



1 MAGGIO 2024

Città intelligenti, decisioni “biased” e rischi di esclusione

di Beatrice Baldini

Dottoranda di ricerca in Diritto, politica e sviluppo
Scuola Superiore Sant’Anna



Città intelligenti, decisioni “biased” e rischi di esclusione*

di Beatrice Baldini

Dottoranda di ricerca in Diritto, politica e sviluppo
Scuola Superiore Sant’Anna

Abstract [It]: Le *smart cities* fronteggiano le più complesse sfide urbane facendo leva sulla diffusione di strumenti tecnologici volti a raccogliere dati funzionali a orientare l’elaborazione di politiche pubbliche. Sebbene siano molteplici i vantaggi connessi all’impiego dell’IA da parte dell’amministrazione, altrettanto plurimi sono i rischi derivanti dalla raccolta massiccia di dati. In particolare, il contributo si sofferma sulle c.d. “*biased decisions*”, esaminando la regolazione di matrice europea e calandola nel contesto cittadino.

Title: Smart cities, “biased” decisions, and risks of exclusion

Abstract [En]: Smart cities face the most complex urban challenges by leveraging the diffusion of technological tools aimed at collecting functional data to guide the development of public policies. While there are many benefits to using AI by the administration, there are also many risks associated with massive data collection. In particular, the paper focuses on the so-called “biased decisions”, examining the European regulation and placing it in the urban context.

Parole chiave: città intelligenti, pregiudizi, regolazione europea, sorveglianza umana

Keywords: smart cities, bias, European regulation, human oversight

Sommario: 1. Verso la digitalizzazione della società? 2. Fronteggiare gli errori umani con i sistemi di IA... 2.1. ... con il rischio di incorrere in condotte pregiudizievoli 3. Il perpetuarsi di *bias* nelle città 3.1. Alcuni esempi tratti dall’esperienza straniera 3.2. Trattare della *smart mobility* in Italia mettendosi al riparo dai rischi 4. Le ambizioni europee nella regolazione dello sviluppo tecnologico 4.1. Una riflessione a partire dal General Data Protection Regulation 4.2. Alcune considerazioni alla luce della Proposta di Regolamento sull’intelligenza artificiale 5. *Smart cities* e *human oversight*: il complesso ruolo del soggetto pubblico 6. *L’human in the loop* per l’eguale trattamento dei cittadini: uno scenario plausibile?

1. Verso la digitalizzazione della società?

Gli strumenti digitali, insinuatisi in ogni aspetto della vita quotidiana, stanno portando a una progressiva digitalizzazione della società¹, sino a influenzare il modo di concepire gli spazi urbani, nonché i rapporti tra cittadini e autorità pubbliche.

Cresce infatti incessantemente l’attenzione riposta sulla progettazione in chiave *smart* delle città² per far leva sui concetti di «velocità, efficienza, ottimizzazione», promuovendo «l’uso delle tecnologie

* Articolo sottoposto a referaggio.

¹ E. SPILLER, *Il diritto di comprendere, il dovere di spiegare. Explainability e intelligenza artificiale costituzionalmente orientata*, in *Riv. di BioDiritto*, 2, 2021, p. 419 parla addirittura di «progressiva automatizzazione della società».

² L. BUSATTA, *Partecipazione, inclusione e interoperability: l’ottimizzazione dei servizi alla persona nella smart city sostenibile*, in G.F. FERRARI, *Le smart cities al tempo della resilienza*, (a cura di), Mimesis, 2021, p. 587.

dell'informazione e della comunicazione (TIC)»³. I *big data*, l'*Internet of Things* (IoT), l'Intelligenza Artificiale (IA) hanno condotto alla «evoluzione dei legami tra tecnologia e dimensione urbana»⁴ tradottasi nella diffusione delle *smart cities*.

Tuttavia, se, da un lato, molteplici sono i vantaggi offerti dalla raccolta massiva di dati e dal ricorso all'IA, dall'altro, altrettanto plurimi sono i rischi, che mettono a dura prova la tenuta dei diritti e delle libertà della persona, derivanti da un utilizzo incontrollato e distorto della tecnologia⁵.

Tra tali rischi, un profilo problematico, già da tempo discusso in dottrina, attiene ai pregiudizi e alle discriminazioni perpetrati tramite l'utilizzo di algoritmi. Si tratta di conseguenze derivanti dai sistemi di IA che, sfruttando «dati incompleti, obsoleti o *biased*»⁶, compiono un'attività di profilazione degli utenti errata e riproducono, «nella sfera digitale, i pregiudizi e gli stereotipi già esistenti nella realtà»⁷.

Ciò ha assunto una «portata globale e drasticamente pervasiva» sia per il «campo di applicazione [...] vastissimo»⁸ di tali condotte pregiudizievoli, che per le diverse tipologie di vittime da esse prodotte. Peraltro, alla luce del rapido evolversi della tecnologia, per lungo tempo non si è assistito a disposizioni giuridiche «formalizzate» e «consolidate» sul piano operativo⁹, riguardanti tale specifico aspetto.

Il contributo intende inserirsi in questo quadro, soffermandosi, dapprima, sui concetti di IA, algoritmi e *bias*, per poi calarsi nello spazio urbano con la presentazione di esempi tratti dal più ristretto ambito della *smart mobility*, sia relativi all'esperienza straniera, che nazionale. Dopodiché, l'analisi delinea gli strumenti messi a disposizione sul piano giuridico, rilevando il registrarsi di un'intersezione tra sfera pubblica e privata quando sono prese in considerazione le decisioni assunte tramite l'utilizzo di algoritmi. Infine, il lavoro mostra come le citate sfere siano oramai tagliate trasversalmente dalla Proposta di Regolamento

³ M. CAPORALE, *L'intelligenza si ripartisce o si condivide? A proposito di smartness, livelli di governo e una certa idea di città*, in *IdF.*, 4, 2015,

⁴ M. DEMICHELIS, *Il governo digitale del territorio. Dagli usi temporanei alla rigenerazione urbana*, Jovene, 2023, p. 146.

⁵ Senza poter citare tutti gli A. che hanno trattato dei rischi derivanti dall'uso dell'IA, rendono l'idea di due fra le più grandi complessità da governare (protezione della *privacy* e rimozione del divario sociale) G. MALGIERI, *Vulnerability and Data Protection Law*, in *Oxford Data Protection & Privacy Law*, 2023 e L. ATTIAS, G. SCORZA, *La consapevolezza digitale al servizio dell'etica*, in *Dir. inf.*, 6, 2019, pp. 1196. Questi ultimi affermano emblematicamente che «la rivoluzione tecnologica in corso farà sì [...] che la mancanza di adeguate competenze e abilità digitali, specie di base, in assenza di adeguati correttivi, taglierà fuori milioni di cittadini dalla partecipazione alla vita sociale, economica, culturale e democratica del Paese».

⁶ G. GOMETZ, *Intelligenza artificiale, profilazione e nuove forme di discriminazione*, in *Teoria e storia del diritto privato*, NS, 2022, p. 4.

⁷ G. GIORGINI PIGNATIELLO, *Il contrasto alle discriminazioni algoritmiche: dall'anarchia giuridica alle Digital Authorities?*, in *federalismi.it*, 16, 2022, p. 166.

⁸ G. GIORGINI PIGNATIELLO, *Il contrasto*, cit., p. 167.

⁹ Come messo in luce da B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *Riv. di Biodiritto*, 3, 2021, p. 418: l'IA «costituisce un oggetto difficile da regolare, sia perché, ancor più di altre tecnologie innovative, è caratterizzata da incessanti sviluppi che rendono rapidamente obsoleta qualsiasi disciplina volta a regolarla, sia perché, nei suoi sistemi più avanzati (*machine learning*, *deep learning*, *neural networks*) si contraddistingue per una forte dose di autonomia e imprevedibilità di funzionamento, la quale, accompagnata all'inspiegabilità dei processi interni (fenomeno della *black box*) può rappresentare una potenziale fonte di rischi, non calcolabili *ex ante*».

dell'Intelligenza Artificiale (c.d. Ai Act)¹⁰, che, una volta entrata in vigore in via definitiva, costituirà la base di partenza anche per la regolamentazione *smart* delle città in ragione della sua applicazione generalizzata ai sistemi di IA.

2. Fronteggiare gli errori umani con i sistemi di IA...

Le pubbliche amministrazioni (PA) si avvalgono in misura crescente dei sistemi di IA per svolgere i propri compiti, nell'ottica di semplificare e garantire il buon andamento dell'attività e dell'organizzazione amministrativa¹¹: dalla regolazione economica, all'erogazione dei servizi pubblici, all'adozione «delle proprie decisioni autoritative»¹². In tal senso, anche nelle PA i sistemi di IA divengono una «protesi elettronica dell'intelligenza umana»¹³ in grado di percepire l'ambiente circostante e di intraprendere «autonomamente azioni che massimizzano la possibilità di ottenere con successo gli obiettivi prefissati»¹⁴. Nel processo di digitalizzazione in corso, l'IA affianca l'intelligenza umana per velocizzare i processi decisori e, potenzialmente, ridurre gli errori. Il risultato dovrebbe essere «l'innalzamento del livello di razionalità delle scelte assunte», proprio perché a guida della macchina vi sono «criteri logico-matematici»¹⁵. Ciò si palesa su piccola scala, ovvero a stretto contatto con il cittadino-utente, i cui dati vengono raccolti e archiviati per mezzo di prestazioni tecnologiche sempre più sofisticate e miranti a proiettare la città verso «una crescita sostenibile rispetto agli *standard* qualitativi della vita di coloro che la abitano»¹⁶.

L'IA può assumere le vesti di algoritmi «*machine* (e *deep learning*)», differenziati dagli «algoritmi *model based*»¹⁷ in quanto capaci di «autoapprendimento»¹⁸. In particolare, la macchina svolge in modo autonomo i suoi compiti, analizza «dati ignoti», partendo da un «certo *set* di dati conosciuti»¹⁹, e si migliora imparando dalle

¹⁰ Gli emendamenti alla Proposta di Regolamento dell'Intelligenza Artificiale sono stati approvati lo scorso 14 giugno, mentre, lo scorso dicembre, è stato raggiunto l'accordo sopra citato. Per gli ulteriori passaggi dell'*iter* di approvazione a cui è stata sottoposta la Proposta si rinvia al paragrafo 4.2.

¹¹ Statuisce in tal senso il Tar Campania, Sez. III, 14 novembre 2022, n. 7003.

¹² B. MARCHETTI, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, in *Riv. di BioDiritto*, 2, 2021, p. 368.

¹³ F. FALLETTI, *Decisioni automatizzate e diritto alla spiegazione: alcune riflessioni comparatistiche*, cit., p. 170.

¹⁴ B. MARCHETTI, *La garanzia dello human in the loop*, cit., p. 368.

¹⁵ N. PAOLANTONI, *Il potere discrezionale della pubblica automazione. Sconcerto e stilemi. (Sul controllo giudiziario delle "decisioni algoritmiche")*, in *Dir. amm.*, 4, 2021, p. 820.

¹⁶ M. DEMICHELI, *Il governo digitale del territorio*, cit., p. 147.

¹⁷ O algoritmi condizionali «che non fanno altro, [...] che applicare regole (informatiche)» e il cui «codice risulta in ogni caso perfettamente intellegibile per un essere umano dotato delle necessarie competenze tecniche» come spiega G. CARULLO, *Decisione amministrativa e intelligenza artificiale*, in *Dir. inf.* 3, 2021, p. 435.

