La garanzia dello *human in the loop* alla prova della decisione amministrativa algoritmica

Barbara Marchetti*

THE ALGORITHMIC ADMINISTRATIVE DECISION AND THE HUMAN IN THE LOOP

ABSTRACT: The article aims to examine the use of artificial intelligence by public administrations in a legal context characterized by the absence of general regulation and centered on the Courts ex post regulation. Starting from a case relating to a Ministry of Education's plan based on an algorithm, it analyzes the effectiveness of the European Data Protection Regulation and the actual role of the "human in the loop" principle, considering the different types of algorithm applicable and the legal background and expertise of public officials operating with artificial intelligence.

KEYWORDS: Artificial intelligence; public administration; administrative law guarantees; human in the loop; judicial review

SOMMARIO: 1. L'impiego dell'Intelligenza artificiale da parte della pubblica amministrazione e l'assenza di una regolazione generale – 2. Le difficoltà di una regolazione ex ante e il tentativo di una regolazione ex post: gli approdi della giurisprudenza amministrativa in ordine alla ammissibilità delle decisioni algoritmiche – 3. La pubblica amministrazione e l'utilizzo degli algoritmi – 4. Lo "human in the loop" al cospetto dell'autorità amministrativa – 5. Algoritmi model based e sistemi di machine learning alla prova delle garanzie del regolamento europeo sulla protezione dei dati personali – 6. Criticità e possibili correttivi: il ricorso avverso la decisione automatizzata dinanzi ad un decisore umano e l'importanza di un'amministrazione capace di governare l'Intelligenza artificiale.

1. L'impiego dell'Intelligenza artificiale da parte della pubblica amministrazione e l'assenza di una regolazione generale

«There is an old joke among pilots that says the ideal flight crew is a computer, a pilot and a dog. The computer's job is to fly the plane. The pilot is there to feed the dog. And the dog's job is to bite the pilot if he tries to touch the computer»[1]. Il rapporto tra l'umano e la macchina che questa storia ci consegna indica una fiducia assoluta nel funzionamento dell'intelligenza artificiale, che, in realtà, oggi hanno in pochi. Al tempo stesso, nessuno di noi pensa che l'uomo possa fare a meno dei sistemi di intelligenza artificiale, dati i benefici e le potenzialità che, nelle loro molteplici applicazioni, dischiudono per la vita dell'uomo. È dunque evidente che l'interazione tra l'uomo e la macchina costituisce la combinazione necessaria per assicurare i vantaggi e minimizzare i rischi dell'impiego dell'IA.

* *Professore ordinario, Facoltà di giurisprudenza, Università di Trento. Mail: barbara.marchetti@unitn.it.* *Contributo sottoposto a doppio referaggio anonimo.*

[1] Da *Humans may not always grasp why AIs act. Don't panic,* in *The Economist*, 15 febbraio 2018.

ISSN 2284-4503

Tra i soggetti che si avvalgono dei sistemi di IA per svolgere i propri compiti vi è, naturalmente, anche la pubblica amministrazione, che può utilizzarli nella propria attività di regolazione economica[2], nello svolgimento dei servizi pubblici (pensiamo al servizio sanitario), nell'adozione delle proprie decisioni autoritative[3]. In questo scritto si proverà a riflettere in particolare sull'impiego degli algoritmi nell'esercizio della funzione amministrativa e sulle garanzie che la giurisprudenza amministrativa, chiamata a pronunciarsi sulla nota vicenda del piano straordinario della scuola, ha individuato, da ultimo nella sentenza n. 881 del 2020, per conciliare il ricorso a procedure completamente automatizzate con l'esigenza di assicurare adeguate garanzie per il cittadino.

Prima di addentrarci in questa specifica questione occorre, però, muovere da una premessa sui fenomeni che qui si intende indagare, a partire dalla definizione di intelligenza artificiale. Quest'ultima è, come noto, definizione piuttosto controversa: tipicamente essa si rifà, anche nella terminologia adottata, all'idea di intelligenza umana e contempla diverse abilità che comprendono la capacità di apprendere, di astrarre, di ragionare, di usare il linguaggio. Accogliendo una definizione semplice ma basilare, in questa sede potremmo definire IA l'insieme dei sistemi (degli agenti intelligenti) che percepiscono l'ambiente che li circonda e intraprendono autonomamente azioni che massimizzano la possibilità di ottenere con successo obiettivi prefissati[4]. Data questa definizione, conviene subito dire che resteranno escluse dal nostro discorso sia la decisione informatica[5], in cui il supporto tecnologico rileva per il solo profilo della "forma" dell'atto, sia più in generale l'amministrazione digitale[6], in cui pur essendo

[2] In argomento cfr. F. BASSAN, *Potere dell'algoritmo e resistenza dei mercati in Italia. La sovranità perduta sui servizi*, Soveria Mannelli, 2019.

[3] Per una ricognizione dei diversi impieghi dell'IA da parte della p.a. cfr. G. AVANZINI, *Decisioni amministrative e algoritmi informativi. Predeterminazione analisi predittiva e nuove forme di intellegibilità*, Napoli, 2019. Per una riflessione generale sull'ascesa degli algoritmi e sul rapporto tra Stato, diritto e tecnologia in Italia cfr. A. SANTOSUOSSO, *Intelligenza artificiale e diritto. Perché le tecnologie di IA sono una grande opportunità per il diritto*, Milano, 2020; A. PAJNO e al., *AI: profili giuridici. Intelligenza artificiale: criticità emergenti e nuove sfide per i giuristi*, in *Biolaw Journal*, 3, 2019, 205; A. SIMONCINI, *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 4, 2019; M. BASSINI, L. LIGUORI, O. POLLICINO, *Sistemi di intelligenza artificiale, responsabilità e accountability. Verso nuovi paradigmi?*, in F. PIZZETTI (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Torino, 2018; C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, in *Biolaw Journal,* Special Issue 2, 2019, 711; C. CASONATO, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, 2019, 101; C. CASONATO, *Intelligenza artificiale e giustizia: potenzialità e rischi*, in *DPCE online*, 3, 2020, 3369; M. FASAN, *Intelligenza artificiale e pluralismo: uso delle tecniche di profilazione nello spazio pubblico democratico*, in *Biolaw Journal*, 1, 2019, 107; L. CASINI, *Lo Stato nell'era di Google. Frontiere e sfide globali*, Milano, 2020.

[4] Questa è la definizione ricavabile dal Rapporto della Commissione europea, *AI WATCH. Defining Artificial Intelligence*. In generale sul confronto esistente sulle diverse definizioni di IA v. N. PETIT, *Law and Regulation of AI and Robots: Conceptual Framework and Normative Implications*, working paper disponibile alla pagina https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2931339; S.J. RUSSELL, P. NORVIG, Artificial *Intelligence: A Modern Approach*, 2nd ed., New Jersey, 2003; K. WARWICK, *Intelligenza artificiale. Le basi*, Palermo, 2015.

[5] Se con essa si intende l'atto redatto con lo strumento informatico, questa fuoriesce dall'ambito dell'IA; diverso è il caso invece dell'atto il cui contenuto decisorio è determinato dalla macchina (elaborato elettronicamente o informaticamente), ossia il caso della decisione algoritmica. In proposito v. TAR Lazio, sez. III-bis, 22 marzo 2017, n. 3769.

[6] Sulla distinzione tra amministrazione digitale e amministrazione digitale algoritmica cfr. A. SIMONCINI, Amministrazione digitale algoritmica. Il quadro costituzionale, in R. CAVALLO PERIN, D.U. GALETTA, *Il diritto dell'amministrazione pubblica digitale*, Torino, 2020.

ISSN 2284-4503

decisivo il ricorso alla tecnologia nello svolgimento del lavoro delle p.a., le attività conoscitive e decisorie necessarie per l'esercizio dei compiti pubblici continuano a restare affidate ai funzionari amministrativi e dunque all'elemento umano.

Da una prima analisi della realtà delle amministrazioni italiane e utilizzando come banco di prova il contenzioso amministrativo potremmo essere portati a pensare che, se si considera la definizione di intelligenza artificiale appena contemplata, esse non facciano ad oggi un uso significativo di sistemi di IA.

Tuttavia, a ben guardare, in numerosi settori pubblici l'IA gioca già un ruolo significativo, sebbene in qualche caso non del tutto visibile. Troviamo, infatti, l'impiego di algoritmi nel settore della sicurezza (algoritmi-poliziotto), in quello bancario (algoritmi che segnalano anomalie indicative di possibile riciclaggio), nell'amministrazione fiscale (algoritmi che segnalano possibili rischi di frodi fiscali), in ambito sanitario[7], in materia di contratti pubblici, nei servizi pubblici (ad esempio alcuni algoritmi sono adoperati per la determinazione delle tariffe di acqua ed energia), nell'impiego pubblico e perfino nella classificazione delle zone rosse, arancioni e gialle utilizzate dal governo per differenziare l'applicazione delle misure anti-Covid 19[8].

A tale utilizzo da parte delle amministrazioni pubbliche di sistemi di IA non corrisponde, fino ad oggi, alcuna regolazione pubblica nazionale, nel senso che non è presente una disciplina legislativa generale. Esistono, come è noto, numerosi atti di soft law emanati dalle Istituzioni europee[9], sono applicabili alcune disposizioni del regolamento europeo per la garanzia dei dati personali e la cybersicurezza e sono stati adottati documenti programmatici e strategici a livello nazionale, ma manca ancora, su un piano generale, una disciplina compiuta, così come è solo agli inizi una riflessione sui caratteri che una regolazione dell'IA deve avere per essere efficace e tempestiva.

