

FONDAZIONE GIANDOMENICO ROMAGNOSI
S C U O L A D I G O V E R N O L O C A L E

Note della Fondazione Giandomenico Romagnosi

Nota 18-2020

**Linee guida etiche per un'intelligenza
artificiale meritevole di fiducia pubblica:
molte questioni, incerte prospettive**

**Enrico Bagli, Stefano Regondi
e Giuseppe Carlo Ricciardi**

Settembre 2020

Fondazione Scuola di Governo Locale Giandomenico Romagnosi

Presidente: Andrea Zatti.

Responsabile Scientifico delle Note Brevi: Andrea Zatti.

Nota 18-2020, settembre 2020.

Autori: Enrico Bagli, Stefano Regondi e Giuseppe Carlo Ricciardi.

Linee guida etiche per un'intelligenza artificiale meritevole di fiducia pubblica: molte questioni, incerte prospettive.

Linee guida etiche per un'intelligenza artificiale meritevole di fiducia pubblica: molte questioni, incerte prospettive¹

di Enrico Bagli*, Stefano Regondi** e Giuseppe Carlo Ricciardi***

1. Potenzialità e rischi dell'Intelligenza Artificiale (IA). Un'introduzione.

I sistemi di intelligenza artificiale (d'ora innanzi "IA") stanno gradualmente intervenendo nella nostra vita quotidiana, promettendo di generare incrementi di produttività, di migliorare l'efficienza e di ridurre i costi dei processi produttivi².

Accanto alla definizione classica di J. Copeland³, negli ultimi tempi si fanno strada definizioni di IA che sottolineano gli effetti benefici delle applicazioni di *machine learning* anche sul potenziamento delle capacità umane⁴, corroborando un rapporto già di per sé reso evidente dalle applicazioni concrete dell'IA, come accaduto recentemente per la robot antropomorfa *Sophie*, cui è stata addirittura conferita la cittadinanza saudita⁵. La prospettiva di una nuova declinazione del rapporto sussistente tra uomo e robotica ovvero tra "personalità" della decisione umana ed automazione della decisione artificiale sottende un ripensamento della tassonomia di tali

¹ Il contributo è frutto della riflessione congiunta degli autori.

* Enrico Bagli è Data scientist presso CRIF S.p.A.

** Stefano Regondi è Direttore operativo e dello sviluppo dei Centri clinici NeMO e legale rappresentante di Nemo Lab.

*** Giuseppe Carlo Ricciardi è assegnista in Diritto amministrativo italiano e comparato presso l'Università di Pavia e membro del Comitato scientifico di Fondazione Romagnosi-Scuola di Governo Locale.

² OECD, *Artificial Intelligence in Society*, 2019, OECD Publishing, Paris, <https://doi.org/10.1787/eedfee77-en>

³ J. Copeland, *Artificial Intelligence: Philosophical Introduction*, New Jersey, 1993.

⁴ Accenture, *Why is artificial intelligence important?*, 2017, in https://www.accenture.com/_acnmedia/pdf-54/accenture-artificial-intelligence-ai-overview.pdf

⁵ B. Waters, *Citizen Sophia: It's (Past) Time to Legislate Robotics Regulation*, in *Georgetown Law Technology Review*, November 2017, in <https://georgetownlawtechreview.org/citizen-sophia-its-past-time-to-legislate-robotics-regulation/GLTR-11-2017/>

relazioni sia per quanto attiene alle organizzazioni private, sia per le pubbliche amministrazioni. Essa assume contorni difficilmente definibili *ex ante*, così come aprioristicamente non preventivabili risultano essere le innumerevoli potenzialità ed applicazioni concrete astrattamente percorribili nell'ambito della sfera pubblica e privata.

Accanto ai benefici attesi, tuttavia, l'IA ingenera nuove forme di perplessità giuridiche, etiche e in termini di aspetti equitativi. Infatti,

[...] i timori che l'IA possa essere in grado di sostituire e/o annullare decine di migliaia di posti di lavoro per gli esseri umani, o i timori relativi all'essere maltrattati da attori malintenzionati, o in alternativa le preoccupazioni inerenti l'elusione delle responsabilità previste dai Codici vigenti sono argomenti comparsi nelle pubblicazioni scientifiche o divulgative correlate all'argomento⁶.