¹⁸ Le tre citazioni appartengono a B. MARCHETTI, *La garanzia dello human in the loop*, cit., p. 378.

¹⁹ G. CARULLO, *Decisione amministrativa*, cit., pp. 43-438 è la fase di apprendimento «(anche nota con il termine inglese *training*), ossia un insieme di parametri numerici da utilizzare nella successiva fase di esecuzione». Questa componente è inintelligibile, mentre l'altra «componente fondamentale» è costituita «da un codice sorgente» scritto in un determinato linguaggio di programmazione.

acquisizioni dell'esperienza passata, per compiere predizioni e classificazioni, o addirittura, nel caso di *large generative AI model*, per creare nuovi dati, come immagini, testi e audio²⁰. L'avanzamento tecnologico è così rapido, che si assiste al diffondersi di algoritmi strutturati su reti di apprendimento elaborate sulla scorta della «rete neuronale del cervello umano»²¹, la cui peculiarità è mantenere una logica «sfumata», cioè non vincolata da criteri di verità o di falsità²².

Il tratto caratteristico di questi algoritmi è l'opacità, la quale deriva da un funzionamento del tutto imprevedibile, imperscrutabile, nonché da una programmazione effettuata con criteri spesso non comprensibili neppure dall'ingegnere che ne ha curato la strutturazione²³. I dati raccolti dagli algoritmi vengono inoltre elaborati in procedimenti non visibili (c.d. *black box*)²⁴, in cui non si assiste ad alcun intervento umano. Si è, dunque, dinanzi a un *iter* di calcolo non è ricostruibile²⁵, il cui *output* – che «può essere dei più diversi tipi»²⁶ – è comunque idoneo a fungere da base informativa di condizionamento della decisione umana.

2.1. ... con il rischio di incorrere in condotte pregiudizievoli

Alla luce delle caratteristiche fin qui descritte, è chiaro che quando l'IA coinvolge l'attività pubblica l'esigenza di trasparenza, preposta al buon andamento dell'amministrazione, si accentua. E si accentua anche in ragione delle svariate criticità che l'uso degli algoritmi presenta. Tra queste, si annovera spesso il problema delle discriminazioni algoritmiche, causate da «sistemi di IA che sbagliano»²⁷ e che, al pari di ogni discriminazione, «determinano trattamenti svantaggiosi per ragioni di razza, sesso e altre caratteristiche personali protette»²⁸ giuridicamente. Come accennato, ciò può derivare dall'utilizzo di dati incompleti, obsoleti o *biased*²⁹, nonché da «calcoli matriciali effettuati dall'IA nelle *black box*». Tali calcoli

²⁰ Negli ultimi anni sono emersi modelli ancora più avanzati di IA, ovvero i c.d. “*large generative AI model*”, che si distinguono dai modelli di *machine learning* più tradizionali in quanto questi ultimi «*are designed to make predictions or classifications*» in P. HACKER, A. ENGEL, M. MAUER, *Regulating ChatGPT and other Large Generative AI Models*, in *arxiv.org*, 6, 2023.

²¹ Si tratta delle reti neurali. Per una spiegazione si v. A. LIEVORE, *Self driving cars: pericoli alla prossima curva? Profili assicurativi, etici e giuridici della mobilità intelligente*, in *Trento BioLaw selected student paper*, 2018-2019, p. 8.

²² S. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *Riv. di BioDiritto*, 1, 2019 p. 79.

²³ S. SIMONCINI, *L'algoritmo incostituzionale*, cit., p. 78.

²⁴ Sul problema della *black box* v. R. GUIDOTTI, A. MONREALE, S. RUGGIERI, D. PEDRESCHI, F. TURINI, F. GIANNOTTI, *Local Rule-Based Explanations of Black Box Decision Systems*, in *arxiv.org*, 2018, pp. 1-10.

²⁵ F. FALLETTI, *Decisioni automatizzate e diritto alla spiegazione: alcune riflessioni comparatistiche*, in *Dir. Inf.*, 2, 2020, pp. 169-175.

²⁶ G. CARULLO, *Decisione amministrativa*, cit., p. 437 spiega che la decisione algoritmica può trattarsi «di un valore booleano (*true-false* [...]), ma può anche consistere in un testo discorsivo, od un qualsiasi diverso modo di presentare dati secondo le esigenze del caso».

²⁷ G. GOMETZ, *Intelligenza artificiale profilazione e nuove forme di discriminazione*, in *Teoria e storia del diritto privato*, NS, 2022, cit., p. 5.

²⁸ G. GOMETZ, *Intelligenza artificiale*, cit., p. 14.

²⁹ *Ibid.*, p. 4.

«non si limitano a ripetere le esperienze già immagazzinate, ma le ricombinano», con il «il rischio di perpetuare i *bias* e i possibili elementi discriminatori in esse presenti»³⁰.

Il termine *bias* oggi, diversamente dal passato, in cui era usato in modo per lo più neutrale per indicare una deviazione da uno *standard*, assume un'accezione negativa, in quanto indica qualcosa che deve essere evitato e di necessariamente problematico³¹.

La letteratura di diverse discipline³² ha articolato una tassonomia delle categorie e sottocategorie di *bias*. In generale, la stessa suole distinguere tra i c.d. *bias* cognitivi dalle ipotesi in cui gli errori del sistema IA derivano dall'utilizzo di dati incompleti.

La prima espressione è usata per riferirsi agli errori derivanti da percezioni sbagliate e deformate a causa di pregiudizi o ideologie³³. I pregiudizi, da sempre presenti in ogni contesto sociale, possono essere il frutto di schemi tipici della mente umana, «derivanti anche dalla necessità di velocizzare le reazioni»³⁴ per adeguarsi ai cambiamenti o rispondere al verificarsi di eventi, oppure possono essere il prodotto di rappresentazioni cognitive e impressioni elaborate con riguardo a determinati soggetti. In sostanza, questi ultimi, anziché essere percepiti individualmente, sono giudicati in ragione dell'appartenenza a una specifica categoria. Gli psicologi sociali si riferiscono a un simile processo mentale, spesso alle origini di comportamenti discriminatori, parlando di stereotipi³⁵.

Ecco che, queste inclinazioni distorsive della realtà, già di per sé «perniciose nel lungo termine»³⁶, se riprodotte dai sistemi di IA, possono subire una progressiva accentuazione³⁷, sia per l'oscurità del funzionamento degli algoritmi, sia per l'elevato livello di fiducia spesso riposto su di essi. Tali pregiudizi sono perpetrabili negli algoritmi dal suo programmatore se, inconsciamente o volontariamente (per esempio, per la presenza di ragioni di profitto)³⁸, ne struttura i modelli in modo discriminatorio. A partire

³⁰ F. FALLETTI, *Decisioni automatizzate*, cit., p. 175.

³¹ D. DANKS, A.J. London, *Algorithmic Bias in Autonomous Systems*, in *JCAI*, 2017, p. 2.

³² Come spiega R. XENIDIS, *Tuning EU equality law to algorithmic discrimination: Three pathways to resilience*, in *Maastricht Journal of European and Comparative Law*, 27(6), 2020, p. 737 «algorithmic bias' has been the subject of a growing strand of literature in various disciplines like computer science, ethics, social sciences and law».

³³ L. ALEXANDER, *What makes wrongful discrimination wrong? Biases, preferences, stereotypes, and proxies*, in *University of Pennsylvania Law review*, 141/149, 1992, pp. 157-165.

³⁴ Per una ricostruzione più dettagliate delle diverse classificazioni v. S. La Rosa, *Automation Bias. La sostituzione del giudizio e altri istupidimenti*, in *MaCSIS Working Paper Series*, 1, 2020, p. 8.

³⁵ A. FORZA, G. MENEGONI, R. RUMIATI, *Il giudice emotivo. La decisione tra ragione ed emozione*, il Mulino, Bologna, 2017, pp. 108-109.

³⁶ S. LA ROSA, *Automation Bias. La sostituzione del giudizio e altri istupidimenti*, p. 8.

³⁷ G. GIORGINI PIGNATIELLO, *Il contrasto alle discriminazioni algoritmiche: dall'anarchia giuridica alle Digital Authorities?*, in *federalismi.it*, 16, 2021, p. 167.

³⁸ Il riferimento va a quelle ipotesi in cui volontariamente il programmatore è mosso da logiche economiche di massimizzazione dei profitti. Per esempio, con riguardo alla progettazione di veicoli dotati di guida autonoma è stato evidenziato come «Key stakeholders in the AV value-chain can be motivated by perverse incentives to design AV algorithms in ways to favour the safety of certain groups of road users over others» H. SI MIN LIM, A. TAEIHAGH, *Algorithmic Decision-Making in AVs: Understanding Ethical and Technical Concerns for Smart Cities*, in *Sustainability*, 11, 2019, p. 19.

dall'ambiente sociale di riferimento, i pregiudizi e le percezioni errate sono cioè assorbite dalla struttura dell'algoritmo per alimentare il funzionamento della macchina.

In alternativa, un algoritmo in teoria neutrale può essere messo in moto da dati incompleti perché, per esempio, attinenti a un contesto non interamente rappresentativo della popolazione³⁹. Oppure, l'insieme di dati raccolti dalla macchina può derivare da analisi statistiche che, in quanto basate sulla probabilità⁴⁰, non sono necessariamente fedeli alle singole situazioni concrete⁴¹.

Si comprende, dunque, che le premesse per trattamenti discriminatori si pongono nella fase di immissione dei dati nell'algoritmo o, ancor prima, nella fase della sua programmazione. Questo perché lo sviluppatore dall'algoritmo può articolarne il funzionamento a partire da specifiche caratteristiche della persona, appartenente anche a categorie protette e suscettibile di essere stereotipata; oppure, perché il *set* di dati utilizzato è fin dall'origine incongruo, con la conseguenza che, in ragione della logica digitale *garbage in – garbage out*, sarà inesatto anche l'*output*.

Il risultato ultimo di tale processo è che gli errori della macchina prendono il posto degli errori umani, nonostante proprio la macchina si fosse prefissata l'obiettivo di risolverli⁴².

3. Il perpetuarsi di *bias* nelle città

Elevati sono gli sforzi compiuti per permettere alla transizione digitale di progredire, malgrado, al contempo, sia crescente la consapevolezza delle sue ripercussioni negative. Questo ambivalente atteggiamento di fiducia e di timore verso la tecnologia si palesa nelle *smart cities* di cui, da un lato, ne vengono elogiate l'efficienza, la resilienza, la capacità di contrastare il cambiamento climatico e di aumentare la qualità della vita⁴³; mentre, dall'altro lato, si sono mosse diverse voci critiche che ne sottolineano gli effetti distopici, tra i quali l'inasprimento delle diseguaglianze⁴⁴.

³⁹ H. SI MIN LIM, A. TAEIHAGH, *Algorithmic Decision-Making in AVs*, cit., p. 7; G. GOMETZ, *op.*, cit., p. 14 rileva che si tratta di «situazioni che risultano da errori degli (o negli) algoritmi, le cui prestazioni scadono 'selettivamente' con riguardo ai soggetti appartenenti a certe minoranze».

⁴⁰ È in sostanza sempre necessario considerare i dati partendo «dal significato che loro stessi hanno nel rappresentare la realtà», con la consapevolezza che «il dato raccolto nel dominio dei fenomeni sociali non descrive in maniera precisa la realtà ma rappresenta informazioni in qualche modo legate alla realtà» come sostenuto da F. MAGGINO, G. CICERCHIA, *Algoritmi, etica e diritto*, in *Dir. Inf.*, 6, 2019, p. 1161.