[7] Su cui v., tra gli altri, P. GUARDA, *Ok Google, am I sick?: Artificial Intelligence, e-health and data protection Regulation*, in *Biolaw Journal*, 1, 2019, 359.

[8] In argomento cfr. G. AVANZINI, *Decisioni amministrative e algoritmi informatici. Predeterminazione, analisi predittiva e nuove forme di intellegibilità*, Napoli, 2019, 35-76, in cui si illustrano le applicazioni di algoritmi nel settore pubblico; D.U. GALETTA, J.G. CORVALÀN, *Intelligenza artificiale per una pubblica amministrazione 4.0? Potenzialità, rischi e sfide della rivoluzione tecnologica in atto*, in *Federalismi.it*, 3, 2, 2019; F. BASSAN, *Potere dell'algoritmo e resistenza dei mercati in Italia. La sovranità perduta sui servizi*, Soveria Mannelli, 2019; F. COSTANTINO, *Rischi e opportunità del ricorso delle amministrazioni alle predizioni dei Big Data*, in *Diritto pubblico*., 2019, i, 43; sulle diverse applicazioni dell'IA e le ricadute sulla discrezionalità amministrativa cfr. A. CASSATELLA, *La discrezionalità amministrativa nell'età digitale, in corso di stampa*; in generale v. anche S. CIVITARESE MATTEUCCI, L. TORCHIA, *La tecnificazione*, Firenze, 2016.

[9] Per una ricognizione della strategia europea in materia di IA, anche in prospettiva comparata con l'esperienza statunitense, cfr. E. CHITI, B. MARCHETTI, *Divergenti? Le strategie di Unione europea e Stati Uniti in materia di intelligenza artificiale*, in *Rivista della regolazione dei mercati*, 1, 2020, 29 ss.; L. PARONA, *Prospettive europee e internazionali di regolazione dell'intelligenza artificiale tra principi etici, soft law e self regulation*, in *Rivista della regolazione dei mercati*, 1, 2020, 70; S. FRANCA, *La regolazione dell'intelligenza artificiale in Germania: stato dell'arte e prospettive future*, in *Rivista della regolazione dei mercati*, 1, 51, 2020.

2. Le difficoltà di una regolazione ex ante e il tentativo di una regolazione ex post: gli approdi della giurisprudenza amministrativa in ordine alla ammissibilità delle decisioni algoritmiche.

La circostanza che l'oggetto da regolare sia in continua evoluzione e abbia tratti inediti – ad esempio i sistemi di IA possono essere autonomi e imprevedibili, ciò che li rende parzialmente incontrollabili dall'uomo – rende estremamente complesso disciplinarli[10]: da un lato il Parlamento, e dunque la fonte legislativa, non ha la flessibilità né la velocità necessarie per seguire l'evoluzione tecnologica; dall'altro una regolazione del fenomeno solo da parte dell'esecutivo, attraverso fonti secondarie e in mancanza di una cornice legislativa, appare altrettanto discutibile in ragione dei possibili rischi per le libertà fondamentali e la *privacy*. Inoltre, regolare l'IA significa tante cose diverse tra loro: può significare individuare un regime giuridico per la robotica, introdurre una disciplina della responsabilità civile o penale dei sistemi di IA, prevedere limiti all'impiego di algoritmi, stabilire una regolamentazione per specifiche applicazioni di IA, come ad esempio i sistemi di riconoscimento facciale e così via.

A ciò si aggiunga la difficoltà di individuare il livello ideale per tale regolazione: che difficilmente potrà essere solo nazionale, ma che richiederà certamente un coordinamento sovranazionale o addirittura globale.

In un quadro normativo così incerto, e abitato sostanzialmente da atti di soft law[11], il ruolo delle Corti appare dunque fondamentale in termini di *ex post regulation*: sono, infatti, i giudici a trovarsi nella condizione di dover ricavare dall'ordinamento i principi e le regole che disciplinano il fenomeno e ad assicurare la protezione dei diritti e degli interessi lesi dall'applicazione dell'IA, non solo per i profili di responsabilità (civile e penale) ma anche – per ciò che qui maggiormente interessa – per stabilire se e in quale misura l'impiego di sistemi di IA sia compatibile con i valori costituzionali di un dato ordinamento e con le garanzie, nello specifico, tipiche del diritto pubblico[12].

Esistono numerose questioni a quest'ultimo proposito. Una prima questione riguarda il modo in cui la pubblica amministrazione si può procurare i sistemi di intelligenza artificiale di cui si avvale nella propria azione. Si potrebbe essere portati a ritenere che l'amministrazione acquisti all'esterno i sistemi di

[10] Negli Stati Uniti esiste un dibattito significativo sulla regolazione dell'IA: si veda per esempio M.U. SCHERER, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies and Strategies*, in *Harvard Journal Law and Technology*, 2016, Spring, 353; J. DANAHER, *Is Effective regulation of AI Possible? Eight Potential Regulatory Problems*, scaricabile alla pagina https://perma.cc/2gbn-Fvmm .

[11] Tra i molti atti di soft law si ricordano OECD, Council Recommendation on Artificial Intelligence, del giugno 2019; UNESCO, Preliminary study on the technical and legal aspects relating to the desirability of a standard-setting instrument on the ethics of artificial intelligence, del marzo 2019; UNESCO, First draft of the Recommendation on Ethics of Artificial Intelligence, del settembre 2020; Council of Europe, Committee of Ministers, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, dell'aprile 2020; Committee of Ministers, Declaration of the Committee of Ministers on the manipulative capabilities of algorithmic processes, del febbraio 2019; Commissioner for Human Rights, Unboxing AI: 10 steps to protect human rights -Recommendation of the Commissioner for Human Rights, del Maggio 2019; Consultative Committee of the Convention for the protection of individuals with regard to the Automatic Processing of Personal Data, Guidelines on Artificial Intelligence and Data Protection, del gennaio 2019; CEPEJ, European Ethical Charter for the use of artificial intelligence in judicial systems and their environment, del dicembre 2018. Da ultimo sulle prospettive regolatorie si veda Council of Europe Ad Hoc Committee on artificial Intelligence (CAHAI), *Towards Regulation of AI Systems*, reperibile al sito www.coe.int/cahai.

[12] Cfr. C. CASONATO, *Costituzione e intelligenza artificiale: un'agenda per il prossimo futuro*, cit., 711.

IA per i propri usi, sfruttando l'*expertise* dei soggetti privati che investono nella ricerca e nello sviluppo dell'IA, ma tale opzione – per quanto possa apparire per certi versi naturale – non è l'unica possibile e potrebbe rivelarsi non sempre opportuna. Le agenzie federali statunitensi, per esempio, realizzano *in house* la metà circa dei sistemi di IA che impiegano perché, perfino in un sistema tradizionalmente propenso alle esternalizzazioni e privatizzazioni, vi è il fondato timore che la scelta di affidare sempre e comunque a soggetti privati la costruzione di IA per il settore pubblico possa portare ad applicazioni che non rispondono adeguatamente alle esigenze organizzative e regolatorie della pubblica amministrazione[13].

Un'altra questione concerne la necessità di ripensare le competenze e le abilità di una parte della forza lavoro in servizio alla pubblica amministrazione: quando e se l'IA diventa uno strumento indispensabile di lavoro ed una tecnologia destinata a permeare l'azione pubblica e la prestazione dei servizi, occorre riflettere sulla opportunità di reclutare personale che possa governare tali nuovi strumenti, poiché, come detto e come vedremo più avanti, una delle condizioni fondamentali per uno sviluppo antropocentrico dell'IA è la meta-autonomia[14], ossia il mantenimento di un equilibrio tra la componente umana e la componente macchina, reso attraverso la garanzia dello "*human in the loop*".

Vi è infine la questione del modo in cui l'impiego di sistemi di IA incide sulla relazione tra amministrazione e cittadino, sullo statuto giuridico delle decisioni pubbliche adottate con algoritmi, sul patrimonio consolidato di garanzie che deve essere riconosciuto ai privati di fronte al potere pubblico[15]. A questi

[13] D. FREEMAN ENGSTROM, D.E. HO, C. M. SHARKEY, M.-F. CUÉLLAR, *Government by Algorithm: Artificial Intelligence in Federal Administrative Agencies*, scaricabile all'indirizzo https://www-cdn.law.stanford.edu/wp-content/uploads/2020/02/ACUS-AI-Report.pdf; secondo cui «building internal capacity, rather than simply embracing a default practice of contracting out for technical capacity, will be crucial to realizing algorithmic governance's promise and avoiding its perils». Se si guarda al sistema americano, si può vedere come la metà circa delle agenzie federali (45%, 64) si avvalgano di sistemi di IA per compiere analisi e monitoraggio della regolazione (Food and Drug Administration), ma anche per compiti di enforcement (ad esempio, è il caso della Security and Exchange Commission e della Custom and Border Protection), per la gestione di pubblici servizi, per finalità organizzative e per la fase di adjudication, soprattutto per procedimenti standardizzati (Social Security Administration e US Patent and Trademark Office).