Nel quadro complesso sinora delineato, molte Istituzioni pubbliche hanno ravvisato la necessità di individuare almeno alcuni principi-cardine alla luce dei quali sviluppare i sistemi di IA, al fine di offrire una regolamentazione di massima propedeutica alla prevenzione dei rischi connessi all'*empowerment* organizzativo ed applicativo intrinseco a tali tecnologie.

Tali principi sono stati sinora raccolti in testi di *soft law*, generalmente denominati "linee guida".

2. Linee Guida etiche per un'Intelligenza Artificiale meritevole di fiducia. Principi di base.

Nel novero dei numerosi contributi riscontrabili nell'ambito delle linee guida etiche esistenti per l'IA⁷, la Commissione europea ha definito una propria specifica visione volta a sostenere «un'IA etica, sicura e all'avanguardia prodotta in Europa»⁸. In ragione di ciò, la Commissione ha istituito un gruppo di lavoro composto da esperti di alto livello sull'intelligenza artificiale incaricato

⁶ Jobin, A., Ienca, M. & Vayena, E. *The global landscape of AI ethics guidelines*. Nat Mach Intell 1, 389–399 (2019). <https://doi.org/10.1038/s42256-019-0088-2>

⁷ *Ibid.*

⁸ COM(2018)237 e COM(2018)795. Da notare che il termine "made in Europe" è utilizzato nelle comunicazioni ufficiali della Commissione. L'obiettivo di queste Linee Guida non è unicamente quello di comprendere i sistemi di IA prodotti in Europa, ma anche quello di comprendere tutti quei Sistemi sviluppati ovunque ed implementati o usati in Europa.

di redigere due documenti, vale a dire gli orientamenti etici⁹ e le raccomandazioni politiche e di investimento¹⁰.

Per il gruppo di esperti, i sistemi di intelligenza artificiale devono focalizzarsi sulla centralità e sul ruolo primario dell'essere umano, con il dichiarato e programmatico obiettivo di incrementare il benessere della società mediante processi trasparenti e meritevoli di fiducia. In particolare, per essere "affidabile", l'IA deve essere lecita, vale a dire rispettare tutte le leggi e i regolamenti applicabili, nonché etica, ovvero assicurare l'adesione a principi e valori etici robusti, sia dal punto di vista tecnico che sociale. All'interno della figura 1 si presenta il quadro per un "IA degna di fiducia", con i quattro principi etici che derivano dalla tabella dei diritti fondamentali dell'Unione Europea¹¹ e con i sette requisiti chiave per realizzare l'IA e il passo finale necessario per valutare l'affidabilità di una specifica applicazione AI.

A seguito del piano coordinato sullo sviluppo e l'uso dell'IA realizzato in Europa, i Governi degli Stati appartenenti all'Unione Europea hanno iniziato a strutturare il proprio operato sull'intelligenza artificiale. In Italia, il Ministero dello sviluppo economico (MISE) ha recentemente pubblicato la "*Strategia italiana per l'Intelligenza Artificiale*"¹². Il documento presenta una visione di lungo periodo sull'intelligenza artificiale, nonché le misure suggerite per realizzare l'ambizioso progetto europeo finalizzato a «[...] diventare la regione leader a livello mondiale per lo sviluppo e la diffusione di un'IA all'avanguardia, etica e sicura, promuovendo un approccio incentrato sull'uomo nel mondo contestato».

All'interno del documento programmatico la Commissione preposta dell'Unione Europea redige i presupposti etico-giuridici per un'IA meritevole di fiducia, classificandoli in quattro categorie: **(i) Rispetto per l'autonomia umana, (ii) Prevenzione del danno, (iii) Buona fede, (iv) Giustificabilità.**

