strumenti di «command and control», caratterizzate dal binomio attività «consentite/vietate» e normalmente assistite da sanzioni applicabili in via amministrativa o giudiziaria mondo della tecnologia, soprattutto nel settore della circolazione dei dati, dell'informazione e della comunicazione (ICT) sono caratterizzate da un elevato tasso di ineffettività.
Si pensi, ad esempio, al fallimento sostanziale della applicazione della normativa antitrust alle cosiddette "Big Tech". Il tentativo di far rientrare un fenomeno nuovo ed inedito come quello delle "piattaforme tecnologiche", nelle nozioni classiche di "abuso di posizione dominante" ovvero di "intese restrittive della concorrenza", ha fatto sì che, anche nei casi in cui la normativa antitrust è stata effettivamente applicata e sono state irrogate sanzioni anche molto rilevanti sul piano economico, l'applicazione della regolazione non è riuscita in alcun modo a scalfire la posizione sostanziale di monopolisti che attualmente tali industrie ricoprono.
Ma vi è un esempio più calzante visto il nostro tema specifico.
Si pensi, infatti, al GDPR, il noto regolamento europeo n. 679 in materia di dati personali entrato in vigore nel 2016; se si legge il Regolamento, si scopre che in realtà all'interno del vigente diritto "eurounitario" esiste già una norma generale in materia di decisioni adottate da sistemi di IA. Il principio, scolpito nell'art. 22 del GDPR – che a sua volta riprende l'art. 15 della precedente direttiva 95/46/CE – afferma che «l'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona».
In effetti questa norma esprime un principio generale di grandissima importanza con riferimento ai sistemi di trattamento automatico dei dati, tra i quali si inseriscono quelli di IA: nessuna decisione che interferisca significativamente con una persona può essere presa "unicamente" – *solely* – da un algoritmo (principio di "non esclusività"). Questa regola, per il suo tenore testuale, risolverebbe già

moltissime delle possibili controversie sull'utilizzo di algoritmi decisionali, ad esempio nel settore della giustizia. Tutte le volte che una decisione automatica non prevede un intervento umano, è vietata.
Chiediamoci, quante volte è stato applicato l'art. 22 del GDPR oppure – visto che ha un tenore letterale praticamente identico – l'art. 15 della direttiva 95/46, così come recepita dai vari stati?
Ebbene, le applicazioni giudiziarie o amministrative di queste regole sono state rarissime se non del tutto inesistenti.
Le ragioni di tale insuccesso sono molteplici. Innanzitutto, la stessa formulazione dell'art. 22 del GDPR. Il principio di "non esclusività", infatti, nel GDPR è soggetto a numerosissime eccezioni.
In base al testo vigente, una decisione algoritmica può legittimamente incidere sui diritti della persona anche senza alcun intervento umano:
a) quando sia necessaria per concludere o eseguire un contratto tra interessato e titolare del trattamento;
b) quando sia stata autorizzata dal diritto dell'Unione o dello stato membro;
c) quando vi sia il consenso esplicito dell'interessato.
È evidente che la portata di queste eccezioni nei fatti è amplissima, tanto che verrebbe da chiedersi se, in realtà, residui uno spazio effettivo per la regola. Per questo il principio di "non esclusività" pur esistendo da oltre 25 anni nei fatti ha avuto una scarsissima applicazione pratica.
Ma vi è un altro fattore che rende l'utilizzo delle forme classiche di regolamentazione spesso poco efficace nel controllo di queste tecnologie. La ragione è quella che altrove ho chiamato la travolgente "forza pratica" degli algoritmi. Torniamo all'art. 22 del GDPR, il fatto che la norma vieti l'uso di una tecnologia del tutto sostitutiva di una decisione umana ovviamente rende come unicamente praticabile l'impego di tecnologie "non sostitutive", ma semplicemente di supporto alla decisione. Ed in effetti oggi la tendenza più diffusa – soprattutto nella cosiddetta Social AI – non è quella di rimpiazzare il decisore umano con un robot che decida al suo posto, quanto di fornire al decisore umano valutazioni, suggerimenti, previsioni, che possano aiutarlo nella decisione.
L'effetto dell'art. 15 prima e dell'art. 22 ora, è stato rendere lecita solo quella che viene chiamata «collaborative AI» [8]. Ma qui si rafforza il dubbio sulla effettività di questo principio. Il fattore ignorato, infatti, è la forza *pratica* dei sistemi tecnologici.
Formulerei così l'azione di questa forza: una volta introdotto un sistema automatico di decisione all'interno di un processo decisionale umano, il sistema automatico tende, nel tempo, a catturare completamente la decisione stessa.
Questo accade, di badi, non per ragioni di valore scientifico, di accuratezza predittiva o di affidabilità tecnica dell'automatismo, ma eminentemente per ragioni di convenienza pratica.
Innanzitutto, per chi è stato leso in un suo diritto o in una libertà, dimostrare che la decisione lesiva sia stata presa unicamente sulla base di un algoritmo – e non "anche" sulla base – è spesso una *probatio diabolica*; chi ha preso la decisione (un funzionario di banca, un dirigente amministrativo, un giudice), infatti, avrà sempre la possibilità di sostenere che nel decidere si è solo *avvalso* dell'algoritmo, ma non