Il dato interessante è che negli Stati Uniti le agenzie provvedono in buona parte in house allo sviluppo dei sistemi di IA (84 su 157). Talvolta però problemi di budget possono ostacolare o limitare lo sviluppo di sistemi di IA. Allora però oltre al contracting out è prevista la possibilità di prendere in prestito (borrow) collaborando con soggetti non commerciali o con altre amministrazioni.

Inoltre, l'esternalizzazione può porsi in tensione con alcuni valori centrali dell'amministrazione (il cui obiettivo non è solo economico ma concerne l'accesso per tutti, la qualità, la correttezza dell'agire amministrativo). Per un'analisi dei diversi impieghi dei sistemi di IA da parte dell'amministrazione federale cfr. Government by algorithm: Artificial Intelligence in federal Administrative Agencies, cit. Per un commento di tale report cfr. L. PARONA, Government by Algorithm: un contributo allo studio del ricorso all'IA nell'esercizio delle funzioni amministrative, in GDA, 2021, 1, in stampa.

[14] L. FLORIDI, J. COWLS, *A Unified Framework of Five Principles for AI in Society*, Nov. 2019, scaricabile all'indirizzo https://assets.pubpub.org/8kfahmow/c8d3cba5-8f10-4a00-894c-3a3b886ad844.pdf che parlano di «meta-autonomy, or a "decide-to-delegate" model: humans should retain the power to decide which decisions to take: exercising the freedom to choose where necessary, and ceding it in cases where overriding reasons, such as efficacy, may outweigh the loss of control over decision-making. Any delegation should also remain overridable in principle».

[15] Per un inquadramento costituzionale dell'amministrazione algoritmica cfr. A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, cit., 29.

ultimi profili si riferisce, in particolare, la giurisprudenza amministrativa da cui si muove in questo scritto, la quale ha esplorato le possibilità della amministrazione algoritmica e ne ha vagliato l'ammissibilità alla luce del quadro complessivo dei principi e delle garanzie del diritto amministrativo, primi fra tutti la trasparenza, i diritti di partecipazione, la motivazione, il diritto alla tutela giurisdizionale effettiva.

3. La pubblica amministrazione e l'utilizzo degli algoritmi.

Prima di esaminare gli approdi della giurisprudenza amministrativa nella vicenda ormai nota dell'algoritmo utilizzato per le assegnazioni delle sedi del piano straordinario della scuola, è bene introdurre un'altra distinzione. Essa riguarda l'uso strumentale o decisorio dell'algoritmo, intendendo con ciò il fatto che l'impiego di sistemi esperti possa intervenire sia nell'ambito dell'attività preparatoria e strumentale dell'amministrazione sia nel momento decisorio vero e proprio, porsi cioè come strumento per l'assunzione della decisione in sostituzione del funzionario.

Nel primo caso, il sistema di IA si limita a fornire informazioni o a effettuare segnalazioni che mettono poi in moto la decisione umana[16].

Potremmo ritenere queste ipotesi meno problematiche di quelle in cui l'algoritmo sostituisce la decisione e così in effetti è, soprattutto quando siamo di fronte ad algoritmi *model based*, funzionanti cioè secondo *hard rules*, ossia istruzioni ben definite e non ambigue che, eseguite correttamente dalla macchina, portano ad un risultato certo e predefinito. Si pensi ad alcuni algoritmi la cui applicazione è praticamente "invisibile" e non genera problemi di sorta, dall'auto-velox, a quelli per la valutazione di prove a crocette nei concorsi pubblici, a quelli che stabiliscono il traffico sulla rete ferroviaria, a quelli che valutano le offerte nelle procedure di gara.

Tuttavia, quando il sistema di IA che assiste l'umano è *machine learning*, capace cioè di auto-apprendere e di adottare decisioni in autonomia (con ciò che questo comporta in termini di prevedibilità e opacità delle decisioni), il rapporto tra tale sistema e la decisione umana può avere implicazioni assai più problematiche.

Si pensi ad un sistema *machine learning* di segnalazione di un'anomalia nelle operazioni bancarie che determini l'avvio di un'attività di indagine da parte di un funzionario dell'autorità di vigilanza bancaria; oppure ad un algoritmo che, incrociando i dati personali del passeggero (PNR) di un volo aereo con i dati dei social media conduca l'agente della polizia di frontiera ad effettuare ispezioni supplementari nei confronti di quella persona[17]. Teoricamente i funzionari sono liberi, in entrambe le situazioni, di valutare l'output del sistema, di interpretarne il contenuto e di decidere di conseguenza. Tuttavia, le possibilità che l'agente di polizia possa effettivamente controllare gli esiti del funzionamento dell'algoritmo ed eventualmente individuarne gli errori sono molto ridotte; ad esempio, gli agenti

[16] Sul punto cfr. C. COGLIANESE, D. LEHR, *Regulating by Robot: Administrative Decision Making in the Machine-Learning Era*, in Geo. I.J., 105, 2017, 1167; C. COGLIANESE, D. LEHR, *Transparency and Algorithmic Governance*, in *Administrative Law Review*, 71, 7, 2019. Secondo questi A. l'impiego di algoritmi machine-learning è per ora funzionale ad assistere il decisore umano, che dunque adotta la decisione, ma è inevitabile che gradualmente i sistemi machine learning verranno utilizzati anche per l'adjudication e la regulation.

[17] L'esempio è riportato da B. WAGNER, *Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems,* in *Policy and Internet*, 11, 1, 2019, 110.

dell'agenzia europea Frontex hanno circa 12 secondi di tempo per decidere se il viaggiatore segnalato dal sistema sia o no da ispezionare ed è improbabile che decidano di non seguire le indicazioni provenienti dal sistema. E lo stesso varrà per la decisione del funzionario dell'autorità bancaria circa l'opportunità di avviare o meno un'indagine.

È stato cioè osservato, a seguito ad alcuni studi di psicologia comportamentale, che esiste nel rapporto tra umano e macchina un *anchoring effect* (o come lo chiama Garapon un *effet moutonnier*[18]) dovuto al forte condizionamento esercitato dalla macchina sull'umano: applicato al rapporto tra utilizzo strumentale dell'IA da parte dell'amministrazione e fase decisoria in capo al funzionario, questo significa un peso potenzialmente determinante dell'IA sull'autorità chiamata ad adottare la decisione finale.

Sempre dagli studi condotti da alcuni studiosi di psicologia, ad esempio, risulta che una persona sia portata nell'88% dei casi a rimanere della propria opinione se si confronta con un'altra persona, e invece a mutarla (nel 66 % dei casi) se a pensarla diversamente è un sistema di IA. E ciò sul presupposto di una fiducia tendenzialmente elevata nella infallibilità dei sistemi esperti[19].

L'effetto di aggancio prodotto dalla fiducia nella macchina è ben esemplificato in un altro esperimento[20] in cui è stata testata l'influenza del noto sistema COMPAS di misurazione del rischio di recidiva assegnando gli stessi profili di imputati a due distinti gruppi di valutatori. Al primo gruppo è stato detto che COMPAS aveva classificato tali profili come *high risk*, al secondo gruppo che si trattava di profili secondo COMPAS *low risk*. La valutazione umana è risultata fortemente influenzata dalla classificazione della macchina, perché i medesimi profili sono risultati oggetto di una valutazione di rischio del 42 % superiore da parte del primo gruppo rispetto a quanto ritenuto dal secondo.

Occorre dunque considerare attentamente il modo in cui macchina e umano interagiscono anche quando l'impiego dell'IA è strumentale e preparatorio rispetto alla decisione, e non solo quando il sistema esperto sostituisce la decisione umana[21].

Se si passa all'ipotesi in cui, invece, l'algoritmo prende anche formalmente la decisione al posto del funzionario, e dunque stabilisce l'assetto del rapporto, si ha una relazione più diretta tra la macchina e il cittadino, poiché la sfera degli interessi e dei diritti di quest'ultimo dipende direttamente da come l'algoritmo è stato costruito, da quali dati gli sono stati forniti, da come è stato allenato e da come è giunto all'output finale. Se dunque i diritti del privato dipendono direttamente dall'impiego di un sistema di IA, ancor più diventa necessario garantire nei confronti di quest'ultimo trasparenza, partecipazione, diritto alla motivazione e tutela giurisdizionale.

[18] A. GARAPON, J. LASSEGUE, *Justice digital*, Paris, 2018.

[19] J.M. LOGG, J.A. MINSON, A. MOORE, *Algorithm appreciation: people prefer algorithmic to human judgment, Organizational Behavior and Human Decision Processes*, 151, 2019, 90-103.

[20] M. VACCARO AND J. WALDO, *The effects of Mixing Machine Learning and Human Judgment, Communication of the ACM*, 62, 11, 2019, 104-110.

[21] In questo senso cfr. C. CASONATO, *Per un'intelligenza artificiale costituzionalmente orientata*, in A. D'ALOIA (a cura di), *Diritto e intelligenza artificiale*, Milano, in corso di pubblicazione, p. 104 secondo cui «di fronte ad una tecnica percepita diffusamente, anche se erroneamente, come neutrale, oggettiva e sempre esatta è possibile, e forse probabile, che la persona incaricata di presiedere la procedura consideri più comodo e prudente non opporsi al risultato algoritmico, evitando di assumersi una responsabilità che la esporrebbe, ad esempio, a dover motivare il proprio comportamento. Il rischio, insomma, è che si sviluppi un «effet moutonnier» (che potremmo tradure come effetto pecorone) in base al quale la decisione sarebbe "catturata" dalla AI e il cd. diritto al human in the loop verrebbe svuotato di qualsiasi contenuto garantista.