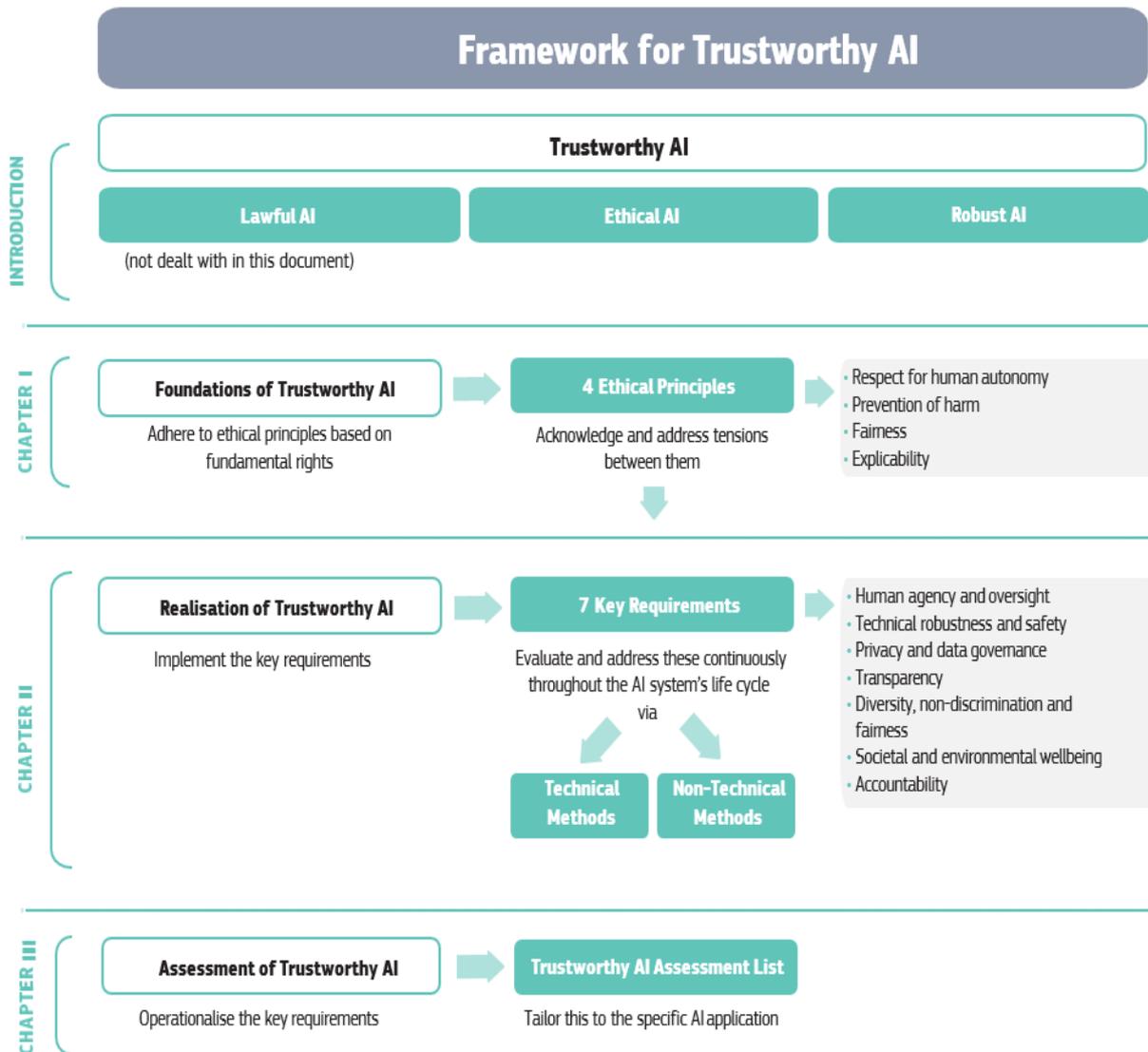
⁹ Gruppo di Esperti scientifici su IA promosso dalla Commissione Europea, *Ethics guidelines for trustworthy AI*, Brussels, 2019, <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

¹⁰ Gruppo di Esperti scientifici su IA promosso dalla Commissione Europea, *Policy and investment recommendations for trustworthy Artificial Intelligence*, 2019, Brussels, <https://ec.europa.eu/digital-single-market/en/news/policy-and-investment-recommendations-trustworthy-artificial-intelligence>

¹¹ Parlamento Europeo e Ufficio per le Pubblicazioni ufficiali della Comunità Europea, *Charter of fundamental rights of the European Union*, 2000, Luxembourg.