⁴¹ S. TOMMASI, *Algoritmi e nuove forme di discriminazioni: uno sguardo al diritto europeo*, *Rev. Dir. Brasileira*, 10, 2021, p. 115.

⁴² S. LA ROSA, *Automation Bias*, cit., p. 15.

⁴³ La letteratura in materia è molto vasta. Pertanto, si rinvia a una recente pubblicazione tratta dalla letteratura straniera che compie una ricognizione degli autori che si sono espressi in termini positivi sul tema segnalando la capacità delle *smart cities* di aumentare la qualità della vita dei cittadini e di favorire la crescita economica: J. SANCHEZ GRACIAS, G. S. PARNELL, E. SPECKING, E. A. POHL, R. BUCHANAN, *Smart Cities—A Structured Literature Review*, in *Smart cities*, 6, 2023, p. 1732. A livello nazionale si segnala G. FRANCO FERRARI, *Innovazione e sostenibilità per il futuro delle smart cities*, Mimesis, Milano, 2023.

⁴⁴ A. CARAGLIU, C. F. DEL BO, *Smart cities and urban inequality*, in *Regional Studies*, 56, 2022, p. 1098; S. POOLE, *The truth about smart cities: "In the end, they will destroy democracy"*, in *theguardian.com*, 2014; J. REMES, *Can smart cities be equitable?*,

Tali opposte conseguenze derivano sempre dal medesimo punto di partenza, ovvero dalla raccolta di un enorme mole di dati⁴⁵ al fine di ottimizzare, in linea di principio, qualsiasi ambito della città: per accrescere la sicurezza in chiave di polizia predittiva⁴⁶, per la pianificazione urbana, ad esempio, con l'uso dei gemelli digitali (c.d. *digital twin*)⁴⁷, ma anche per il riciclo intelligente dei rifiuti o per ridurre la congestione del traffico, per stimare il numero degli incidenti stradali e prevenirli, per introdurre *smart grid*, per efficientare l'energia degli edifici, per compiere un monitoraggio ambientale, per sviluppare servizi di maggior gradimento per i cittadini alla luce delle preferenze da loro espresse tramite *report* o, ancora, per fornire supporto clinico ai pazienti⁴⁸.

Anzitutto, il ricorso senza precedenti alla tecnologia dell'informazione e della comunicazione, a seconda del contesto sociale, economico e istituzionale di riferimento⁴⁹, può acuire le disparità tra cittadini, per esempio, tra coloro che possiedono le competenze di usare gli strumenti digitali e coloro che, invece, per ragioni economiche, di età, percorso di studio ne sono privi. Oppure, le differenziazioni sono destinate a emergere a causa della progettazione – in un'ottica di investimento per lo sviluppo urbano – delle tecnologie più sofisticate solo nel centro città, con la conseguenza che esse non possono essere sfruttate in eguale misura da tutti i residenti, in particolare, dalle persone meno abbienti poste ai margini della città⁵⁰.

Peraltro, le condizioni diseguali correlate all'accesso alla tecnologia si accentuano a causa delle discriminazioni perpetrate direttamente dagli algoritmi, le quali derivano da un'acquisizione senza precedenti di dati di vario tipo. Addirittura, in dottrina, si è parlato, con riguardo a talune città che si presentano come un agglomerato di sensori distribuiti per le strade, di «*technocratic rhetoric*»⁵¹, per riferirsi a quella svolta distopica che «*could take humans out of the loop and turn them into passive rather than active agents*»⁵².

in *Brookings.edu*, 2018; N. LEE, A. R. POSE, *Innovation and spatial inequality in Europe and USA*, in *Journal of Economic Geography*, 13,2013, pp. 1-22.

⁴⁵ Come spiega E. PALMERINI, *Dalle smart cities allo scoring del cittadino*, in *I confini del digitale. Nuovi scenari per la protezione dei dati*, Atti del Convegno, 2019, p. 21, nelle *smart cities* «l'accento è posto soprattutto sui rischi intrinseci allo sviluppo di grandi capacità di raccolta, aggregazione e uso dei dati».

⁴⁶ Per un approfondimento in materia di polizia predittiva si rinvia a M. B. ARMIENTO, *La polizia predittiva come strumento di attuazione amministrativa delle regole*, in *Dir. amm.*, 4, 2020, pp. 990 ss.

⁴⁷ E. PENNACCHIA, F. CINQUEPALMI, *Twin digital cities, la vera 'intelligenza' della città digitale nel XXI secolo*, in *Contesti. Città, Territori, Progetti*, 1, 2020, pp. 52–61.

⁴⁸ P. D. KÖNIG, *Citizen-centered data governance in the smart city: From ethics to accountability*, in *Sustainable Cities and Society*, 75, 2021, p. 2. V. anche European Parliament, *Artificial Intelligence and Urban Development*, 2021, p. 20.

⁴⁹ N. LEE, A. R. POSE, *Innovation and spatial inequality*, p. 2. Gli autori affermano «*These processes will depend significantly on the particular socio-economic and institutional contexts*».

⁵⁰ Compie un'analisi dettagliata della letteratura A. CARAGLIU, C. F. DEL BO, *Smart cities and urban inequality*, cit., pp. 1097-1101.

⁵¹ Fino a giungere al rischio di un «ricatto tecnocratico»: C. ACOCELLA, *Città intelligente e diritti: nuove prospettive di consumo nel prima della socialità*, in *PA persona e amministrazione*, 2, 2021, p. 113.

⁵² M. P. RODRÍGUEZ BOLÍVAR, *Smart Cities: Big Cities, Complex Governance?*, in M. P. RODRÍGUEZ BOLÍVAR, *Transforming City Governments for Successful Smart Cities* (Ed.), Springer, 2015, p. 4.

Ebbene, proprio il fatto che nelle *smart cities* il funzionamento di determinati sistemi sembra «prescindere dalla sorveglianza umana», che in ogni caso «non sarebbe svolta dall'amministrazione, ma dai privati che sviluppano e gestiscono le piattaforme»⁵³, può avere notevoli ripercussioni sulla capacità di valutare in modo neutrale, *ex ante*, la qualità dei dati immessi nell'algoritmo e di rimuovere, *ex post*, gli errori prodotti dalla macchina. Nell'era dei *big data*, si deve considerare che la raccolta, l'integrazione e l'utilizzo in maniera massiccia dei dati funzionale alla progettazione, alla gestione urbana nonché all'erogazione dei servizi, avviene spesso a opera di aziende private le cui finalità ultime non sono necessariamente conosciute dagli attori pubblici⁵⁴. Questo aspetto potrebbe essere idoneo a incrinare la strumentalità delle decisioni all'interesse generale, nonché la trasparenza dell'azione amministrativa in assenza di un'adeguata sorveglianza da parte del decisore pubblico sull'operatore economico. Ciò è ancor più vero se si considera la possibilità di incorrere in errori, finanche discriminazioni, a causa dell'inconsapevolezza circa la modalità di funzionamento degli algoritmi dei sistemi di IA programmati da soggetti terzi e poi utilizzati dalla pubblica amministrazione.

3.1. Alcuni esempi tratti dall'esperienza straniera

Per comprendere la portata del problema e le modalità in cui lo stesso si presenta, è utile considerare alcuni esempi tratti dall'esperienza straniera.

Un primo motivo per cui si creano *output* errati deriva dal c.d. «*signal problem*», ovvero la convinzione che i «dati riflettano accuratamente il tessuto sociale, e invece presentano lacune significative». Nel 2013, Crawford spiegava che a Boston, per risolvere un problema di buche per strada, veniva utilizzato l'approccio intelligente, cioè una specifica applicazione per smartphone (StreetBump) che «attinge[va] ai dati dell'accelerometro e del GPS per rilevare passivamente le buche» e «segnalarle» nell'immediato⁵⁵. Sennonché, un simile meccanismo, per funzionare, non poteva che servirsi dei soli dati delle persone munite di *smartphone*, lasciando indietro i soggetti a basso reddito, ma anche le persone anziane, con minore probabilità di esserne in possesso. L'applicazione accettava una potenziale situazione di *digital divide*⁵⁶ idonea a consolidare la vulnerabilità di soggetti già emarginati, ma anche di determinare una

⁵³ E. CHITI, B. MARCHETTI, N. RANGONE, *L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali*, in *Riv. BioDiritto*, 2, 2022, p. 504.

⁵⁴ C. E. KONTOKOSTA, B. HONG, *Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions*, in *Sustainable Cities and Society*, 64, 2021, p. 2.

⁵⁵ S. LA ROSA, *op. cit.*, p. 16.

⁵⁶ L. ATTIAS, M. MELCHIONDA, M. PIACITELLI, A. RUGGIERO, *Mind the gap*, in *Riv. elet. dir. eco., man.*, n. 3-2017, pp. 94-103 affermano: «con l'espressione “*digital divide*” ci si riferisce al divario esistente fra singoli individui, gruppi sociali, imprese e Stati nazionali, in riferimento all'accesso, reperimento e fruibilità delle informazioni, all'utilizzo consapevole delle nuove tecnologie, alla concreta disponibilità di adeguati strumenti tecnologici, all'interazione conscia con i *social network*, all'impiego responsabile dei social media ed al mantenimento di un'efficiente capacità di avvalersi, soprattutto tra i professionisti del digitale, delle tecnologie innovative e delle potenzialità della società dell'informazione».

risoluzione non pienamente efficiente del problema delle buche stradali, lasciando le condizioni urbane più periferiche sprovviste di manutenzione. Così, una volta scoperte queste ripercussioni, le autorità locali incaricarono «degli addetti per ispezionare tutte le zone», lasciando che i dati degli utenti divenissero delle mere «informazioni addizionali»⁵⁷.

Similmente, nel 2014, altri Autori analizzavano un servizio di *e-government* (sistema 311⁵⁸) volto a consentire ai residenti di città americane come New York, San Francisco, Washington, di chiedere assistenza per la riparazione dai parchimetri, dei marciapiedi, delle strade o della segnaletica stradale⁵⁹. Si trattava di un sistema di co-produzione teso a stimolare l'interazione tra governatori e residenti, ma suscettibile di sottovalutare gli effetti di una scarsa rappresentazione di soggetti che, non necessariamente per scelta⁶⁰, erano impossibilitati a presentare richiesta di assistenza con mezzi tecnologici. Tali effetti, oltre a riverberarsi sull'efficacia degli interventi urbani, comportano l'esclusione dal processo decisionale di una parte dei residenti.

In entrambi i casi presentati, le conseguenze pregiudizievoli derivano dalla fruizione non eguale degli strumenti digitali, nonché dalla contestuale raccolta incompleta di dati da parte degli algoritmi.