Benché sia ancora limitato il ricorso a decisioni algoritmiche[22], la vicenda legata alle assunzioni nell'ambito della fase C del piano straordinario della scuola (l. 107/2015) ha prodotto – come è noto – una giurisprudenza del Consiglio di Stato che, superando un approccio di iniziale chiusura del TAR Lazio, ha saputo elaborare una prima riflessione originale sul regime giuridico applicabile alla decisione algoritmica. Su tali pronunce la letteratura di commento è stata particolarmente ricca e molte delle diverse questioni al centro delle decisioni dei giudici sono state considerate[23]. Sia consentito, tuttavia, muovere qui dalla sentenza che da ultimo ha dato atto del cammino intrapreso dal giudice amministrativo (la n. 881 del 4 febbraio 2020 della VI sezione), "consolidando" le precedenti decisioni (n. 2270/2019 e n. 8472/2019) per approfondire uno degli elementi centrali su cui essa si fonda, ossia la questione della meta-autonomia.

Ciò che, in particolare, interessa capire è in che misura la garanzia dello *human in the loop* sia effettiva e praticabile nei confronti di istruttorie e decisioni adottate in base a procedure automatizzate ed algoritmiche, fino a che punto cioè l'interazione uomo-macchina (che deve esprimersi nel potere di revoca della delega da parte dell'umano) sia garantita all'interno della relazione IA-amministrazione-cittadino.

Non saranno quindi esaminate in questa sede altre questioni centrali legate all'impiego dell'IA e affrontate dal Consiglio di Stato, quale quella del rapporto tra diritto di accesso (al codice sorgente) e diritto di proprietà intellettuale; delle implicazioni della decisione algoritmica in termini di tutela dei dati; della portata del principio di legalità rispetto all'esercizio del potere amministrativo attraverso algoritmi[24]; della esternalizzazione ai privati dello sviluppo dei sistemi di IA e delle innumerevoli implicazioni giuridiche, per esempio sotto il profilo della responsabilità[25].

Il Consiglio di Stato ha fissato le condizioni di ammissibilità delle decisioni algoritmiche in un percorso argomentativo che possiamo sintetizzare in cinque passaggi fondamentali: innanzitutto l'uso di sistemi esperti da parte della p.a. è ammesso ed è considerato un passaggio necessario e opportuno in ragione

[22] Il contenzioso può originarsi anche nel settore delle procedure di gara e di concorso, in ragione della presenza di software che determinano l'esclusione automatica di concorrenti per ragioni non imputabili all'interessato ma al malfunzionamento del sistema di accettazione delle domande (V., ad esempio, Cons. Stato, sez. VI, 18 maggio 2020 n. 3148).

[23] F. DONATI, *Intelligenza artificiale e giustizia*, in *Rivista AIC*, 2020, 1, 415; A. SIMONCINI, *L'algoritmo incostituzionale: l'intelligenza artificiale e il futuro delle libertà*, in *BioLaw Journal*, 1, 2019; M.C. CAVALLARO, G. SMORTO, *Decisione pubblica e responsabilità dell'amministrazione nella società dell'algoritmo*, in *Federalismi.it*, 6, 2019; S. TRANQUILLI, *Rapporto pubblico-privato nell'adozione e nel controllo della decisione amministrativa algoritmica*, in *Diritto e società*, 2, 2020 281; F. LAVIOLA, *Algoritmico, troppo algoritmico: decisioni amministrative automatizzate, protezione dei dati personali e tutela delle libertà dei cittadini alla luce della più recente giurisprudenza amministrativa*, in *Biolaw Journal*, 3, 2020, 389; A. VALSECCHI, *Algoritmo, discrezionalità amministrativa e discrezionalità del* giudice (nota a Cons. Stato, sez. VI, 4 febbraio 2020, n. 881), in *iusinitinere.it*, settembre 2020; N. MUCIACCIA, *Algoritmi e procedimento decisionale: alcuni recenti arresti della giustizia amministrativa*, in *Federalismi.it*, 15 aprile 2020; A. SOLA, *La giurisprudenza e la sfida dell'utilizzo di algoritmi nel procedimento amministrativo*, in *Giustamm.it*, novembre 2020; V. CANALINI, *L'algoritmo come "atto amministrativo informatico" e il sindacato del giudice*, in *Giornale di Diritto Amministrativo*, 6, 2019, 781; A. NICOTRA, V. VARONE, *L'algoritmo, intelligente ma non troppo*, in *Rivista AIC*, 4, 2019, 86.

[24] Su cui specificamente cfr. S. CIVITARESE MATTEUCCI, *Umano, troppo umano. Decisioni amministrative automatizzate e principio di legalità*, in *Diritto pubblico*, 1, 5, 2019.

[25] E. PICOZZA, *Intelligenza artificiale e diritto. Politica, diritto amministrativo and artificial intelligence*, in *Giurisprudenza italiana*, 7, 2019, 1657 ss.

dei vantaggi che producono in termini di rapidità, certezza e neutralità, soprattutto a fronte di enormi quantità di dati o di procedure standardizzate.

In secondo luogo, la circostanza che a tali strumenti di IA non sia applicabile la disciplina della legge 241/90, con le garanzie che essa dischiude, non può essere una ragione per vietare l'impiego degli algoritmi, poiché tale normativa nasce in un momento storico in cui la pubblica amministrazione non aveva l'opportunità di avvalersi dell'IA.

In terzo luogo, L'ammissibilità dei sistemi di IA non deve essere limitata alle decisioni vincolate, ma si estende anche a quelle caratterizzate da discrezionalità, quantomeno tecnica.

In quarto luogo, il ricorso agli algoritmi va inquadrato in termini di modello organizzativo: si tratta dunque di una diversa modalità di esercizio del potere autoritativo previsto dall'ordinamento, e dunque la scelta di tali strumenti non richiede una specifica ed espressa previsione legislativa.

Infine, l'apertura agli strumenti di IA nella funzione pubblica non significa sottrazione delle decisioni algoritmiche ad un regime pubblicistico di controllo e di *explainability*, ma la loro sottoposizione a regole, diverse da quelle individuate dalla legge 241/90, in grado di assicurare la compatibilità delle procedure e decisioni algoritmiche con il patrimonio delle garanzie proprie di un diritto amministrativo costituzionalmente orientato[26].

La ricerca di tali regole porta il Consiglio di Stato a considerare la disciplina contenuta nel Regolamento europeo per la tutela della privacy (General Data Protection Regulation, d'ora in avanti GDPR) scomponendola nelle seguenti garanzie: a) la pubblica amministrazione deve garantire *la piena conoscibilità dei criteri applicati* e del modello utilizzato; b) la decisione deve poter essere *imputata all'organo titolare del potere*; c) quest'ultimo deve poter svolgere *la necessaria verifica di logicità e legittimità della scelta e degli esiti* affidati[27].

Più specificamente, secondo il Consiglio di Stato, la conoscibilità deve intendersi sia con riferimento alla p.a. che decide di affidarsi ad una procedura basata su un algoritmo, sia avendo riguardo al destinatario degli esiti della decisione automatizzata. Essa inoltre implica conoscibilità di tutti gli aspetti, compresi i suoi autori, i dati immessi (input) e considerati rilevanti, il procedimento usato per l'elaborazione, il meccanismo di decisione, le priorità assegnate. Infatti, i criteri, i presupposti e gli esiti di tale procedura devono essere conformi alle prescrizioni e alle finalità stabilite dalla legge e devono, inoltre, poter essere sindacati dinanzi ad un giudice (p.to 10).

Ciò implica che la formula tecnica incorporata nell'algoritmo (la cui comprensione presuppone abilità informatiche, statistiche, matematiche) deve essere spiegata in modo da tradurla in "regola giuridica".

[26] Pensa che le categorie tradizionali del diritto amministrativo siano inadeguate ad inglobare le innovazioni prodotte dall'avvento dei sistemi di IA nell'amministrazione e che l'innovazione prodotta dai sistemi esperti modificherà il diritto amministrativo E. PICOZZA, *Intelligenza artificiale e diritto. Politica, diritto amministrativo and artificial intelligence*, cit., 1658.

[27] Sulle garanzie previste dal GDPR nei confronti delle decisioni automatizzate cfr. A. SIMONCINI, *Profili costituzionali della amministrazione algoritmica*, in *Rivista trimestrale di diritto pubblico*, 2019, 4; P. HACKER, *Teaching Fairness to Artificial Intelligence: Existing and Novel Strategies against Algorithmic Discrimination Under EU Law*, in *CMLR*, 2018, 55, 1143; B. CASEY, A. FARHANGI, R. VOGL, *Rethinking Explainable Machines: The GDPR's "Right to Explanation" Debate and the Rise of Algorithmic Audits in Enterprise*, in *Berkeley Technology Law J.,* 2019, 34:143, 144; S. WATCHER ed al., *Why a Right to explanation of Automated Decision-making Does Not Exist in the General Data Protection Regulation*, in *Int'l Data Privacy L.,* 2017, 76.