¹² MISE-Ministero dello Sviluppo Economico, *Proposte per una Strategia italiana per l'intelligenza artificiale*, luglio 2020, in https://www.mise.gov.it/images/stories/documenti/Proposte_per_una_Strategia_italiana_AI.pdf. Il documento completa il lavoro inaugurato con la bozza per la consultazione del 31 luglio 2019 MISE, *Strategia italiana per l'intelligenza artificiale*, luglio 2019, in <https://www.mise.gov.it/images/stories/documenti/Strategia-Nazionale-Intelligenza-Artificiale-Bozza-Consultazione.pdf>.

Figura 1: Le linee guida come diagramma di lavoro per una IA degna di fiducia.



Fonte: The Guidelines as a framework for Trustworthy AI (documento citato in referenza n.8, n.9).

La tassonomia proposta può essere meglio compresa alla luce di quanto segue.

(i) Il principio del rispetto per l'autonomia umana antepone il primato della persona nel porre il discrimine del proprio bene e del proprio male. In uno Stato di democrazia pluralista quale è il nostro, si rende necessario «[...] rispettare l'autonomia di un soggetto agente e riconoscere doverosamente le capacità e le prospettive della persona di fare determinate scelte e di prendere certe decisioni basate su convinzioni e valori personali»¹³. Qualora il soggetto non dovesse essere "agente",

¹³ T. L. Beauchamp, *Principi della bioetica: autonomia, beneficalità, giustizia*, in G. Russo (ed.), *Bioetica fondamentale e generale*, SEI, Torino, 1995, p. 84.

potranno essere adottati i criteri di soluzione proposti da Viafora C.¹⁴: a) Supposizione del miglior interesse per il soggetto passivo nella misura in cui l'IA adempie ai propri obiettivi costitutivi; b) Evocazione della storia passata del soggetto passivo per cercare di comprendere quale scelta egli farebbe in tale e specifica situazione se fosse in grado di elaborare ed enunciare la decisione.

(ii) Il principio della prevenzione del danno si può ben sintetizzare con l'espressione latina: *primum non nocere*, che da secoli rappresenta un brocardo caro ai codici deontologici di giuristi e sanitari. Tale brocardo enuncia il principio di beneficalità, il quale invita all'astensione dal cagionare il male e/o il danno a terzi, a prevenire il male e/o il danno a terzi, a promuovere la realizzazione di pratiche orientate al bene e non al male e/o al danno. Tale principio si concilia con la Costituzione della Repubblica italiana che all'articolo 32 recita:

La Repubblica tutela la salute come fondamentale diritto dell'individuo e interesse della collettività, e garantisce cure gratuite agli indigenti. Nessuno può essere obbligato a un determinato trattamento sanitario se non per disposizione di legge. La legge non può in nessun caso violare i limiti imposti dal rispetto della persona umana.

Esso rinviene un ancoraggio anche in alcune disposizioni del codice civile preordinate, al contrario, ad imputare la responsabilità per eventi dannosi già cagionati (artt. 2048 c.c., 2050 c.c., 2052 c.c.) ovvero in alcuni principi che attengono alla responsabilità del costruttore o produttore (art. 114 cod. consumo). Inoltre, esso si pone in sintonia con il fondamentale principio di precauzione di derivazione europea, del quale è stata fatta ampia applicazione in molti casi legati ai temi delle biotecnologie, delle valutazioni scientifiche, di alcune procedure sanitarie e, in generale, in tutti quei casi nei quali natura e tecnologia si intrecciano con le questioni giuridiche afferenti alla salute, all'ambiente e all'alimentazione, generalmente di pertinenza del biodiritto e del diritto pubblico generale.

(iii) Il principio della buona fede ha un ruolo di primaria importanza nella disciplina (non solo) contrattuale nel diritto italiano, rinvenendo le sue radici profonde nel diritto romano ed assumendo la natura di principio generale, applicabile sia ai rapporti pubblicistici che privatistici. I soggetti contraenti, in applicazione di tale principio, devono valutare accuratamente le condotte da mantenere prima, durante e successivamente alla finalizzazione del

¹⁴ C. Viafora, *Fondamenti di bioetica*, CEA, 1989.