Vi sono ipotesi in cui, invece, l'*output* dell'algoritmo può essere il frutto della scarsa qualità del *data set* su cui è stato programmato l'algoritmo, ma anche di *bias cognitivi* del *designer* o degli *stakeholders* del prodotto digitale considerato. Esemplicative sono le tecnologie dei veicoli autonomi, ovvero quelle vetture che si guidano senza bisogno dell'intervento umano, ideate per aumentare la sicurezza, l'efficienza del traffico, la mobilità per gli anziani e i disabili, nonché ridurre l'impatto ambientale. Già nel 1939, le auto elettriche telecomandate confluivano nell'immaginazione dei pensatori delle città future⁶¹ e, oggi, si sono sempre più diffuse tra i progetti delle case automobilistiche, ma anche delle aziende informatiche (prime fra tutte Google)⁶², al fine di favorire una transizione *smart* della mobilità urbana.

e dei nuovi dispositivi, al trascorrere del tempo, al mutare del contesto sociale, culturale e politico ed in seguito a variazioni delle condizioni lavorative o economiche». Per una ricognizione dei problemi connessi al *digital divide* v. F. FAINI, S. PIETROPAOLI, *Scienza giuridica e tecnologie informatiche*, Giappichelli, Torino, 2017, p. 84; P. TARALLO (a cura di), *Digitale divide. La nuova frontiera dello sviluppo globale*, Franco Angeli, Milano, 2003.

⁵⁷ E. PALMERINI, *Dalle smart cities allo scoring del cittadino*, cit., p. 21.

⁵⁸ Che è passato dall'essere un numero telefonico per contattare la polizia a un servizio telefonico per ricevere assistenza e servizi urbani in J. WINBOURNE, *311 Services – A Key Step in the Global Transition to e-Government Services*, in *winbourne.it*.

⁵⁹ C. E. KONTOKOSTAA, B. HONG, *Bias in smart city governance: How socio-spatial disparities in 311 complaint behavior impact the fairness of data-driven decisions*, in *Sustainable Cities and Society*, 64, 2021, p. 2.

⁶⁰ Per esempio, in un recente studio, è stato sottolineato come la partecipazione da parte dei residenti nel processo co-decisionario dipende molto dall'impegno politico di ciascuno con la conseguenza che il sistema 311 potrebbe non assicurare una piena partecipazione civica in A. WHITE, K. S. TRUMP, *The Promises and Pitfalls of 311 Data*, in *Urban Affairs Review*, 54, 2018, 794-823.

⁶¹ «Nel 1939 Norman Bel Geddes, un *designer* teatrale e industriale americano, espose all' Expo di New York "Futurama": un'installazione che rappresentava come avrebbe potuto essere la configurazione urbanistica delle città americane nel prossimo futuro» e in cui erano «raffigurate auto elettriche radiocomandate». in A. LIEVORE, *Self driving cars: pericoli alla prossima curva? Profili assicurativi, etici e giuridici della mobilità intelligente*, cit., p. 2.

⁶² A. LIEVORE, *Self driving cars: pericoli alla prossima curva?*, cit., pp. 5-7.

I problemi si pongono perché le *self driving cars* funzionano sulla base di algoritmi *deep learning* che acquisiscono una grande mole di dati sulla base di un «apprendimento per rinforzo»⁶³, ovvero un processo automatizzato della realtà circostante che avviene con telecamere, sensori, GPS e, soprattutto, con il c.d. *Computer Vision*, riproduttivo della vista umana. Tuttavia, la modalità di funzionamento del veicolo autonomo non è esattamente uguale a quella di un uomo, in quanto dipende dai *data set* con cui l'algoritmo è stato addestrato, *set* di dati che non necessariamente raffigura in modo completo e accurato ogni possibile scenario del traffico di una città. Per esempio, il *data set* potrebbe non comprendere determinate persone perché assenti al momento della ripresa effettuata dai sensori. Ciò provoca il peggioramento della *performance* del veicolo, ma anche rischi alla sicurezza «quando a essere coinvolti sono i sistemi di rilevazione del pedone»⁶⁴. Oppure, risultati non totalmente rappresentativi della popolazione di riferimento sono suscettibili di derivare da correlazioni matematiche, cui vengono sottoposti i dati per l'apprendimento della macchina che, in ragione di schemi mentali del programmatore o per ragioni di profitto del produttore, privilegiano o penalizzano determinate caratteristiche personali dell'autista. Al di là delle ripercussioni sui modelli assicurativi esistenti, nonché degli interrogativi circa l'individuazione di colui a cui imputare la responsabilità nel caso di incidenti, l'impossibilità di intervenire nella struttura del procedimento seguito dalla macchina apre ulteriori questioni connesse a problemi di carattere etico: si tratta della necessità di decidere verso quale soggetto indirizzare una condotta lesiva quando il danno è inevitabile⁶⁵. In particolare, nel caso di collusione inevitabile, qual è il criterio alla base della scelta compiuta dal veicolo autonomo? Dal momento che questi mezzi funzionano in ragione della modalità di programmazione, occorre sottolineare come i loro produttori, traendo profitto dalla maggiore sicurezza degli acquirenti, potrebbero essere indotti ad allocare il rischio di una minor sicurezza sui soggetti terzi⁶⁶. È un'eventualità su cui occorre riflettere in quanto sono i soggetti privati a governare la

⁶³ A. LIEVORE, *op. cit.*, p. 9.

⁶⁴ G. GROTTI, *Veicoli autonomi: viaggio tra algoritmi ed etica*, tesi di laurea in ingegneria meccanica, 2022, pp. 27-28. In particolare, presenta l'esempio della incapacità del veicolo autonomo di riconoscere tutte le tipologie di pedoni a causa della mancata rilevazione, al momento della programmazione dell'algoritmo, di pedoni di giovane età. Tale discriminazione sulla base dell'età anagrafica dipendeva dal fatto che le riprese erano state condotte quando i bambini e gli adolescenti erano a scuola. Sul punto si rinvia anche a B. WILSON, J. HOFFMAN, J. MORGENSTERN, *Predictive Inequity in Object Detection*, in *arxiv.org*, 2019, pp. 1-13.

⁶⁵ Si tratta di una questione vicina al noto "Trolley Problem" esaminata da A. LIEVORE, *op. cit.*, pp. 45-46 e a H-Y LIU, *Three Types of Structural Discrimination Introduced by Autonomous Vehicles*, in *UC Davis Law Review Online*, 51, 2018, pp. 149-180.

⁶⁶ Per una trattazione più specifica delle situazioni di vantaggio che potrebbero essere introdotte dai produttori dei veicoli in questione, per ragioni di carattere economico, si rinvia a H. SI MIN LIM, A. TAEIHAGH, *Algorithmic Decision-Making in AVs: Understanding Ethical and Technical Concerns for Smart Cities*, in *Sustainability*, 18, 2019, pp. 1-28.

realizzazione di mezzi preordinati a tutelare la sicurezza (in questo caso) stradale⁶⁷. Del resto, le logiche a guida degli investimenti delle compagnie private non possono che essere di carattere economico.

Un altro esempio, estrapolabile dal settore della mobilità urbana, proviene dalla città di Amsterdam⁶⁸. Per rendere la città più vivibile e accessibile ai ciclisti e ai pedoni, è stato ridotto il numero di macchine autorizzate a parcheggiare in città con un servizio basato su un algoritmo *machine learning*, diffuso in più di 150.000 parcheggi. Per verificare l'eventuale ingresso in città di soggetti non autorizzati, tale sistema interviene, in via preliminare, processando i dati ricavati tramite sensori e telecamere; dopodiché, vengono identificate le varie targhe delle vetture e, all'esito del processo automatizzato, l'algoritmo è in grado di determinare la validità dei diritti dei parcheggiatori.

La particolarità deriva dal fatto che il sistema in questione viene reputato a basso rischio per i diritti dei cittadini, ma comunque, qualora il risultato non sia positivo, ovvero sia accertata l'assenza di una previa autorizzazione, si attiva una fase in cui interviene un ispettore umano chiamato a controllare le immagini acquisite e verificare se vi siano ragioni per non ritenere corretto l'*output* prodotto dalla macchina. Successivamente, se l'*output* risulta privo di errori, l'ispettore può richiedere alle autorità amministrative competenti di irrogare una sanzione, in caso contrario, può rinviare i dati alla macchina per sottoporli a una procedura depurata da errori. All'esito di questa procedura, deve seguire, nell'arco di 48 ore, l'eliminazione dei dati assemblati scorrettamente.

Nonostante l'*iter* seguito dall'algoritmo non sia comprensibile dall'ispettore, la previsione di una forma di controllo umano rappresenta un presidio essenziale per verificare che il procedimento sia giunto a un *output* corretto. È invero fondamentale scongiurare eventuali discriminazioni fondate su determinate caratteristiche dei titolari della vettura (come, per esempio, il modello della sua auto, l'età, o generale, il suo profilo) sulla base delle quali sia stato allenato l'algoritmo. E se, da una parte, può non essere comprensibile la modalità di raccolta ed elaborazione dei dati per l'opacità dell'algoritmo, d'altra parte, il controllore umano può stabilire a posteriori se sono presenti motivi validi – che non fanno leva sulle caratteristiche personali dell'utente – per parcheggiare in città. Questa verifica permette di discostarsi dal risultato incongruo, così che, a prescindere dalla modalità con cui l'algoritmo ha processato le informazioni, l'errore prodottosi sia superabile. Ciò, però, presuppone l'assenza di atteggiamenti lassisti, timorosi, o di estrema fiducia, nei confronti di quello che per molti è il nuovo «oracolo», la «fonte

⁶⁷ A. CARAGLIU, C. F. DEL BO, *Smart cities and urban inequality*, cit., p. 1100 illustrano, tra le varie cause di disuguaglianze perpetrate nelle *smart cities*, il coinvolgimento di soggetti privati nell'implementazione di *policies* e progetti poiché tenderanno a incentivare investimenti in TIC con lo scopo principale di massimizzare i profitti.

⁶⁸ N. SAWHNEY, *Contestations in urban mobility: rights, risks, and responsibilities for Urban AI*, in *AI & SOCIETY*, 38, 2023, p. 1091, che rinvia alla pagina web del servizio olandese "City of Amsterdam Algorithm Register".

automatica di verità»⁶⁹, ovvero la tecnologia. Ed è per tale motivo che le autorità locali di Amsterdam cercano di contrastare simile problematiche consentendo al privato interessato, accedendo a un apposito servizio *online*, di presentare un reclamo scritto entro sei settimane dall'irrogazione della sanzione.

3.2. Trattare della *smart mobility* in Italia mettendosi al riparo dai rischi

Anche in Italia molte città stanno sviluppando sistemi di IA a supporto del processo decisionale delle amministrazioni per efficientare il settore della mobilità e aumentare la sostenibilità ambientale, basandosi sul monitoraggio in tempo reale della vita dei cittadini, tramite la raccolta dei loro dati personali. Le città italiane stanno cioè sempre più seguendo la tendenza europea di dotarsi di *urban data platform* o di applicazioni funzionali all'evoluzione intelligente dei trasporti⁷⁰.

Per esempio, la città di Bologna ha avviato, a partire dall'11 maggio 2022 e per la durata di sei settimane, l'*app* IoPollicino, che è fondata su meccanismi innovativi di *e-government*. Essa vuole rappresentare uno strumento in mano ai residenti, con età maggiore di 16 anni, per condividere spontaneamente e in forma anonima i dati sui propri spostamenti, sì da contribuire all'elaborazione futura di politiche di mobilità sostenibili.

Altra città capofila nell'elaborazione di un sistema di mobilità urbana intelligente è Venezia che, a partire dal 2020, ha creato, con un'intensa collaborazione con aziende private (Tim e Abertis), una *Smart Control Room*. Si tratta di una centrale operativa dove confluiscono in tempo reale una pluralità di dati raccolti con sensori dislocati in varie zone della città. Parte di questi dati «sono prodotti dalla città», «altri sono forniti dai soggetti terzi di natura privata»⁷¹ sopra menzionati. Gli scopi perseguiti da tale strumento sono il monitoraggio dei flussi in entrata e in uscita – con riguardo sia al transito delle imbarcazioni nei canali e dei mezzi pubblici o privati, che degli abitanti e dei turisti –, la definizione della disponibilità dei parcheggi e, infine, l'attivazione dell'intervento delle forze dell'ordine.