Secondo il Consiglio di Stato, in considerazione di questa esigenza, non può assumere rilevanza, in senso contrario, la riservatezza delle imprese produttrici (o le regole sulla proprietà intellettuale), le quali devono cedere di fronte alle esigenze di trasparenza della p.a.

Sono del resto gli artt. 13 e 14 del Regolamento europeo 679/2016 a stabilire che, di fronte ad una procedura automatizzata, il privato deve poter conoscere le «informazioni significative sulla logica utilizzata, nonché l'importanza e le conseguenze previste da tale trattamento per l'interessato»[28]. E tale principio di conoscibilità deve intendersi in termini di principio di comprensibilità.

Tale garanzia è completata dall'art. 15 del medesimo regolamento, il quale non solo prevede un diritto azionabile in capo all'interessato (e non il mero obbligo in capo al titolare del trattamento), ma anche di acquisire informazioni per tutto il corso della procedura automatizzata e anche in seguito all'adozione della decisione.

Il diritto di accesso alle informazioni si accompagna poi all'altra garanzia essenziale a fronte di una decisione automatizzata che incide sui suoi interessi, ossia il diritto della persona a non essere sottoposta ad una decisione interamente automatizzata senza il coinvolgimento umano, il diritto cioè alla supervisione di un umano. Questo non significa solo che deve essere assicurata l'imputabilità della decisione in capo ad un umano (profilo della responsabilità), ma che questo umano deve essere in grado di verificare la logicità e la legittimità della decisione algoritmica.

Questa condizione – intesa come espressione del principio di non esclusività della decisione algoritmica (p.to 11.2) – comporta, secondo il Consiglio di Stato, che, nel processo decisionale, debba comunque esserci una persona che possa «controllare, validare o smentire (discostarsi) dalla decisione automatica», ossia che possa revocare la delega alla macchina e riappropriarsi della decisione.

L'affermazione di queste nuove garanzie si accompagna, peraltro, ad una rinuncia cruciale: escono, infatti, di scena per questi procedimenti gli *hearing rights*, cioè i diritti di partecipazione. Il passaggio, per quanto necessario e forse naturale in considerazione delle caratteristiche delle decisioni algoritmiche, le quali non ammettono una partecipazione nelle forme classiche di un contraddittorio (orale o scritto), costituisce un sacrificio molto significativo per il destinatario della misura, che forse in un'ottica di bilanciamento potrebbe non essere sempre *giustificato*: per esempio, se a fronte di procedimenti standardizzati a basso contenuto discrezionale la sua compressione potrebbe essere ragionevole, lo stesso potrebbe non valere per provvedimenti afflittivi con un contenuto discrezionale significativo, i quali dunque non si presterebbero, in questa logica, ad essere sostituiti da provvedimenti del tutto automatizzati.

[28] Sul diritto alla explainability e sulla sua portata controversa (nel senso cioè che esso non implica necessariamente un diritto alla spiegazione della decisione individuale) cfr. M.E. KAMINSKI, G. MALGIERI, *Multi-layered Explanations from Algorithmic Impact Assessments in the GDPR*; F. SOVRANO, F. VITALI, M. PALMIRANI, *The difference between Explainable and Explaining: requirements and challenges under the GDPR*, in XAILA 2019 Explainable AI in Law 2019. Proceedings of the 2nd Explainable AI in Law Workshop (XAILA 2019) co-located with 32nd International Conference on Legal Knowledge and Information Systems (JURIX 2019), Aachen, CEUR-WS.org, «Ceur Workshop Proceedings», 2019, 2681, 1; F. SOVRANO, F. VITALI, M. PALMIRANI, *Modelling GDPR-Compliant Explanations for Trustworthy AI*, *in Electronic Government and the Information Systems Perspective*, Cham, «Lecture Notes In Artificial Intelligence», 2020, 12394, 219-233.

4. Lo "human in the loop" al cospetto dell'autorità amministrativa.

Prima di esaminare in che misura la garanzia dello *human in the loop* possa operare al cospetto di una pubblica amministrazione che faccia uso di un sistema di IA e interrogarsi sul ruolo che il funzionario pubblico può concretamente svolgere in termini di guardiano della macchina, una prima precisazione riguarda il *richiamo* e al contempo il *superamento* della disciplina contenuta nell'art.22 del Regolamento europeo da parte del nostro giudice amministrativo.

Nel GDPR, infatti, il divieto di decisioni esclusivamente automatizzate soffre di tre eccezioni di portata amplissima, rappresentate dai casi in cui la decisione automatizzata sia necessaria per concludere un contratto (a), oppure sia autorizzata dal diritto dell'Unione o dello Stato membro (b) o ancora sia intervenuto il consenso esplicito dell'interessato[29].

È evidente che si tratta di eccezioni così estese da poter neutralizzare la garanzia stessa prevista dall'art. 22.

La giurisprudenza del Consiglio di Stato, invece, è stata chiara nell'affermare che una decisione completamente automatizzata suscettibile di incidere sui diritti degli interessati risulta di per sé *incostituzionale*, a prescindere dalla violazione della disciplina fissata nella legge 241/90: così facendo, i giudici hanno imposto – a livello nazionale – una tutela maggiore di quella europea[30].

Per verificare se, nel caso di decisioni pubbliche, possa dirsi presente la garanzia di un umano in grado di controllare la decisione, di disporre dell'autorità e della competenza necessarie a cambiare la decisione e di considerare tutti i dati rilevanti al fine di correggere la macchina quando questa sbaglia, occorre affrontare tre diversi profili.

Il primo è di carattere generale, e impone di interrogarsi su come si concili la garanzia dello HITL con sistemi di *machine learning*, cioè con sistemi di IA che funzionano sulla base di algoritmi capaci di autoapprendere. A fronte di sistemi esperti con tali caratteristiche, operanti cioè secondo processi decisionali che restano opachi e non prevedibili per gli stessi programmatori – contraddistinti cioè dalla c.d. *black box* – occorre verificare le effettive possibilità per l'umano di mettere in discussione i risultati cui è pervenuta la macchina, controllare la correttezza dei dati immessi, e correggere gli output (o eliminare il *bias*, cioè l'effetto discriminatorio)[31].

Il secondo ha a che vedere con la circostanza che la p.a. generalmente esternalizza a soggetti privati la costruzione dei sistemi di IA che utilizza, sicché, rispetto alla macchina, occorre domandarsi chi sia – tra il funzionario amministrativo e il costruttore del codice sorgente – l'umano che deve mantenere il controllo sul funzionamento del processo.

[29] Sottolinea l'eccessiva latitudine di queste eccezioni A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, cit., 30.

[30] A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, cit., 29, il quale ricorda come il divieto di una decisione completamente automatizzata venga fatto derivare dai valori costituzionali ricavabili dagli artt. 3, 24 e 97, oltre che dall'art. 6 della Convenzione europea dei diritti dell'Uomo.

[31] Secondo quanto si legge nelle linee guida stabilite dall'art. 29 Working party «To qualify as human intervention, the controller must ensure that any oversight of the decision is meaningful, rather just a token gesture. It should be carried out by someone who has the authority and competence to change the decision. As part of the analysis, they should consider all the available input and output data»: così B. WAGNER, *Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems*, in *Policy and Internet*, 11, 1, 2019, 104.

Infine, se deve essere il funzionario pubblico, come parrebbe, cioè l'utilizzatore finale del sistema, a controllare l'IA, occorre riflettere sul tipo di competenze matematiche ed informatiche che dovrebbero sussistere in capo al medesimo per evitare la cattura dell'umano da parte della macchina (c.d. effetto aggancio prima richiamato), sia quando l'intelligenza artificiale opera in fase istruttoria sia quando sostituisce del tutto la decisione umana.

Per affrontare tali questioni vorrei riprendere anzitutto la definizione di decisione algoritmica, perché nella giurisprudenza che si interroga sulla ammissibilità di tali decisioni e sul loro regime giuridico, esse vengono intese come un insieme indistinto.

In realtà, stabilire le regole cui questo tipo di atti devono rispettare e riflettere sulla necessità di garantire rispetto ad esse lo HITL non può prescindere da una distinzione fondamentale tra algoritmi *model based*, fondati su *hard rules* (che sono relativamente semplici) e sistemi *machine* (e *deep*) *learning*, anch'essi fondati su algoritmi, ma caratterizzati – come si diceva poc'anzi – dalla capacità di autoapprendimento della macchina, di autonomia di giudizio, il cui funzionamento è imperscrutabile e imprevedibile allo stesso disegnatore[32].

I sistemi *deep learning* si basano, per esempio, su reti neurali (artificiali) le quali simulano – secondo modelli matematici – i collegamenti dei neuroni cerebrali umani, riuscendo a ricavare, da una quantità enorme di dati (input) conseguenze (output) che sono l'esito della capacità della macchina di autoapprendere e di selezionare (sulla base dell'esperienza) gli output ottimali per raggiungere un determinato obiettivo.

Le due tipologie di sistemi di IA – *model based* e *machine* (*deep learning*) – pongono problemi e sfide del tutto diversi quando impiegati dalla pubblica amministrazione. Un esempio ci aiuta a chiarire.

5. (Segue). Algoritmi model based e sistemi di machine learning alla prova delle garanzie del regolamento europeo sulla protezione dei dati personali

Si consideri un algoritmo creato per stabilire quali studenti ammettere in una data Università (ma in termini analoghi potremmo pensare ad un algoritmo costruito per decidere chi ha diritto ad un sussidio o a ricevere una data autorizzazione). Cominciamo con il considerare il funzionamento di un algoritmo funzionante in base ad *hard rules* e individuiamo questa come l'ipotesi A.