sinallagma così come dovranno prestare la medesima attenzione all'attività di interpretazione della volontà negoziale. L'espressione "buona fede", intesa quale regola di comportamento dei contraenti, ricorre non a caso in molte disposizioni del codice civile (Artt. 1175 c.c., 1337 c.c., 1338 c.c., 1366 c.c. e 1375 c.c.). In Italia il principio di buona fede ha acquisito nel corso dei secoli una forza propulsiva originata dalla spinta dottrinale che ha caratterizzato l'evoluzione dell'ordinamento giuridico italiano dall'Impero Romano al Medio Evo – per il tramite dei giureconsulti prima e dei glossatori poi – passando per il Rinascimento e la *Communis Opinio*, sino all'unificazione dello Stato italiano ed al consolidamento dell'attuale sistema normativo fondato su leggi scritte sottoposte alla Costituzione repubblicana, integrate dal formante giurisprudenziale¹⁵. Tale forza si è tradotta in una regola di comportamento autonoma con la quale si può interpretare o addirittura integrare l'accordo contrattuale in modo tale da richiedere alle parti che interagiscono un comportamento improntato alla buona fede nei rapporti contrattuali.

(iv) Il principio di giustificabilità unisce i principi sopra menzionati nella prospettiva di inserire l'obbligazione etica di giustificare, o meglio di rendere conto, dell'attività di intelligenza artificiale correlando la relativa applicazione e circostanziando la relativa messa in esercizio.

Come può facilmente evincersi dalla rassegna di principi proposta ci si muove in un ambito indefinito, nel quale l'ancoraggio a principi generali supplisce all'assenza di una disciplina efficace, rispetto ad un fenomeno del quale non si conoscono appieno le reali implicazioni.

Ciò non deve stupire. Del resto, se oggetto della norma è una fattispecie, la conoscenza della fattispecie rappresenta l'antecedente logico sul quale plasmare una disciplina ritenuta accettabile o desiderabile per la comunità destinataria di quelle norme.

Il progresso tecnologico e l'incedere delle applicazioni concrete dei sistemi di IA, in verità, conservano un tasso di opacità ancora troppo marcato e sovvertono le tradizionali regole di produzione del diritto, sicché l'unico modo di prevenire applicazioni distorte, se non aberranti, di tali tecnologie risiede senz'altro nell'ancoraggio ai principi generali della convivenza civile, specialmente allorquando tali principi siano accolti in testi costituzionali o

¹⁵ Tale orientamento è stato avallato anche dalla giurisprudenza della Corte di Cassazione, la quale in più pronunce ha evidenziato come il principio di buona fede debba permeare l'intero ambito delle relazioni contrattuali integrando la volontà negoziale delle parti coinvolte.

comunque assumano valenza generale ed interdisciplinare, tanto in quanto diffusamente accolti nella società chiamata ad applicarli.

3. I principi sono necessari ma non sufficienti. La necessità di una consapevolezza collettiva.

Il legame tra i principi generali e il tessuto sociale che li dovrà applicare, tuttavia, non si risolve nella mera previsione degli stessi all'interno di codici di comportamento, di codici deontologici o di linee guida comunque denominate.

Infatti il rango e la natura di simile fonte giuridica, avente natura di *soft law*, non consentono di ritenere sufficiente la previsione di tali principi ai fini di scongiurare derive non desiderabili nell'utilizzo dell'IA: tali principi, a ben vedere, da un lato sarebbero comunque applicabili, ancorquando non espressamente richiamati, attesa la natura di principi generali, talora anche di carattere costituzionale; dall'altro lato, per converso, le fonti di *soft law* non hanno comunque carattere di piena cogenza e, dunque, non consentono di incidere nella realtà dei rapporti interpersonali o interorganizzativi come pure dovrebbe accadere per questioni di tale portata.