Anche l'Aquila ha innovato il proprio sistema di mobilità mirando a creare una infrastruttura interconnessa in cui le informazioni siano condivise e coordinate, tramite l'utilizzo di applicazioni, con finalità, oltre che di condizionamento di scelte «*eco-friendly*», di prevenzione e predizione⁷².

Ebbene, nel primo caso, i dati raccolti per elaborare le politiche urbane non necessariamente appartengono a tutti i residenti in possesso delle caratteristiche anagrafiche richieste per partecipare

⁶⁹ C. MILANI, V. GARCÍA, *L'evoluzione dell'Intelligenza Artificiale: dall'automazione del lavoro al condizionamento reciproco*, in *Mondo Digitale*, 2023, p. 9.

⁷⁰ In questa direzione si segnalano, oltre agli esempi che verranno illustrati nel paragrafo, anche Firenze e Roma nell'ambito dei rispettivi Piani di Azione per Energia Sostenibile e Clima.

⁷¹ E. CHITI, B. MARCHETTI, N. RANGONE, *op. cit.*, p. 500 che trattano dell'esempio ora presentato.

⁷² M. AUTILI, A. DI SALLE, F. GALLO, C. POMPILIO, M. TIVOLI, *A Choreography-Based and Collaborative Road Mobility System for L'Aquila City*, in *Future Internet*, 11, 2019, p. 9 da cui è tratto l'esempio.

all'attività di *report*. Potrebbero discenderne effetti analoghi a quelli presentati da Crawford nel 2013, nel senso che le soluzioni sostenibili andrebbero a privilegiare quelle aree urbane in cui risultano essere stati effettuati “maggiori spostamenti”, senza che però ciò sia totalmente fedele alla realtà concreta. Pur vero è, che, nel caso di specie, il servizio intende – in modo esplicito – basarsi sulla spontaneità della partecipazione, addossando sul singolo le implicazioni di una decisione non integralmente rappresentativa della popolazione residente.

Tuttavia, il miglioramento della qualità urbana – e contestualmente della qualità della vita dei residenti – dovrebbe andare a favore di ciascun cittadino, con la consapevolezza che vi sono condizioni tali da impedire una contribuzione volontaria all'attuazione di detto servizio.

Negli ultimi due esempi, il problema può invece derivare da *bias* cognitivi, inconsci o meno che siano. Invero, il controllo dei flussi degli abitanti, dei turisti, dei mezzi di trasporto in entrata e uscita, così come lo svolgimento dell'attività di predizione e prevenzione, potrebbe avvenire in modo distorto, ovvero, privilegiando condizioni di tipo economico o sociale, ma anche caratteristiche legate alla razza o al sesso. Se da una parte si tratta di ipotesi al limite ed estreme, ma che per la loro assoluta meschinità devono essere prevenute in anticipo con ogni mezzo, dall'altro, risulta complesso scongiurare simili effetti una volta prodottisi. In tal senso, si rilevarebbe utile un intervento di un operatore chiamato a garantire un'eguale distribuzione dei servizi, a ridurre i divari esistenti, dunque a compiere un controllo preliminare sulla qualità e rappresentatività dei dati, nonché a correggere gli *output* falsamente rappresentativi del contesto urbano di riferimento o, addirittura, perpetuanti modelli storici di discriminazioni. Ciò garantirebbe che valori umani fondamentali, quali l'eguaglianza, siano reintrodotti in un percorso logico, rigido e automatizzato come quello su cui si basa l'algoritmo.

A tal fine è necessario che, da un lato, il controllore abbia effettivo interesse a rimuovere le situazioni di diseguaglianza create dalla macchina; dall'altro, occorre ricercare un supervisore totalmente neutrale, il cui modo di pensare non sia affetto da *bias* cognitivi, che, come accennato, sono l'esito di processi mentali umani, prima ancora che artificiali.

Dinanzi a questi profili controversi, la soluzione maggiormente prospettata in dottrina, in giurisprudenza e, da ultimo, anche dal nuovo Codice dei contratti pubblici (d.lgs. 36/2023)⁷³ è infatti il c.d. *human in the*

⁷³ Cons. Stato, sez.VI, 4.02.2020, n. 881; Cons. Stato sez. VI, 8.09.2021, n. 1206. Molteplici sono ormai gli scritti sul tema, si rinvia a titolo esemplificativo a E. FALLETTI, *Mai accettare caramelle né atti amministrativi da sconosciuti, ancorché algoritmi, nota a TAR Campania, Napoli, 14.11.2022, n. 7003*, in *Dir. Inf.* 1, 2023, pp. 91-107; B. MARCETTI, *La garanzia dello human in the loop alla prova della decisione amministrativa algoritmica*, cit., pp. 370-372; E. CARLONI, *I principi della legalità algoritmica. Le decisioni automatizzate di fronte al giudice amministrativo*, in *Dir. Amm.*, 2, 2020, pp. 271 ss. R. FERRARA, *Il giudice amministrativo e gli algoritmi. Note estemporanee a margine di un recente dibattito giurisprudenziale*, in *Dir. Amm.*, 4, 2019, pp. 773 ss.; peraltro, anche l'art. 30 del nuovo Codice dei Contratti Pubblici (d.lgs. 36/2023), dopo aver espresso il *favor* per l'utilizzo di strumenti digitali da parte delle stazioni appaltanti, impone che le decisioni assunte con sistemi di IA siano conoscibili e comprensibili (co. 3), che rispettino il principio di non esclusività della decisione algoritmica (co. 3) e, infine, che non diano luogo a discriminazioni (co. 3 e co. 4). Sul punto si rinvia al recente contributo di L. IANNOTTA,

loop⁷⁴. Esso consiste nella garanzia di un «contributo umano capace di controllare, validare ovvero smentire la decisione automatica»⁷⁵e, assieme al principio di non discriminazione algoritmica, nonché alla necessaria piena conoscibilità a monte del modulo utilizzato e dei criteri applicati, rappresenta un requisito per giustificare il diffuso favore al ricorso di strumenti digitali.

La non esclusività della decisione algoritmica rappresenta così al contempo «un limite al potere amministrativo», in grado di porre il privato in «una posizione di *immunità*, cioè di sottrazione al potere automatizzato dell'amministrazione»⁷⁶, e un criterio guida per lo sviluppo *smart* del contesto urbano ogniqualevolta l'attore pubblico assuma degli atti in grado di incidere sulla posizione dei cittadini.

4. Le ambizioni europee nella regolazione dello sviluppo tecnologico

La garanzia di un procedimento amministrativo non automatizzato, espressa in primo luogo dalla giurisprudenza nazionale e destinata a riverberarsi nella gestione delle città, trae la sua origine dal diritto europeo. Peraltro, essa rileva dei momenti di intersezione tra branche del diritto diverse, derivanti dalla sovrapposizione del piano strettamente pubblicistico con quello privatistico: essendo i dati personali la fonte primaria del funzionamento degli algoritmi, è inevitabile il confronto con la disciplina preposta alla loro salvaguardia. Non è un caso se, per definire i principi guida della decisione pubblica algoritmica, lo spunto principale sia provenuto proprio dal GDPR (Reg. 2016/679).

Si tratta della base di partenza a cui oggi si aggiunge la normativa europea di settore tesa a disciplinare in modo armonizzato i sistemi di IA⁷⁷, assumendo «i dati come risorsa fondamentale dell'economia e della società digitale»⁷⁸, nonché sfruttando, con flessibilità e consapevolezza, le opportunità offerte dallo sviluppo tecnologico. Il riferimento è cioè all'elaborazione di un «pacchetto digitale UE»⁷⁹ che, nell'aprile del 2021, ha condotto ad avanzare la Proposta di Regolamento sull'intelligenza artificiale (c.d. Ai Act).

Decisioni algoritmiche e valutazione dell'offerta: la digitalizzazione nel settore dei contratti pubblici, tra strumenti digitali e contributo umano, in *federalismi.it*, 5, 2024, pp. 33-60.

⁷⁴ Peraltro, quando si ha a che fare con decisioni algoritmiche amministrative è stata sottolineato che profili psicologici, più che giuridici, possono giocare un ruolo fondamentale; in particolare un atto confezionato con un sistema di IA potrebbe essere accettato con più difficoltà rispetto a un atto per il quale si assiste all'intervento umano: così G. CARULLO, *op. cit.*, p. 460.

⁷⁵ A. SIMONCINI, *L'algoritmo incostituzionale: intelligenza artificiale e il futuro delle libertà*, in *Riv. BioDiritto*, 1, 2019, p. 79.

⁷⁶ E. FRAGALE, *La cittadinanza amministrativa al tempo della digitalizzazione*, in *Dir. amm.*, 2, 2022, pp. 496-497.

⁷⁷ Per una ricognizione sul punto si rinvia a C. SILVANO, *Prospettive di regolazione della decisione amministrativa algoritmica: un'analisi comparata*, in *Riv. it. di dir. pub.*, 2/3, 2022, pp. 265. Ma anche come mette in luce C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, *cit.*, p. 416. G. FINOCCHIARO, *La proposta di regolazione sull'intelligenza artificiale: il modello europeo basato sulla gestione del rischio*, in *Dir. inf.* 2, 2022, pp. 303-322.

⁷⁸ G. RESTA, *La regolazione digitale nell'Unione Europea – Pubblico, privato, collettivo nel sistema europeo di governo dei dati*, in *Riv. trim. dir. pub.*, 4, 2022, p. 971.

⁷⁹ Insieme al Digital Service Act, al Digital Market Act e il Data Governance Act come rilevato da G. RESTA, *La regolazione digitale nell'Unione Europea*, *cit.*, p. 971.

Essa «disciplina l'immissione in commercio e l'utilizzo degli algoritmi di intelligenza artificiale in una prospettiva di compatibilità con il quadro dei diritti fondamentali europei»⁸⁰.

Dei passi fondamentali all'avanzamento dell'AI Act sono stati impressi dal Parlamento Europeo nel 2023 con l'approvazione di alcuni emendamenti e, poi, con il raggiungimento di un accordo su una bozza di Proposta tra il Consiglio e i negoziatori del Parlamento lo scorso dicembre; infine, con l'approvazione, da parte del Parlamento, dell'accordo raggiunto con gli Stati Membri lo scorso 13 marzo⁸¹. Il testo, che deve ancora entrare definitivamente in vigore, andrà a completare il contenuto del General Data Protection Regulation (GDPR), dal quale ha avuto origine l'obiettivo europeo di proteggere i dati degli utenti del *web*.

4.1. Una riflessione a partire dal General Data Protection Regulation

Il «punto di partenza obbligato»⁸² dell'intera riflessione è dunque il GDPR, al cui interno si scorgono disposizioni a tutela dei diritti dei titolari dei dati sottoposti a processi automatizzati, ossia della persona fisica identificata o identificabile (c.d. «interessato»). Secondo quanto disposto dall'art. 4. n. 4, la specifica categoria di trattamenti automatizzati sui «dati personali», e volta a «valutare determinati aspetti personali relativi a una persona fisica [...]», prende il nome di «profilazione».