In questo caso, l'algoritmo viene costruito sulla base di una sequenza di istruzioni, definite e non ambigue, che possono essere eseguite meccanicamente dalla macchina al fine di produrre un determinato risultato[33]. Supponiamo che per individuare quali studenti ammettere ad una determinata istituzione

[32] In tema cfr. Y. BATHAEE, *"The Artificial Intelligence Black box and the Failure of Intent and Causation"*, in *Harvard Journal of Law and Technology*, 31, 2, spring 2018, 890 ss.

[33] Generalmente queste istruzioni sono inserite secondo la logica booleana classica, che riconosce due valori, vero o falso, acceso o spento, 0 o 1 e così via. Qualche volta invece gli algoritmi sono costruiti secondo la logica fuzzy (o logica sfumata) secondo cui si ammette che un oggetto sia contemporaneamente vero e falso, ma con diversi gradi (o valori) di verità o di appartenenza. Per esempio dato l'insieme da 0 a 1 (per giovane) un neonato avrà un valore di 1, un diciottenne un valore di 0.5, un sessantenne un valore di 0,10. Quando l'algoritmo è costruito secondo una logica fuzzy, secondo i computer scientists, esso potrebbe assicurare un grado di explainability maggiore. In argomento, cfr. R. CHIMATAPU, H. HAGRAS, A. STARKEY, G. OWUSUM, *Explainable AI and Fuzzy Logic Systems, Conference Paper*, in *Conference on theory and practice of Natural Computing*, 2018, TPNC 2018, Lectures Notes in Computer Science, vol. 11324; J. ADAMS, H. HAGRAS, *A Type-2 Fuzzy Logic Approach to*

ISSN 2284-4503

universitaria si utilizzino tre dati: il risultato conseguito nel test universitario, la media dei voti (o il voto finale) della scuola superiore e un terzo punteggio numerico assegnato a ciascun candidato sulla base della difficoltà del percorso scolastico (si può accordare, per esempio, un punteggio più alto per alcuni tipi di liceo, medio per altri tipi di liceo e più basso per gli altri istituti scolastici).
Alla macchina viene detto di moltiplicare per cinque il risultato del test, per sei la media dei voti (o il voto di maturità) e poi di aggiungere il punteggio legato alla difficoltà del percorso scolastico moltiplicato per tre. In esito a queste operazioni di calcolo ponderato, vengono ammessi all'Università gli studenti che si trovano nel 10% alto della lista dei candidati.
A fronte di un algoritmo di questo tipo, come è evidente, il sistema di IA si comporta come un calcolatore potente, che sa elaborare con una velocità di calcolo superiore a quella umana, i dati immessi traducendoli in output prevedibili dal programmatore. Naturalmente ciò non significa che in assoluto questo sistema sarà esente da qualunque problema, né che sarà certamente rispondente all'interesse pubblico: potremmo avere, ad esempio, errori nella creazione dell'algoritmo – ad esempio per l'immissione di istruzioni errate – oppure potrebbe essere irragionevole la scelta operata in fase di determinazione dei criteri se si stabilisse, irragionevolmente, che un certo percorso di studi pesi sproporzionatamente di più di un certo altro percorso di studi.
Ciò che rileva ai nostri fini, tuttavia, è che rispetto a questo tipo di algoritmi, l'esito è *certo e prevedibile*, ed il processo che dagli input porta all'output è stabilito dal programmatore secondo la sequenza "*if - then*"[34].
Per tali ragioni, questi algoritmi *model based* non pongono problemi di *explainability*: l'umano è, infatti, in grado di spiegare quali sono i criteri per l'ammissione all'Università, quali i fattori valutati e in che modo essi sono stati ponderati, così come di correggere eventuali illogicità o illegittimità delle decisioni, ad esempio, accorgendosi che, dati certi input, l'output non è corretto per un'errata impostazione dei dati.
Su un modello simile era fondato l'algoritmo "incriminato" per il piano straordinario della scuola.
In questo caso, resta da capire, a fronte della esternalizzazione della costruzione del sistema di IA, come e se il funzionario amministrativo che in concreto sta impiegando l'algoritmo sia in grado di avere il controllo del funzionamento della macchina e di correggerne l'eventuale esito errato.
Se, infatti, sul piano degli input, sarà la p.a. a decidere (anche esercitando la propria discrezionalità tecnica) i fattori rilevanti e il loro peso – nell'ipotesi A, per esempio, sarà stato *frutto di discrezionalità tecnica la decisione dell'amministrazione di attribuire un determinato punteggio ai diversi percorsi scolastici* (diversi tipi di licei, altri istituti scolastici, istituti tecnici) – e dunque a fornire i criteri per la costruzione del codice sorgente, non è altrettanto sicuro che il funzionario pubblico abbia le competenze necessarie (e adeguate) per controllare, dati gli input, il funzionamento dell'algoritmo e la correttezza degli output.
In questa logica, ci si dovrebbe dunque interrogare sulla opportunità di assicurare un rapporto di collaborazione tra funzionario (incaricato di quel procedimento) e soggetto privato (cui è stata

Explainable AI for Regulatory Compliance, Fair Costumer outcomes, and Market Stability in the Global Financial Sector, 2020 IEEE International Conference on Fuzzy Systems, scaricabile al sito https://ieeexplore.ieee.org/xpl/conhome/9171991/proceeding.

[34] K. WARWICK, *op.cit.*, 93 ss.

ISSN 2284-4503

esternalizzata la creazione dell'algoritmo) non solo in fase di *design* e di trasmissione dei dati (input), ma anche nella fase di applicazione del sistema, a compensare l'eventuale impossibilità per il funzionario amministrativo – che abbia una formazione non specifica – di operare la supervisione necessaria sulla macchina.

Il problema cambia qualitativamente in caso di sistemi *machine* e *deep learning* (basati cioè su algoritmi che auto-apprendono), corrispondente alla ipotesi B.

Prendiamo la medesima procedura di ammissione all'Università utilizzata prima: si danno al computer gli stessi dati (risultato test universitario, media voti scuola superiore, punteggio per difficoltà del percorso scolastico) ma, in aggiunta, si immettono nel sistema i dati dello storico delle ammissioni, consentendo che da questi il computer riceva ed estrapoli ulteriori dati (che potrebbero riguardare, ad esempio, elementi e caratteristiche dei profili degli studenti ammessi, quali il genere femminile o maschile, la zona geografica di provenienza, l'appartenenza ad una certa etnia) i quali permettono al sistema di apprendere e selezionare, attraverso connessioni interne alla propria rete neurale non intellegibili al programmatore (di qui, appunto, l'immagine della *black box*[35]), gli esiti ottimali (output).

Attraverso questo processo decisionale complesso, il sistema giungerà a premiare taluni candidati a svantaggio di altri, senza che sia possibile comprendere, però, la ragione per cui è giunto ad una determinata scelta piuttosto che ad un'altra. In questo caso, è evidente che *transparency* ed *explainability* non sono garantite: anzi, tanto maggiore è la complessità delle reti neurali e la quantità dei dati inseriti, minore sarà la possibilità di comprendere perché, dati quegli input, si è giunti a certi output.

La macchina, infatti, non spiega le ragioni delle proprie scelte.

Ciò è problematico su più fronti: da un lato, in termini generali, per effetto della autonomia della macchina nell'apprendere dai dati, potrebbe generarsi il rischio di *bias*: ad esempio, se nell'ipotesi della procedura di selezione, lo storico delle ammissioni dovesse mostrare una prevalenza di studenti ammessi di sesso maschile, il sistema potrebbe eleggere tale dato di genere come dato rilevante per le successive selezioni, senza che da parte dell'amministrazione fosse presente alcuna volontà di premiare tale elemento, e talvolta persino nell'ipotesi in cui l'Università abbia un interesse specifico ad invertire tale fattore[36].

[35] F. PASQUALE, *The Black Box Society. The Secret Algorithms that Control Money and Information*, Cambridge; London, 2016.

[36] Royal Society, Machine Learning: The Power and Promise of Computers that Learn by Example, Report (Apr. 2017). Interessante, in proposito, è il caso dell'algoritmo Ofqual (Office of Qualification and Examinations Regulation) con cui nel Regno Unito si era deciso di calcolare il voto finale di maturità nell'estate 2020 in ragione della impossibilità dello svolgimento degli esami legata alla Pandemia da Covid-19. La pubblicazione dei risultati ha mostrato come nel 40 % dei casi l'algoritmo avesse prodotto votazioni inferiori alle previsioni fatte dagli insegnanti (anche di due classi inferiori) e come ciò si fosse tradotto in una discriminazione per gli studenti provenienti da scuole pubbliche e da contesti sociali più svantaggiati. La costruzione dell'algoritmo si basava, in particolare, sul riconoscimento di un peso determinante non tanto al curriculum scolastico del singolo studente ma al rendimento accademico delle scuole di provenienza. A seguito delle proteste in Scozia e in Inghilterra, il 17 agosto 2020 il Segretario di Stato per l'istruzione G. Williamson ha stabilito che OFQUAL accetti gli studenti in base al voto degli insegnanti invece che in base all'algoritmo.