[8] Tra i tanti contributi, si veda l'articolo comparso sulla Harvard Business, H.J. WILSON, P. DAUGHERTY, *Collaborative Intelligence: Humans and AI Are Joining Forces,* https://store.hbr.org/product/collaborative-intelligence-humans-and-ai-are-joining-forces/R1804J

ISSN 2284-4503

è stato "sostituito" da esso, avendo in totale autonomia deciso di […] aderire alla valutazione effettuata dall'algoritmo.

Occorrerà certamente attendere la giurisprudenza sul punto, che significativamente tarda, ma quantomeno si dovrebbe ritenere che il principio di non esclusività è rispettato se il decisore umano rimane in grado, comunque, di esprimere una propria motivazione che giustifichi l'adesione alla valutazione effettuata dall'algoritmo o quantomeno di dare informazioni sulla logica utilizzata.

In secondo luogo, l'attuazione di un principio normativo come quello di "non esclusività" si scontra con l'evidente forza "pratica" di qualsiasi automatismo valutativo che, da un lato, solleva il decisore dal *burden of motivation* – dal peso dell'esame e della motivazione –; dall'altro, gli consente di "qualificare" la propria decisione con un crisma di "scientificità" ovvero "neutralità" che oggi circonda la valutazione algoritmica e le conferisce una peculiare "autorità".

Senza poi pensare alla interazione tra questi sistemi di supporto alla decisione e le regole sulla responsabilità (civile, penale, amministrativa, disciplinare) del decisore umano; per intendersi, consideriamo una situazione ipotetica: un medico che deve stilare il referto di un'analisi diagnostica, sarà libero di discostarsi dalla "valutazione predittiva" effettuata da un sistema di IA incorporato nella macchina e che "suggerisce" l'esistenza di una certa patologia? Soprattutto quando, in caso di errore, il medico potrebbe incorrere in una responsabilità professionale. Oppure, l'assicurazione professionale dello stesso dottore coprirà i danni eventualmente causati nel caso in cui il professionista si sia consapevolmente dissociato dalla "predizione" effettuata dalla macchina?

Per ovviare a questi deficit di effettività, la regolamentazione della IA in generale e, a maggior ragione nell'ambito della amministrazione della giustizia, dovrà necessariamente utilizzare un registro normativo a geometria variabile, capace di adeguarsi alle diverse dimensioni del problema. Di qui l'esigenza di un vero e proprio sistema di governo delle opzioni regolatorie, che possa allocare razionalmente le fonti di regolazione al fine di massimizzare la tutela dei diritti e delle libertà nei confronti di queste utilissime ma rischiose nuove tecnologie.