

I codici di condotta di cui alla Proposta di Regolamento UE sull'Intelligenza Artificiale

ipotesi per un'applicazione dello Human Rights-Based Approach

Piergiorgio Belotti*

Abstract Il contributo, dopo aver svolto una ricostruzione storico-giuridica (per ovvie ragioni parziale) dell'evoluzione del c.d. Ai Act, si sofferma sui codici di condotta in questa contenuti, anche tramite un confronto con altri codici di condotta presenti in altri settori. Ciò allo scopo di provare ad ipotizzare un utilizzo di questo istituto per (ri)portare al centro del discorso lo Human Rights Based Approach.

1 Introduzione

Le norme che verranno emanate dall'Unione Europea in materia di Intelligenza Artificiale¹ avranno come obiettivo dichiarato quello di disciplinare le conseguenze² degli utilizzi da parte delle persone e delle organizzazioni di questa (relativamente) nuova famiglia di tecnologie.

Queste saranno sempre più incisive nel corso del ventunesimo secolo grazie alla enorme quantità di dati disponibili, imparagonabile sino ad oggi, ad un costo relativamente basso³. Infatti, come noto, l'Intelligenza Artificiale utilizza dati, in modo diretto nel caso del c.d. Machine Learning e in modo indiretto nei sistemi basati su regole, per ottenere output predittivi. All'aumentare della disponibilità di dati, aumenta anche la quantità e, in alcuni casi, anche la qualità, delle previsioni ad essi associate; le quali potranno essere utilizzate da un numero sempre maggiore di soggetti, di qualunque tipologia, per una più efficace programmazione delle proprie attività⁴ e, dunque, in ultima analisi, per una maggiore pervasività dell'agire. Senza contare, peraltro, che molti sistemi di Intelligenza Artificiale non avranno – e non hanno – nemmeno bisogno di operatori umani per poter svolgere le proprie azioni e, anzi, operano in sostituzione delle persone nello svolgimento di diversi compiti e spesso anche nel prendere numerose decisioni⁵.

L'Unione Europea, nell'introdurre i documenti preparatori alla Proposta di Regolamento, individua sempre due tipologie di conseguenze dalle quali parte per sviluppare il proprio discorso.

*Studio Legale Vincenzi, Brescia; ✉ piergiorgio.belotti@gmail.com

1. Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts
2. VON DER LEYEN U., *Un'Unione più ambiziosa. Il mio programma per l'Europa*, 16.7.2019
3. BERNABE' F., TREMONTI G. e IANNINI A., *La nuova economia fondata sui dati*, in "Big Data e Privacy. La nuova geografia dei poteri", Garante per la protezione dei dati personali, 2017: "Nel 2015 l'umanità ha prodotto più informazioni di quante ne fossero state create in tutti gli anni precedenti della civiltà umana"
4. Si veda un esempio risalente di efficace utilizzo dei dati effettuato dalla compagnia Netflix in CARR D., *Giving Viewers What They Want*, The New York Times, 24.2.2013
5. Le parole dell'AI - Etica e Intelligenza Artificiale nelle parole di Pasquale Stanzone <https://youtu.be/IfQQZsWWME4>

In primo luogo, cita sempre le conseguenze a favore di un migliore sviluppo della persona e della società umana, dovuto proprio alla natura di tali famiglie di tecnologie. Tra queste se ne sono rinvenute sin dall'inizio numerose: *“dal trattamento delle malattie croniche o dalla riduzione dei tassi di incidenti stradali mortali alla lotta contro il cambiamento climatico o alla prevenzione delle minacce alla sicurezza informatica”*⁶. E a tal fine ha sempre mostrato, nelle scelte di politica legislativa, una particolare attenzione agli investimenti utili allo sviluppo dell'Intelligenza Artificiale e alla creazione di un quadro di riferimento normativo che possa offrire a questi investitori la certezza del diritto, con lo scopo di raggiungere, in ultima istanza, la c.d. sovranità tecnologica⁷.

In secondo luogo, viceversa, evidenzia anche i pericoli e i rischi che possono derivare da un utilizzo malevolo o comunque disaffetto dell'Intelligenza Artificiale⁸. Esemplicando: i problemi legati alla discriminazione di minoranze, alla manipolazione di persone, alla selezione di persone effettivamente utili alla società (c.d. social scoring), alla protezione dei dati personali e molte altre ancora.

Si potrebbe affermare che l'approccio iniziale dell'Unione Europea al tema sia efficacemente riassumibile nella formula *“L'essere umano come il fine, non come il mezzo”*⁹. In conclusione, un'Intelligenza Artificiale al servizio dell'essere umano e per lo sviluppo della società umana. Si vedrà più avanti come l'Unione Europea ha concretamente inteso applicare questa idea fondamentale, sostanzialmente – si ritiene – disapplicando quanto appena descritto.