Inoltre, la mancanza di trasparenza e di "spiegabilità" impedisce ai candidati non ammessi di difendersi di fronte all'esclusione, non potendo contare sulla comprensibilità delle ragioni alla base della decisione adottata dall'istituzione universitaria per poterle contestare dinanzi ad un giudice[37].

Infine, nella prospettiva che qui si è privilegiata della esigibilità della garanzia dello *human in the loop*, pare evidente che la *explainability* costituisce anche la condizione necessaria per la supervisione e l'eventuale correzione del funzionamento del sistema[38]. Tendenzialmente, infatti, solo se l'umano è in grado di comprendere il modo in cui il sistema di IA decide (e può avere fiducia in esso solo se ne comprende il funzionamento), può operare un controllo effettivo sulla macchina e, nel caso, revocarle la delega e correggerne gli output[39].

Da questo punto di vista, per i sistemi *machine learning* le difficoltà di garantire lo HITL sono ben maggiori rispetto a quanto osservato per i più semplici sistemi *model based*, perché nonostante gli sforzi dei *computer scientists* per rendere più comprensibile il funzionamento di tali sistemi esperti, non solo il funzionario amministrativo che impiega il sistema *machine learning*, ma lo stesso disegnatore e sviluppatore incontra un ostacolo nella imprevedibilità dei processi decisionali che ne rende impossibile la supervisione e la correzione.

Se questo è vero, la garanzia assicurata dallo *Human in the Loop*, ossia dalla interazione uomo-macchina vacilla e, stando agli approdi della giurisprudenza amministrativa, diventa necessaria una riflessione supplementare in ordine alla compatibilità tra sistemi *machine learning* e azione amministrativa. In questa riflessione, entrano diversi piani.

Per le ragioni che si sono evidenziate, anzitutto, andrebbe promosso, in generale, uno sviluppo delle competenze interne alla pubblica amministrazione, in grado di assicurare una maggiore consapevolezza e padronanza dei sistemi di IA, sia in fase di progettazione e conferimento dei dati, sia sul fronte del loro concreto utilizzo: la introduzione di abilità informatiche e matematiche adeguate nella forza lavoro della p.a. avrebbe l'effetto di rendere possibile, una volta sviluppato il sistema di intelligenza artificiale, un controllo effettivo sul suo funzionamento e sui suoi *output*[40].

In secondo luogo, occorrerebbe procedere alla formulazione di principi e linee guida che orientino la p.a. non solo sul versante del rapporto con il cittadino ma sul versante interno, fornendo una cornice unitaria entro cui collocare le scelte sul ricorso all'IA, i vantaggi e gli svantaggi connessi ai diversi sistemi esperti, la tipologia di procedure sostituibili con procedure automatizzate, la questione dei dati

[37] Ancora prima, il provvedimento sarebbe illegittimo perché senza motivazione. Cfr. C. CASONATO, *Per un'intelligenza artificiale costituzionalmente orientata*, cit., 96; A. SIMONCINI, *Amministrazione digitale algoritmica. Il quadro costituzionale*, cit., 32.

[38] L. FLORIDI, J. COWLS, *A Unified Framework of Five Principles for AI in Society*, cit.; per un adattamento del principio di trasparenza al fine di ammettere sistemi machine learning nella pubblica amministrazione cfr. C. COGLIANESE, D. LEHR, *Transparency and Algorithmic Governance*, in *Administrative Law Review*, 71, 1, 2019.

[39] Tale controllo, peraltro, è anche il presupposto necessario per contenere i rischi di discriminazione algoritmica legati al machine learning: è sempre il Consiglio di Stato p.to 11, sentenza 881/2020), richiamando il considerando 71 del regolamento 679/2016, ad esigere dal titolare del trattamento le misure tecniche e organizzative adeguate al fine di scongiurare i possibili effetti discriminatori delle decisioni automatizzate e a presumere una "necessaria Cooperazione" a tal fine tra la pubblica amministrazione e chi ha istruito la macchina che ha prodotto la decisione.

[40] In questi termini, cfr. anche S. TRANQUILLI, *Rapporto pubblico-privato nell'adozione e nel controllo della decisione amministrativa robotica*, cit., 298.

utilizzabili, i rapporti con il privato affidatario, le implicazioni dell'impiego di intelligenza artificiale sul piano della responsabilità del funzionario e così via.

In questo senso, per esempio, il recente *Executive Order on Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government* statunitense, emanato il 3 dicembre del 2020, può essere d'ispirazione, nella formulazione dei suoi nove principi-guida.

Tramite esso, si obbliga l'amministrazione ad un ricorso all'IA il più possibile consapevole e orientato ai valori pubblicistici. In particolare, è richiesto alle agenzie di operare un bilanciamento costi benefici (in cui i benefici devono superare i rischi ed i rischi devono comunque essere accertabili e gestibili) prima di ricorrere ad un sistema esperto; le agenzie sono tenute ad assicurare che le operazioni e gli esiti delle applicazioni di IA siano "sufficientemente comprensibili" da parte degli esperti e degli utilizzatori; inoltre, è previsto che le responsabilità siano ben definite, comprese e correttamente assegnate per il design, lo sviluppo, l'acquisizione e l'uso dell'IA, così come gli input e output devono essere il più possibile tracciabili e documentati; ancora, ogni applicazione di intelligenza artificiale deve essere *costantemente* monitorata, supervisionata, ed eventualmente disattivata se contrastante con alcuno di tali principi.

L'*order* fissa poi un principio di trasparenza in capo alle agenzie in ordine all'impiego di applicazioni di IA (sia rispetto al Congresso che rispetto al pubblico), che peraltro deve essere compatibile con la protezione della privacy e con le normative in materia di polizia, di sicurezza nazionale o di informazioni protette.

Infine, è stabilito che le agenzie federali apprestino adeguate salvaguardie per un impiego corretto dei sistemi esperti oltre che una formazione del proprio personale funzionale ad assicurare design, sviluppo, acquisizione ed impiego responsabili delle applicazioni di IA.

Ora, benché alcuni di tali principi appaiano vaghi e non del tutto definiti, essi ciò nonostante fissano una prima cornice di riferimento utile per il settore pubblico, entro cui stabilire regole più di dettaglio relative alle diverse agenzie federali e alle specifiche applicazioni di IA in uso al loro interno.

Inoltre, l'esempio nordamericano pone l'accento sulla prospettiva organizzativa interna, cioè quella che guarda alle scelte a monte in tema di affidamento esterno o *in house* dei sistemi di IA, di individuazione delle responsabilità tra umano e macchina nei processi decisionali, di formazione della forza lavoro per governare i sistemi esperti, tutte questioni che sono ancora poco sviluppate nel dibattito italiano e che tuttavia appaiono cruciali per un utilizzo consapevole, razionale e costituzionalmente orientato dei sistemi di IA nel settore pubblico.

6. Criticità e possibili correttivi: il ricorso avverso la decisione automatizzata dinanzi ad un decisore umano e l'importanza di un'amministrazione capace di governare l'Intelligenza artificiale.

L'applicazione di sistemi di IA nei processi decisionali della p.a. pone sfide inedite e richiede un ripensamento delle tradizionali garanzie offerte ai privati di fronte all'azione amministrativa. Questo ha portato il Consiglio di Stato a rimodulare i principi del giusto procedimento contenuti nella legge 241/90 per adattarli alla realtà delle decisioni basate su algoritmi.

In questo scritto si è cercato di riflettere, in particolare, sulla applicazione problematica del principio di non esclusività della decisione automatizzata, in ragione, da un lato, della mancanza di adeguate abilità informatiche e matematiche all'interno della p.a. e, dall'altro, della oggettiva impossibilità di assicurare l'interazione con l'umano, lo *human in the loop,* in caso di sistemi *deep* e *machine learning*.
In ordine al primo aspetto, pare urgente avviare una riflessione sulla opportunità di dotare l'amministrazione di esperti in grado di governare i sistemi di intelligenza artificiale: ciò assicurerebbe non solo una maggiore consapevolezza in ordine ai rischi e ai benefici legati alle scelte circa l'impiego di IA nel settore pubblico, ma anche una capacità concreta di controllarne il funzionamento (e correggerne gli esiti) nelle applicazioni pratiche.
In ordine al secondo aspetto, di complessità maggiore, si possono immaginare due opzioni: la prima va nella direzione dell'esclusione dell'utilizzo di sistemi *machine-learning* da parte della p.a. per la loro inconciliabilità con il principio di *explainability*[41] e con la necessità di garantire lo HITL. Questa posizione di chiusura rischia, però, di limitare fortemente il potenziale dell'IA nel settore pubblico, dato che renderebbe tendenzialmente inutilizzabili tali sistemi perfino quando questi operano in funzione strumentale e preparatoria.
La seconda non ne precluderebbe l'applicazione, ma imporrebbe la previsione di ulteriori meccanismi e procedure in grado di compensare il loro deficit in termini di *explainability* e di contenimento dei rischi, anche discriminatori, che possono accompagnarne l'utilizzo nell'azione amministrativa.
Su questo fronte, oltre ai progressivi sforzi dei *computer scientists* sul piano dell'*explainable AI*, si può provare ad indicare alcune misure specifiche per il settore pubblico, frutto delle riflessioni compiute anche in altri Paesi.
Una potrebbe essere la previsione (già contemplata in Francia dal *Conseil Constitutionnel*[42]) di un ricorso amministrativo avverso la decisione automatizzata dinanzi ad un decisore umano. Tale ricorso, interno alla p.a., dovrebbe essere possibilmente rapido, flessibile, informale ed economico. Attraverso di esso, la garanzia dello HITL sarebbe recuperata in un secondo momento, assicurando una verifica successiva da parte dell'umano. Tale soluzione presenta, però, degli inconvenienti: da un lato, proprio in ragione della complessità del sistema esperto, la correzione dell'*output* richiede in capo all'umano la capacità di replicare il processo decisionale compiuto dall'algoritmo.
D'altro canto, di fronte ad un sistema *machine-learning* che ha funzionato male o che contiene *bias*, il ricorso – pur se attivato da un singolo individuo – avrà probabilmente l'effetto di inficiare l'intero esito della procedura automatizzata di massa, vanificando le finalità di celerità ed efficienza per cui si era fatto ricorso all'IA.