Accanto ai richiamati presupposti etico-giuridici, dunque, nella formulazione di nuove linee guida in materia di Intelligenza Artificiale, occorre formulare due considerazioni, il cui necessario sviluppo dovrà orientare la formulazione sopra menzionata:

- a) Al fine di rendere "operativi" tali principi occorre correlare tali enunciati con parametri atti a monitorare l'adempimento parziale, totale o nullo del principio. Questo aspetto si pone in linea di continuità, evidentemente, con il tema dei controlli e delle autorità a ciò deputate: occorre dunque che ne derivi una reclinazione anche in termini giuridici in grado di definire procedimenti e competenze oggettive.
- b) Al fine di rendere "operativo" e "funzionale" l'intero processo di *mise en terre* di un'IA meritevole di fiducia occorre riprogettare porzioni di percorsi formativi e allocare risorse finanziarie e umane alla realizzazione di formazione dedicata. La logica di inserire una tale innovazione senza seguire il processo di assorbimento, di collaudo, di aggiustamento può condurre a errori macroscopici per l'intera società italiana ed europea (sono questi gli ambiti di riferimento in oggetto al presente *instant paper*), così come la logica sanzionatoria non potrà condurre a un effettivo

giovamento se non seguirà un graduale percorso di avvicinamento e di formazione a tale innovazione.

Infatti, uno dei rischi più rilevanti che sembrano ancora troppo poco considerati dagli operatori è quello della mancanza di consapevolezza nella società che si accinge ad accogliere ed utilizzare le applicazioni dell'IA.

Da un lato, giova sottolineare che alcune prospettazioni giuridiche conducono al riconoscimento di un primo livello di soggettività a delle macchine¹⁶, quali sono i robot – anche quando espressione di IA – soprattutto allorché riconducono il tema della responsabilità ai loro comportamenti prescindendo dalla considerazione che tali macchine costituiscono elaborazioni e prodotti di esseri umani (i produttori e gli sviluppatori), elemento che dovrebbe far riflettere sulla deriva che potrebbe assumere la tematica.

Dall'altro, gli stessi consociati spesso si espongono volontariamente – il che, a dire il vero, non significa consapevolmente – ai rischi connessi alle applicazioni ICT. È il caso di quello che è stato definito “consenso consapevole disinformato”¹⁷, vale a dire il consenso reso dai consumatori all'utilizzo di prodotti e applicazioni tecnologiche senza prestare la dovuta attenzione alle condizioni contrattuali di utilizzo, con tutte le inferenze che ne derivano in termini di tutela della *privacy*.

Il tema della consapevolezza, per vero, incide trasversalmente sia sulle dinamiche personali che su quelle dei pubblici poteri. Se a livello individuale sono molti i rischi di diseguaglianze e discriminazioni, nonché di mancato rispetto di diritti fondamentali sostanzialmente legati alle applicazioni di IA, in ordine ai pubblici poteri si pone il problema del procedimento che conduce all'adozione delle decisioni pubbliche, il quale sfuma fatalmente nelle dinamiche della discrezionalità amministrativa. La logica procedimentale connessa all'utilizzo degli algoritmi rappresenta un tema cruciale anche rispetto alla trattazione delle questioni tecnico-scientifiche.

Potenzialmente, l'utilizzo di algoritmi rispetto alle decisioni pubbliche assume connotazioni ed inferenze ancor più ampie, se non altro perché le decisioni amministrative si rivolgono sovente ad una pluralità di soggetti.

Proprio per queste ragioni, un alto livello di consapevolezza può e deve essere perseguito non solo con riferimento alla creazione di una coscienza diffusa dell'IA nel tessuto dei consociati ma, anche e soprattutto, all'interno

¹⁶ A. Santosuosso, *Diritto, scienza, nuove tecnologie*, Padova, 2011, spec. p. 249.

¹⁷ C. Casonato, *Intelligenza artificiale e diritto costituzionale: prime considerazioni*, in *Diritto pubblico comparato ed europeo*, fasc. speciale, maggio 2019, pp. 101 ss., spec. p. 107.

delle istituzioni. Questo significa che sia i decisori politici che il personale tecnico e funzionariale debbono essere accompagnati sin da ora nel percorso verso l'assimilazione dei rischi e delle potenzialità effettivamente legate alle applicazioni di IA, se non altro affinché con il loro operato non fungano essi stessi da cassa di risonanza di quei rischi anche a livello ordinamentale.

Le iniziative formative, divulgative, scientifiche e culturali che verranno in futuro dedicate alle implicazioni dell'IA, dunque, mitigheranno sia i rischi legati all'attuale assenza di un quadro normativo chiaro e definito sia quelli legati alla potenziale predisposizione di un sistema di norme inadeguato rispetto all'oggetto che è chiamato a disciplinare.