Tanto premesso, la norma principe, apparente «freno a possibili invasioni e predominanze dell'algoritmo sulle scelte dell'uomo»⁸³, è l'art. 22 co. 1. Questa statuisce 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». La disposizione, per lungo tempo rimasta nell'ombra, per poi essere esaminata copiosamente dalla dottrina⁸⁴, «consacra il principio della non esclusività della decisione algoritmica»⁸⁵, importato solo di recente nell'ambito pubblicistico. Essa sancisce cioè il presidio dell'*human in the loop*, ogniqualvolta le decisioni algoritmiche incidano «sul libero esercizio di un diritto della persona riconosciuto dall'ordinamento giuridico»⁸⁶.

⁸⁰ G. RESTA, *La regolazione digitale nell'Unione Europea*, cit., p. 971.

⁸¹ Consultabile al link https://www.europarl.europa.eu/doceo/document/A-9-2023-0188-AM-808-808_IT.pdf.

⁸² T. E. FROSINI, *L'orizzonte giuridico dell'intelligenza artificiale*, in *Dir. inf.*, 1, 2022, p. 14.

⁸³ T. E. FROSINI, *L'orizzonte*, cit., p. 11.

⁸⁴ Per un'analisi si rinvia anche a G. MALGIERI, *"Just" Algorithms: Justification (Beyond Explanation) of Automated Decisions Under the General Data Protection Regulation*, in *Law and Business*, 1, 2021, pp. 16-28. E. Palmerini, *Decisioni algoritmiche e diritto dei dati*, in *giudicedonna.it*, 1/2, 2023, pp. 1-20.

⁸⁵ M. Sforza, *Le garanzie di conoscibilità degli algoritmi e l'esigenza di assicurare un controllo umano del procedimento amministrativo (c.d. human in the loop)*. (Nota a Tar Campania, Sez. III, 14 novembre 2022, n. 7003), in www.giustiziasieme.it, 2023.

⁸⁶ E. Troisi, *AI e GDPR, L'automated decision making, la protezione dei dati e il diritto all'intelligibilità dell'algoritmo*, in *European Journal of Privacy Law & Technologies*, 1, 2019, p. 48.

Tuttavia, ciò avviene con «un raggio di tutela non particolarmente esteso»⁸⁷ per una pluralità di ragioni concomitanti.

Anzitutto, la lettera della disposizione ha favorito ricostruzioni restrittive. Alcune di queste ritengono che in ragione dell'uso dell'avverbio “unicamente” siano sussumibili nella fattispecie solo i trattamenti integralmente automatizzati⁸⁸; altre fanno invece leva sull'avverbio “significativamente”, per accentuare la portata degli effetti pregiudizievoli che l'interessato deve aver subito per poter avanzare una pretesa meritevole di tutela; peraltro, il riferimento “alla sua persona” ha determinato l'esclusione dal campo d'azione della norma dei gruppi composti da più soggetti. Si aggiunga, poi, che il riferimento alle sole “decisioni”, rende estranee alla fattispecie le frequenti ipotesi in cui gli strumenti di IA servono a «operare una rappresentazione o una classificazione della realtà»⁸⁹.

In secondo e ultimo luogo, è lo stesso co. 2 dell'articolo a prevedere una serie di ampie eccezioni⁹⁰. Nello specifico, la garanzia non si applica se la decisione «sia necessaria per la conclusione o l'esecuzione di un contratto tra l'interessato e un titolare del trattamento», oppure quando il trattamento «si basi sul consenso esplicito dell'interessato». Ed è in forza di questo ultimo punto che copiosa dottrina ha segnalato la debolezza della norma, rilevando come il consenso sia spesso il frutto di automatismi o comunque di valutazioni non pienamente consapevoli circa le ragioni del trattamento e le sue possibili conseguenze⁹¹.

Peraltro, ai sensi del paragrafo 2, lett. b) dell'articolo in esame, il trattamento automatizzato può addirittura prescindere dalla necessità del consenso dell'interessato se vi sia una norma autorizzativa del diritto dell'Unione o dello Stato Membro cui è soggetto il titolare del trattamento.

Un ulteriore aspetto che limita la portata dell'art. 22 deriva dalla penuria di norme esplicitamente mirate a scongiurare un effetto di influenzamento del supervisore da parte della macchina. Ne deriva che, il complesso compito rimesso al controllore rischi di degradare a un ruolo di rilevanza solo formale. La norma non verrebbe violata da un atteggiamento passivo del supervisore in quanto essa non impone alcun «potere realmente ‘valutativo’» sul funzionamento dell'algoritmo⁹².

Del resto, un simile rischio non è scongiurato neppure dall'art. 5 co. 1 l. d), che richiede al titolare del trattamento di essere in grado «di adottare tutte le misure ragionevoli per cancellare o rettificare

⁸⁷ E. PALMERINI, *Decisioni*, cit., p. 4.

⁸⁸ E. TROISI, *L'automated decision making*, p. 47.

⁸⁹ E. PALMERINI, *Decisioni*, cit., p. 4.

⁹⁰ Sui rapporti tra art. 22 e artt. 13 e 14 GDPR v. G. COMANDÈ, *Leggibilità algoritmica e consenso al trattamento dei dati personali, note a margine di recenti provvedimenti sui dati personali*, in *Danno e resp.*, 2022, 2, pp. 141 ss. che evidenzia comunque anche «il rovescio della medaglia» dato da un'eccessiva trasparenza.

⁹¹ Si tratta del problema del c.d. *blind consent* come spiegato da C. CASONATO, B. MARCHETTI, *Prime osservazioni sulla proposta di regolamento dell'Unione Europea in materia di intelligenza artificiale*, in *Riv. BioDiritto*, 3, 2021, p. 429.

⁹² G. MALGIERI, G. COMANDÈ, *Right to legibility of automated decision-making*, in *International Data Privacy Law*, 7, 2017, n. 4; E. TROISI, *op. cit.*, p. 47.

tempestivamente i dati inesatti rispetto alle finalità per le quali sono trattati». Rilevare la sussistenza di un'inesattezza, peraltro in termini immediati, presuppone, quale condizione imprescindibile per l'efficacia dell'intervento umano, autonomia di giudizio.

A tale profilo si aggiunge la complessità di individuare quale sia l'errore prodotto dalla rielaborazione di un cospicuo ammontare di dati immessi all'interno del sistema, specie alla luce del fatto che tale errore deve essere individuato in relazione alla finalità perseguita dal trattamento.

Così, l'effetto errato, essendone ignote le ragioni di base, in quanto non esplicitabili e difficilmente palesabili all'esito del trattamento, oltre a non essere compreso, anche quando venisse eliminato, si potrebbe presentare in casi analoghi. Ciò è destinato ad avere un impatto notevole nel contesto urbano dove a un medesimo processo automatizzato lesivo sono sottoposti una pluralità di residenti. Il tutto è coronato dal fatto che il considerando 71 del GDPR tratta dell'esigenza che il titolare del trattamento «metta in atto misure tecniche e organizzative per rettificare i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori», ma senza stabilire quali debbano essere «i requisiti o parametri minimi di attendibilità dei risultati»⁹³.

Al problema dell'oscurità dell'algoritmo⁹⁴ cercano di rimediare gli artt. 13 e 14, e 15, prevedendo degli obblighi di informativa, e il contestuale diritto di accesso, sul modo di funzionamento del medesimo algoritmo⁹⁵. L'obiettivo ultimo è assicurare comprensibilità e trasparenza dell'algoritmo, conferendo una serie di informazioni relative alle finalità del trattamento dei dati personali. Specifica, per esempio, la l. h) dell'art. 15 che l'accesso deve strutturarsi garantendo una serie di informazioni *significantive* sulla logica utilizzata, nonché sull'importanza e sulle conseguenze di tale trattamento per l'interessato. Deve quindi essere un'informazione «sui passaggi inferenziali che hanno portato la macchina a quell'*output*».

Questo tipo di informativa, nella logica del GDPR, dovrebbe permettere all'interessato di dialogare con il titolare del trattamento, allorché il primo decida di opporsi all'utilizzo automatizzato dei suoi dati

⁹³ Entrambe le citazioni sono di G. GOMETZ, *op. cit.*, p. 34. A ciò si aggiunga, come messo in luce da E. PALMERINI, *Decisioni algoritmiche*, cit., p. 10, che per evitare di dover spiegare nel dettaglio la modalità di funzionamento di un algoritmo *black box* vengono scelti algoritmi più semplici, dunque, meno validi e più suscettibili di incorrere in errori.

⁹⁴ E. TROISI, *op. cit.*, 56.

⁹⁵ E. TROISI, *op. cit.*, p. 54. Si consideri peraltro che, partendo da una lettura congiunta delle norme richiamate, il Consiglio di Stato ha espresso il principio secondo cui deve essere individuabile «un centro di imputazione e di responsabilità riferibile a una persona “in grado di verificare la legittimità e logicità della decisione dettata dall'algoritmo». Un contributo umano che «deve comunque esistere nel processo decisionale [...] capace di controllare, validare ovvero smentire la decisione automatica» così come spiegato da molti autori, tra cui 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 *Riv. Biodiritto*, 3, 2020, pp. 389-440; E. FALLETTI, *op. cit.*, pp. 190-191; A. SIMONCINI, *op. cit.*, p. 74 che ripercorre uno dei primi casi sottoposti all'attenzione del Consiglio di Stato in materia, riguardante la mobilità degli insegnanti nella L. 107.2015 in cui sono stati affermati principi importanti. Tra questi, quello in base al quale «l'algoritmo, decidendo di fatto la destinazione dell'insegnante, sostanzia il procedimento amministrativo che dà vita all'atto e, dunque, dà diritto all'accesso ai sensi della legge 241/1990». Si v. altresì, C. SILVANO, *Prospettive di regolazione della decisione amministrativa algoritmica: un'analisi comparata*, in *Riv. it. dir. pubb. comp.* 2,3, 2022, pp. 265 ss.

fruendo del potere messogli a disposizione dall'art. 21. Anche in questo caso, però, il diritto è destinato a soccombere in modo variabile se il «titolare del trattamento dimostri l'esistenza di motivi legittimi e cogenti» per procedere all'assunzione della decisione automatizzata. Tali motivi, secondo il disposto della norma da ultimo citata, devono prevalere «sugli interessi, sui diritti e sulle libertà dell'interessato», oppure essere «legati all'accertamento, all'esercizio o alla difesa di un diritto in sede giudiziaria».

4.2. Alcune considerazioni alla luce della Proposta di Regolamento sull'intelligenza artificiale

Quelle finora tratteggiate sono solo alcune delle problematiche che emergono dal trattamento automatizzato dei dati e dall'impianto normativo del GDPR. Un passo in avanti potrebbe essere compiuto dalla recente Proposta della Commissione di dotare l'Unione Europea di un Regolamento, dalla portata orizzontale, sull'intelligenza artificiale. Già i considerando della Proposta, nell'illustrare le maggiori sfide della digitalizzazione, ribadiscono aspetti di notevole rilevanza, tra cui il fatto di dotarsi di dati di elevata qualità, specie quando si è dinanzi a sistemi il cui funzionamento potrebbe diventare «fonte di una discriminazione vietata dal diritto dell'Unione»⁹⁶.