[41] In Francia, con una riforma del 2016 (*Loi pour une République numérique*, finalizzata a garantire la trasparenza a fronte di procedure e decisioni algoritmiche), è stato introdotto l'obbligo per le p.a. sia di pubblicare le informazioni sugli algoritmi utilizzati (obbligo che comprende la pubblicazione dei criteri e della logica cui è informato il funzionamento della macchina) sia l'obbligo di assicurare ai privati interessati la piena conoscibilità dell'algoritmo, con la conseguenza che l'amministrazione sarebbe chiamata a non impiegare algoritmi il cui funzionamento non può essere spiegato compitamente al cittadino. Sul punto cfr. S. TRANQUILLI, *op. cit*., 308.
[42] Conseil Constitutionnel n. 2018-765, riportata da S. TRANQUILLI, *op. cit.,* 309. L'A. riporta anche il caso giunto all'attenzione del Conseil Constitutionnel (CC n.2020-834) relativo al sistema Parcoursup utilizzato per valutare le domande di ammissione all'Università da parte degli studenti delle scuole superiori.

Un altro possibile correttivo (ricavabile anch'esso dal GDPR, che lo prevede con riferimento ai dati trasmessi per la decisione automatizzata nell'art. 35[43]) potrebbe consistere nella sottoposizione dei sistemi esperti ad un *Algorithmic Impact Assessment* (AIA). Tale valutazione potrebbe essere imposta dalla legge alla pubblica amministrazione, la quale prima di avvalersi di procedure basate su sistemi di *machine learning* (ma anche di algoritmi *model based*) sarebbe tenuta a verificarne la tenuta sul piano dei possibili *bias*, della correttezza e della compatibilità con i valori pubblicistici[44]. Ciò potrebbe implicare per esempio la verifica dell'accuratezza dell'algoritmo attraverso alcuni test, che consentano eventuali correzioni del codice sorgente alla luce dei criteri e degli input prescelti dall'amministrazione. Anche in questo caso, la decisione resterebbe interamente automatizzata, ma si assicurerebbe almeno una maggiore accuratezza dell'algoritmo nella fase di progettazione e sviluppo del modello, alla luce dei *public values*.

Il modello proposto da parte della dottrina statunitense[45] prevede, ad esempio, una fase di *notice and comment*, con la pubblicazione dell'algoritmo (codice sorgente, dati utilizzati, training data) ed una di raccolta di osservazioni da parte di ricercatori ed *auditors*. La criticità è data dal fatto che a fronte di algoritmi complessi tale procedura potrebbe non garantire, comunque, un effettivo scrutinio a monte. Inoltre, la comunicazione dell'algoritmo potrebbe incontrare il limite posto dai diritti di proprietà intellettuale opponibili dalle società private che hanno sviluppato l'IA.

In ogni caso, l'applicazione di un sistema esperto da parte della p.a. dovrebbe almeno essere preceduta da una procedura di validazione o certificazione condotta da esperti. In uno studio del 2017[46] sono stati individuati e testati, per esempio, alcuni metodi di verifica che consentono di controllare il funzionamento degli algoritmi e la loro conformità ad un insieme di regole applicate coerentemente in tutti i casi.

Tale verifica può essere affidata, ancora una volta ad un soggetto privato (diverso da quello che ha sviluppato il sistema esperto) oppure ad unità di esperti interni all'amministrazione stessa (potrebbe

[43] Su questo cfr. B. CASEY, A. FARHANGI, R. VOGL, *Rethinking Explainable Machines: The GDPR's "right to explanation" Debate and the Rise of Alghorithmic Audits in Enterprise*, in *Berkeley Technology Law Journal*, 2019, vol. 34, 143. L'art. 35 del GDPR prevede una valutazione di impatto sulla protezione dei dati nel caso di «a) una valutazione sistematica e globale di aspetti personali relativi a persone fisiche basata su un trattamento automatizzato, compresa la profilazione, e sulla quale si fondano decisioni che hanno effetti giuridici o incidono in modo analogo significativamente su dette persone fisiche». Partendo da questa disposizione le linee guida sul *Data Protection Impact Assessment* (DPIA) invitano alla pratica di AIA in forma di self-regulation al fine di evitare bias e discriminazioni e al tempo stesso con l'intento di fornire un primo livello di spiegazioni sul funzionamento dell'IA.

[44] In questo senso si veda la Directive on Automated Decision-Making del Governo canadese del 2019, che prevede una Algorithm Impact Assessment per la propria amministrazione. In argomento cfr. E. MOSS, E.A. WATKINS, J. METCALF, M.C. ELISH, *Governing with Algorithmic Impact Assessments: Six Observations*, scaricabile da SSRN, 28 maggio 2020; F. MCKELVEY, M. MAC DONALD, *Artificial Intelligence Innovations at the Canadian Federal Government*, in *Canadian Journal of Communication*, 2019, 44, 2, 43.

[45] A.D. SELBST, *Disparate Impact in Big Data Policing*, 52 GA. L. REV. 109 (2017), 169, ricordato da M.E. KAMINSKY, G. MALGIERI, *Algorithmic Impact Assessment Under the GDPR: Producing Multi-layered Explanations*, U of Colorado Law Legal Studies Research Paper 19/28, scaricabile da SSRN, 12 ottobre 2020.

[46] J.A. KROLL et al., *Accountable Algorithms,* in *University of Pennsylvania Law Review*, 2017, 14, 33.

trattarsi di un'agenzia specializzata a livello statale[47] o di organi tecnici a livello regionale) in grado di condurre una validazione dei sistemi di IA secondo i principi e i valori propri del settore pubblico[48].

La ricerca di misure compensative di contrasto ad "un'amministrazione black box", cioè ad un'amministrazione che deleghi le proprie decisioni a processi automatizzati autonomi e inspiegabili, dovrebbe insomma andare nella direzione di un potenziamento del controllo umano, di un più consapevole bilanciamento dei costi e benefici legati all'IA e della promozione di inediti processi di verifica e validazione *ex ante* ed *ex post*.

Riprendendo l'immagine con cui si apre questo breve scritto, dunque, l'umano non è tenuto affatto (solo) a nutrire il cane, ma deve trovare il modo, al contrario, di interagire significativamente con la macchina. Le tecniche ed i meccanismi di questa interazione sono da costruire velocemente ed efficacemente, e devono riguardare non solo il momento dell'applicazione, ma prima ancora essere rilevanti in fase di *design* e progettazione. Essi inoltre devono essere corrispondenti alle diverse logiche e caratteristiche dei sistemi di IA cui si riferiscono, e adattarsi al contesto pubblico o privato in cui sono chiamati ad operare. Si tratta di pensare nuovi strumenti e nuove categorie, la cui individuazione, peraltro, presuppone, per essere efficace, un dialogo proficuo tra *computer science* e diritto, tra *policy makers* e tech companies, tra giuristi ed esperti di informatica.

[47] Questa per esempio è l'ipotesi auspicata da parte della dottrina nordamericana. Cfr. A. TUTT, *An FDA for Algorithms*, in *Administrative Law Review*, 2017, 69:1, 83, secondo cui «a federal regulatory agency could be an effective means of dealing with the challenges posed by these kinds of complex algorithms in the future». In particolare, secondo l'A. un'agenzia federale potrebbe «provide a comprehensive means of organizing and classifying algorithms into regulatory categories by their design, complexity and potential for harm (in both ordinary use and misuse)»; inoltre l'agenzia potrebbe stabilire un regime di autorizzazione ex ante per stabilire regolare l'utilizzo di determinati sistemi di IA non solo nell'amministrazione pubblica ma anche nel mercato, più in generale. La creazione di un'agenzia federale con funzioni regolatorie ipotizzata da Tutt sarebbe opportuna non solo per stabilire quali algoritmi siano utilizzabili nel settore pubblico, ma più in generale nel mercato. L'agenzia cui occorrerebbe ispirarsi, secondo l'A., sarebbe la Food and Drug Administration, in ragione della medesima preoccupazione per i rischi e il benessere generale generati dal commercio dei farmaci.

[48] LORD SALES, *Algorithms, Artificial Intelligence and the Law*, June 2020, Judicial Review, Routledge. In tema di AI e public values, cfr. anche L. ANDREWS, *Public administration, public leadership and the Construction of public value in the Age of the Algotithm and Big Data*, scaricabile da https://onlinelibrary.wiley.com/doi/epdf/10.1111/padm.12534.