Così, con l'intento di creare uno strumento malleabile, in grado di adattarsi alle incessanti evoluzioni dell'era digitale, la Commissione compie una distinzione destinata a riflettersi sull'ambito di applicazione delle disposizioni del Regolamento. In particolare, distingue la disciplina in base al livello di rischio (modello c.d. *risk based*) del sistema IA considerato, che può essere inaccettabile, alto, basso o minimo⁹⁷. Mentre gli ultimi due sono individuati in via residuale, la prima categoria di rischi è definita nella misura in cui il sistema «contrasta in radice con i valori dell'Unione»⁹⁸, mentre la seconda classificazione è oggetto di trattazione specifica nel Titolo III del Regolamento.

Ebbene, con riguardo all'esigenza di prevenzione e di correzione degli *outcomes biased* le maggiori risposte all'interno della Proposta si riscontrano proprio nel Titolo III che⁹⁹, all'art. 6, identifica il sistema ad alto rischio per essere «componente di sicurezza di un prodotto, o [per essere] esso stesso un prodotto» indicato nell'allegato II della Proposta (co. 1 lett. a)); nonché per essere soggetto a una «valutazione della

⁹⁶ Considerando (67) Emendamenti del Parlamento Europeo alla proposta di Regolamento della Commissione che stabilisce regole armonizzate sull'intelligenza artificiale e modifica i regolamenti (CE) n. 300/2008, (UE) n. 167/2013, (UE) n. 168/2013, (UE) 2018/858, (UE) 2018/1139 e (UE) 2019/2144 e le direttive 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828.

⁹⁷ Peraltro, una simile flessibilità sarebbe garantita anche grazie alla previsione di un «obbligo di revisione a cinque anni dall'entrata in vigore» del Regolamento C. CASONATO, B. MARCHETTI, *op. cit.*, 419.

⁹⁸ Regolamento del Parlamento Europeo e del Consiglio che stabilisce regole armonizzate sull'Intelligenza Artificiale (Legge sull'Intelligenza Artificiale) e modifica alcuni atti dell'Unione, COM (2021) 206 final, p. 14. Per visionare il testo aggiornato a seguito dell'approvazione degli emendamenti si rinvia al sito ufficiale del Parlamento Europeo.

⁹⁹ Con riguardo, invece alla tutela processuale prestabile nell'Unione Europea a fronte di *outcomes* in grado di produrre discriminazioni lesive occorre ricercare le soluzioni in fonti diverse. A tal proposito, si rinvia a R. XENIDIS, *Tuning EU equality law to algorithmic discrimination: Three pathways to resilience*, in *Maastricht Journal of European and Comparative Law*, 6, 2020, pp. 736 -758 oppure G. GOMETZ, *Intelligenza artificiale, profilazione e nuove forme di discriminazione*, cit., pp. 1-41.

conformità da parte di terzi in relazione ai rischi per la salute e la sicurezza ai fini dell'immissione sul mercato o della messa in servizio» (co. 1 lett. b)); infine, «per rientrare in uno o più settori critici e casi d'uso di cui all'allegato III, se presentano un rischio significativo di danno per la salute umana, la sicurezza o i diritti fondamentali delle persone fisiche» (co. 2).

In primo luogo, l'art. 10, norma specificatamente dedicata alla *governance* dei dati, fissa in otto lettere le misure da adottare – compatibilmente allo specifico segmento di mercato, al campo di applicazione e a quanto risulta fattibile tecnicamente – per soddisfare i criteri di qualità dei dati. La disposizione è già di per sé idonea a enfatizzare l'importanza di un intervento umano *ex ante*, ovvero nella fase di *training* dell'algoritmo. Essa fa leva sul compimento di operazioni tese alla valutazione preliminare della qualità dei dati, all'arricchimento e aggiornamento per evitare distorsioni, lacune o carenze rilevanti, nonché, non dissimilmente da quanto previsto GDPR, alla necessità della trasparenza dell'algoritmo in relazione alla finalità originaria di raccolta dei dati.

La ricerca dell'adeguatezza qualitativa dei dati si completa con le norme dedicate alla trasparenza (art. 13) e alla sorveglianza umana (art. 14)¹⁰⁰.

La prima disposizione citata richiede che «i sistemi di IA ad alto rischio siano progettati e sviluppati in modo tale da garantire che il loro funzionamento sia sufficientemente trasparente da consentire ai *deployer* di interpretare l'*output* del sistema e utilizzarlo adeguatamente». La norma prosegue richiedendo che la trasparenza sia adeguata a «conseguire il rispetto dei pertinenti obblighi del fornitore e del *deployer*».

Per scongiurare interpretazioni restrittive del primo paragrafo, alimentabili dagli avverbi “sufficientemente”, il paragrafo terzo spiega nel dettaglio come garantire la comprensibilità del sistema di IA. Sono previste una serie di informazioni tecniche funzionali a rendere interpretabile l'*output* del sistema di IA sia dall'utente che dal fornitore, tra cui le caratteristiche tecniche, la *performance* in relazione a specifiche persone o gruppi di persone, le informazioni dei dati immessi o altri aspetti rilevanti per comprendere il *training* dell'algoritmo. Queste informazioni, peraltro, devono essere coincise, complete, corrette, chiare, pertinenti, accessibili e comprensibili per il *deployer*.

In tal modo, la previsione vale a specificare quanto dispone l'art. 15 GDPR, ribadendo la necessità che la trasparenza sia intesa come «*smart disclosure*», ovvero come effettiva comprensibilità della logica e dei meccanismi di funzionamento dell'algoritmo»¹⁰¹, da non considerarsi in sé e per sé, ma in relazione all'*output* prodotto più che alla finalità dello stesso. Tuttavia, anche la portata di questa norma può essere compromessa dalla subordinazione di simili garanzie a valutazioni di opportunità in relazione al caso specifico.

¹⁰⁰ Commentato, tra i vari autori, da G. FINOCCHIARO, *La proposta di regolazione sull'intelligenza artificiale*, cit., p. 315.

¹⁰¹ Così M. MACCHIA, A. MASCOLO, *Intelligenza artificiale e regolazione*, in A.P. F. DONATI (a cura di), *Intelligenza artificiale e diritto: una rivoluzione?*, Bologna, il Mulino, 2022, p. 124.

Alla disposizione sulla trasparenza si affianca l'art. 14 sulla sorveglianza umana, presidio non più relegato a un breve inciso di una disposizione, bensì descritto puntualmente con lo scopo di «prevenire o ridurre al minimo i rischi per la salute, la sicurezza o i diritti fondamentali». A tal fine, «i sistemi di IA ad alto rischio sono progettati e sviluppati, anche con strumenti di interfaccia uomo-macchina adeguati, in modo tale da essere efficacemente supervisionati da persone fisiche durante il periodo in cui sono in uso».

Ne deriva che l'*human oversight* è commisurato al livello effettivo di rischio, al livello di autonomia e al contesto di utilizzo del sistema di IA. Di conseguenza, la sua applicazione non è circoscritta – come invece previsto dalle ipotesi di cui all'art. 22 GDPR – alle sole ipotesi di trattamenti interamente automatizzati. Inoltre, l'assenza di un riferimento individualizzato ai diritti del soggetto interessato al trattamento consente una potenziale applicazione della disposizione a tutela dei gruppi di persone.

La novità di maggior rilievo si scorge poi nell'indicazione dei presupposti di partenza per evitare forme «di distorsione dell'automazione» (paragrafo 4 lett. b)), connessi all'esigenza che il controllore sia in grado «di comprendere correttamente le capacità e i limiti pertinenti del sistema di IA, di monitorarne debitamente il funzionamento, anche al fine di individuare e affrontare anomalie, disfunzioni e prestazioni inattese (lett. a)». La persona fisica a cui viene demandata la sorveglianza deve «rimanere consapevole del rischio di poter fare affidamento automatico sull'*output* prodotto». Pertanto, deve sapere «interpretare correttamente l'*output*» (lett. c)), «decidere, in qualsiasi situazione particolare, di non usare il sistema [...] o altrimenti di ignorare, annullare o ribaltare l'*output*» (co. 4 lett. d)), nonché «intervenire sul funzionamento del sistema di IA ad alto rischio» (lett. e)).

Queste importanti acquisizioni richiedono un intervento sostanziale del sorvegliante umano, e compiono così un decisivo passo in avanti rispetto al GDPR. Affinché però esse non si traducano in mere indicazioni di principio, i singoli Stati Membri dovranno effettuare una scelta ponderata nell'individuazione delle persone fisiche a cui conferire un simile ruolo senza sminuire l'esigenza di garantire elevate competenze tecniche e informatiche.

5. *Smart cities e human oversight*: il complesso ruolo del soggetto pubblico

Calare le disposizioni fin qui esaminate nella realtà urbana intelligente ci pone dinanzi alla necessità di comprendere a chi sia demandato il ruolo di sorvegliante umano. Questo interrogativo per essere affrontato richiede di interfacciarsi con un quadro giuridico in materia di *smart cities* consistente solo una vasta documentazione di *soft law*¹⁰², che oltre a non conferire una definizione univoca di questo modello di città, non predispone ancora strumenti specifici a tutela dei cittadini. In sostanza, posta l'assenza di

¹⁰² Per una ricognizione delle varie Dichiarazioni, Comunicazioni e altri atti di *soft law* adottati dall'Unione in relazione allo sviluppo urbano e, più di recente, allo sviluppo *smart* si rinvia a E. Carloni, M. VAQUERO PIÑEIRO, *Le città intelligenti e l'Europa. Tendenze di fondo e nuove strategie di sviluppo urbano*, in *IdF*, 4, 2015, pp. 865-894.

una normativa di settore dedicata al tema, l'entrata in vigore di un Regolamento dalla portata trasversale in ambito tecnologico potrà assumere una notevole rilevanza.

A ogni modo, malgrado il loro carattere non cogente, è comunque importante sottolineare che gli atti assunti dall'Unione in materia di *smart cities* lasciano emergere lo spirito di fondo nei confronti del processo di digitalizzazione, in quanto si concentrano sulla necessità di compiere il passaggio dal *technology-driven approach*¹⁰³, per lungo tempo caratterizzante le *smart cities*, a un approccio *human-based*. Un passaggio ritenuto doveroso per recidere il legame con la prima fase di progettazione delle *smart cities*, dove gli sforzi si concentravano sulla digitalizzazione massima dei servizi e delle infrastrutture, nell'ottica di favorire la progressione economica soprattutto delle *Tech Companies*¹⁰⁴. L' *human centric paradigm*¹⁰⁵ presuppone il massimo coinvolgimento dei cittadini nel processo decisionale, l'inclusione sociale e loro l'eguaglianza¹⁰⁶, utilizzando strumenti affidabili che prevengano usi distorti della tecnologia.

E' un approccio emergente dalla definizione di città intelligente fornita da ultimo dalla Commissione Europea, in base alla quale «*a smart city is a place where traditional networks and services are made more efficient with the use of digital solutions for the benefit of its inhabitants and businesses.*»¹⁰⁷. E soprattutto, tale modalità di pensare allo sviluppo urbano *smart* si allinea con l'evoluzione più recente della regolazione della tecnologia, volta a far sì che la transizione digitale rimanga un mero “*key enabler*” per lo sviluppo (in questo caso urbano), senza assurgere al suo obiettivo principale.

In questo quadro, il problema dei *bias* e delle discriminazioni algoritmiche si rintraccia in maniera timida solo in qualche documento, per esempio, all'interno dell'European Interoperability Framework for Smart Cities and Communities” (EIF4SCC), adottato negli ultimi mesi del 2022. Questo, nella diciannovesima raccomandazione, richiede alle autorità locali di monitorare e controllare i codici *software*, per assicurare

¹⁰³ Per una ricostruzione dei passaggi evolutive delle *smart cities* in Europa si rinvia all'European Parliamentary Research Service, *Social approach to the transition to smart cities*, STUDY- Panel for the Future of Science and Technology, pp. 5-6, dove si parla di quattro generazioni di *smart cities*.

¹⁰⁴ S. PAROUTIS, L. HERACLEOUS, *A strategic view on smart city technology: The case of IBM Smarter Cities during a - recession*, in *Technological Forecasting and Social Change*, 89, 2014, pp. 262-272; S. POOLE, *The truth about smart cities: In the end they will destroy democracy*, in *theguardian.com.*, mette in luce «*The notion of the smart city in its full contemporary form appears to have originated within these businesses, Greenfield notes in his 2013 book Against the Smart City, “rather than with any party, group or individual recognised for their contributions to the theory or practice of urban planning*» in Oppure si v. E. MOROZOV, F. BRIA, *Rethinking the Smart City. Democratizing Urban Technology*, in *rosalux.nyc.org*, 2018, pp. 5-8.

¹⁰⁵ S. TOPHAN, P. BOSCOLO, M. MULQUIN, *Personal Data-Smart Cities: How Cities Can Utilise Their Citizen's Personal Data to Help Them Become Climate Neutral*, River Publishers, Denmark, 2022, p. 21.

¹⁰⁶ European Parliamentary Research Service, *Social approach to the transition to smart cities*, cit., p. 6 con riferimento alla terza generazione di smart city alla quale sta progressivamente seguendo una terza dove acquisisce maggior rilievo la transizione verde.

¹⁰⁷ V. definizione contenuta nel sito *web* della Commissione Europea dedicato allo sviluppo regionale e urbano nella sezione “*smart cities*”.

un trasparente e responsabile utilizzo degli algoritmi, nonché evitare *bias* in grado di discriminare le persone¹⁰⁸.

Al di là della valenza meramente raccomandatoria, è di rilievo l'attribuzione di un simile compito all'amministrazione. Tale attribuzione, infatti, oltre a lasciare intravedere un apparente *favor* da parte dell'Unione al modello *top-down* di sviluppo intelligente dove sono, cioè, le singole amministrazioni a farsi carico del monitoraggio dell'impiego della tecnologia¹⁰⁹, consente di elevare le stesse amministrazioni a controllori dei sistemi di IA, secondo quanto richiesto dall'art. 14 della Proposta di Regolamento.

Si tratta di un ruolo che consentirebbe al soggetto pubblico di essere il garante ultimo dei diritti fondamentali dei cittadini, ma, che per la sua ambizione, dischiude immediatamente problemi legati al contesto nazionale dove la pubblica amministrazione può non mostrarsi in possesso dei presupposti tecnico-scientifici e finanziari necessari¹¹⁰. Sarà quindi semmai meno idealistico pensare a un maggior coinvolgimento di attori privati (come, per esempio, avvenuto nel caso di Venezia) tramite il ricorso a forme di partenariato pubblico-privato. A ciò si aggiunga l'esigenza di formare l'attore pubblico ad avere un atteggiamento non acritico, cioè distaccato e pronto a un riesame dei risultati raggiunti, nell'utilizzo del sistema IA; aspetto che può rendersi ancor più complesso se si considera come l'impiego della tecnologia sia funzionale proprio alla semplificazione e all'efficientamento degli apparati pubblici.

Infine, si pone il problema di capire come le disposizioni a presidio del controllo umano siano destinate a variare in ragione di possibili diverse interpretazioni del concetto di sistema ad alto rischio da parte di città tra loro eterogenee. Del resto, si è visto, come nel caso di Amsterdam, sebbene il rischio complessivo fosse definito basso dalle autorità locali, la sorveglianza umana costituiva parte integrante del complessivo funzionamento del servizio¹¹¹. Sul punto, quindi, le sfide saranno, da un lato, assicurare un'omogenea applicazione dei criteri di definizione del rischio – se effettivamente s'intende subordinare il controllo umano solo a questo presupposto – dall'altro, comprendere se le previsioni attualmente vigenti a presidio della correttezza della decisione algoritmica rimarranno tali a prescindere dal livello del rischio determinato dall'utilizzo di un certo sistema.

¹⁰⁸ Commission Staff Working Document, *European Interoperability Framework for Smart Cities and Communities* (EIF4SCC), SWD (2022) 710 final, p. 15, raccomandazione n. 19

¹⁰⁹ Secondo il modello *top-down* di matrice europea in G. PANATTONI, *Le Smart Connected Cities: i fattori di insicurezza e (s)fiducia dei cittadini*, in *Dir. Pubb.*, 2, 2023, p. 623; M. SFORNA, *Le garanzie di conoscibilità degli algoritmi e l'esigenza di assicurare un controllo umano del procedimento amministrativo, (c.d. human in the loop)*. (Nota a Tar Campania, Sez. III, 14 novembre 2022, n. 7003), in *giustizjainsieme.it* 2023.

¹¹⁰ E. CHITI, B. MARCHETTI, N. RANGONE, *L'impiego di sistemi di intelligenza artificiale nelle pubbliche amministrazioni italiane: prove generali*, cit., p. 493.

¹¹¹ V. *Automated parking control*, in *algoritmergister.amsterdam.nl*, sezione "Risk Management".

6. *L'human in the loop* per l'eguale trattamento dei cittadini: uno scenario plausibile?

Dalla regolazione fin qui esaminata è emerso come l'AI Act si sofferma con più chiarezza sul presidio dell'*human oversight* rispetto a quanto previsto dal GDPR nel breve inciso contenuto nell'art. 22, garantendone invero un'applicazione più ampia e consapevole. Soprattutto fa in modo che il controllo non si riduca a un mero requisito formale, definendo i contenuti che lo stesso deve assumere per rispettare l'art. 14. Ciò si inserisce nel complesso tentativo, esplicitato già a partire dall'art 10, di garantire la correttezza dei dati immessi nell'algoritmo, nonché l'interpretabilità di quest'ultimo non solo con riguardo ai passaggi interferenziali seguiti, ma anche con riguardo agli effetti prodotti.

È altresì emerso come il «doppio volto del paradigma digitale», particolarmente evidente nelle *smart cities*, si risolve nella ricerca di un approccio antropocentrico¹¹², ovvero nella garanzia che i mezzi digitali rimangano strumentali al miglioramento della qualità di vita delle persone, nonché nella garanzia di un intervento umano nei procedimenti che fanno uso di algoritmi *machine learning*.

All'esito della trattazione si è cercato di mettere in luce come questa, seppure apprezzabile, evoluzione di per sé non sia sufficiente. Si è invero evidenziato come molteplici tra le decisioni algoritmiche affette da pregiudizi e stereotipi perpetuate nell'ambito della *smart mobility* derivino da una pregressa situazione di divario sociale, situazione che tende ad amplificarsi con l'uso della tecnologia. Ciò rende necessario che la pianificazione urbana con mezzi tecnologici «si prenda cura delle ricadute che proprio l'utilizzo della tecnologia può avere in termini di coinvolgimento eguale dei cittadini»¹¹³.

Ne deriva, da un lato, l'ambiziosa necessità di un intervento umano scevro da pregiudizi, in grado di verificare la qualità e la completezza dei dati raccolti e immessi negli algoritmi, dall'altro, la necessità di superare situazioni di *digital divide*. Si è cioè dinanzi a un percorso che richiede un duplice sforzo: in via preliminare, è necessario il conferimento di un diritto all'eguale fruizione dei servizi – anche digitali – offerti dalla città e, dunque, l'eguale comprensione delle informazioni che permettono l'evoluzione *smart* della città medesima¹¹⁴.

Affinché il percorso possa completarsi si pone poi l'esigenza che *l'human in the loop* si elevi a strumento di controllo per l'eguale rappresentazione e inclusione di ogni cittadino nello spazio urbano¹¹⁵, a prescindere

¹¹² E. FRAGALE, *op. cit.*, p. 481 e pp. 486-487 l'A. parla di «trasformazione digitale antropocentrica, fondata sulla solidarietà e sull'inclusione, la libertà di scelta, la partecipazione allo spazio pubblico digitale, la sicurezza e la sostenibilità».

¹¹³ F. ALBERTI, *Nuovi Ecoquartieri smart. L'urbanistica della sostenibilità per territori digitali*, in *Atti della XVIII Conferenza nazionale SIU*, 2015, p. 575.

¹¹⁴ Deve cioè essere garantito, per riprendere la formula usata da C. ACOCELLA, *Città intelligente e diritti: nuove prospettive*, cit., p. 114, «la città come essenziale luogo di radicamento e di esercizio dei diritti». Ciò, a livello operativo, richiede che le Regioni, o più specificatamente le città, si dotino di Agende digitali per garantire livelli minimi di diritti (come già compiuto da alcune Regioni italiane).

¹¹⁵ F. FRACCHIA, P. PANTALONE, *Smart City: condividere per innovare (e con il rischio di escludere?)*, in *federalismi.it*, 22, 2015, p. 11.



dai dati che lo stesso veicola. L'intervento umano diviene molto più di un presidio per il corretto funzionamento dell'algoritmo se, integrandosi all'esigenza di contrastare il divario digitale, viene visto come chiave per un'eguale rappresentazione, dalle decisioni algoritmiche, di tutti i cittadini. Potremmo assumere che, nella misura in cui detto intervento serva a verificare la piena rappresentatività del *set* di dati immessi, nonché l'assenza di *bias* cognitivi, sia un tassello nel percorso per rimuovere gli «ostacoli di ordine economico e sociale» che, «limitando di fatto la libertà e l'eguaglianza dei cittadini, impediscono il pieno sviluppo della persona umana [...]» (art. 3 Cost.). L'*human oversight*, quindi, può essere lo strumento per individuare, correggere gli errori della macchina, e, al contempo, per evitare il diffondersi di decisioni che non tengano conto di valori quali la dignità e l'eguaglianza.

In questo senso, tale strumento viene messo anzitutto nelle mani dell'attore pubblico in ragione della stretta interconnessione che vi è tra trasparenza, controllo sulla decisione automatizzata per la rimozione di distorsioni dell'eguaglianza e tutela effettiva dei diritti dei cittadini nello spazio urbano. È invero l'amministrazione, anche alla luce delle indicazioni europee in materia di gestione dei dati nelle *smart cities*, a doversi attivare, e ciò in virtù della sua qualità di soggetto deputato alla cura dello spazio urbano nell'interesse della collettività. A tal fine, l'attore pubblico è chiamato a mostrare piena consapevolezza degli effetti e del funzionamento dei mezzi digitali adoperati, tanto nella fase di verifica dell'adempimento agli obblighi di cui agli artt. 13 da parte del fornitore del sistema di IA, che nel momento in cui compie il controllo circa l'insussistenza di discriminazioni o errori algoritmici.

Non si può concludere, però, senza rilevare come questi obiettivi, in apparenza lineari, rischino di non essere concretizzabili interamente in ragione della loro complessità. Non solo per le incertezze operative connesse all'approccio basato sul rischio e ai rapporti, ancora da definire, tra GDPR e Ai Act, ma anche e soprattutto perché è ancora fin troppo dubbia la capacità dell'amministrazione pubblica di svolgere un compito così ambizioso, che richiede significativi sviluppi in termini di risorse tecniche, economiche e finanziarie.