2 I Brevi notazioni storiche relative alla Proposta di Regolamento UE sull'intelligenza artificiale

2.1 Primi atti introduttivi

Il primo atto di indirizzo politico proveniente dall'Unione Europea nell'ambito dell'Intelligenza Artificiale avvenne durante il Consiglio Europeo di Tallin del 29 settembre 2017, al termine del quale venne redatta una relazione contenente anche le Conclusioni¹⁰. In particolare, nell'ottica di un'Europa digitale che punti ad ottenere maggiore forza e maggiore coerenza e, nello specifico, nella transizione dell'Europa verso il digitale, sono stati individuati alcuni moduli di intervento delle politiche europee. Tra questi si citano, in particolare, l'adeguamento delle Pubbliche amministrazioni e dei settori pubblici all'era digitale, la necessità di un quadro normativo per il futuro – viene a tal proposito in luce un accenno al Digital Market Act – e molti altri profili. Ciò fino a giungere al punto intitolato *“la consapevolezza dell'urgenza di far fronte alle tendenze emergenti”*, riguardante anche l'Intelligenza Artificiale, che è di fondamentale importanza per due motivi. Anzitutto perché in virtù di questo l'Unione Europea dovrebbe impegnarsi affinché sia garantito, nello sviluppo di tali famiglie tecnologiche, un *“elevato livello di protezione dei dati, diritti digitali e norme etiche”*¹¹. In secondo luogo, tale Consiglio Europeo invita la Commissione a presentare un proprio approccio all'IA, nonché a rafforzare le condizioni quadro allo scopo di *“esplorare nuovi mercati tramite innovazioni radicali basate sul rischio e di riaffermare il suo ruolo guida nel settore industriale”*.

La Commissione rispose a questa chiamata redigendo un documento¹² intitolato *“L'intelligenza artificiale per l'Europa”* del 25.4.2018 nel quale affermava la propria iniziativa secondo una triplice linea di intervento, con lo scopo ultimo di ottenere un'Intelligenza Artificiale sulla quale le persone possano avere fiducia¹³. In primo luogo, riteneva fosse opportuno sostenere gli investimenti, sia privati sia pubblici, nella ricerca e nello sviluppo dell'Intelligenza Artificiale e nell'accesso ai dati. In secondo luogo, prendendo sostanzialmente atto che gli effetti di questa famiglia di tecnologie sarebbero stati dirompenti, poneva

6. COM(2018) 237 final “L'intelligenza artificiale per l'Europa” pag. 1

7. VON DER LEYEN U., Un'Unione più ambiziosa. Il mio programma per l'Europa, 16.7.2019

8. COM(2018) 795 final, “Piano coordinato sull'intelligenza artificiale”

9. Parafrasi di un'espressione tratta da KANT I., *Fondazione della metafisica dei costumi*, 1785.

10. CO EUR 17 (CONCL 5)

11. Si noti sin da subito che non viene evidenziato il c.d. Risk-based Method come primo profilo nella prima proposizione del Consiglio Europeo, ma solamente nella seconda proposizione.

12. COM(2018) 237 final, pag. 3

13. Ibidem, pag. 16

l'attenzione sulla necessità di migliorare l'istruzione¹⁴ e la formazione delle persone per impedire che qualcuno possa "rimanere indietro". Il terzo pilastro dell'iniziativa puntava a garantire l'esistenza di un quadro etico e giuridico adeguato, nel quale vi sia cioè garanzia dei diritti fondamentali raccolti nella Carta dei diritti fondamentali dell'Unione Europea, nonché rispetto dei valori fondativi dell'Unione di cui all'art. 2 del Trattato sull'Unione Europea.

Con il "Piano Coordinato sull'intelligenza artificiale"¹⁵, la Commissione ha messo in pratica le linee guida tracciate mediante la Comunicazione appena descritta. In particolare, sono state individuate alcune concrete azioni tra cui, quella che rileva maggiormente ai nostri fini è quella di cui al par. 2.6, che individua alcune caratteristiche che l'Intelligenza Artificiale dovrebbe possedere: prevedibilità, responsabilità, verificabilità, dovrebbe rispettare i diritti fondamentali e seguire regole etiche. Il piano individua poi due specifiche azioni: la prima è la redazione, da parte del gruppo di esperti di alto livello denominato "European AI Alliance"¹⁶ e di una successiva consultazione pubblica, di un documento contenente orientamenti etici. La seconda è una valutazione della normativa esistente al momento del Piano per verificare se sia flessibile, nonché sufficiente a garantire la sicurezza e la protezione.

2.2 Orientamenti etici per un'IA affidabile

Il successivo passo, quello relativo all'esame dei profili etici dell'IA, si è concretizzato l'8 aprile 2019, quando venne pubblicato il documento definitivo del Gruppo di esperti¹⁷ citato precedentemente. Sin dalle prime pagine di questo viene posto in evidenza che l'obiettivo finale è quello di ottenere una tecnologia affidabile, che consiste in un vero e proprio "prerequisito per lo sviluppo, la distribuzione e l'utilizzo di sistemi di IA". L'affidabilità viene elevata a vero criterio distintivo per una tecnologia di questo tipo, per una semplice ragione: se l'IA e le persone che la sviluppano non fossero affidabili, potrebbero generare danni incommensurabili agli esseri umani e alla società nella quale vivono. Non solo: viene sostenuto che se l'affidabilità non fosse percepita dagli utenti o comunque dall'opinione pubblica, questo potrebbe impedire anche lo sviluppo dell'Intelligenza Artificiale, poiché è chiaro che se una tecnologia non fosse ritenuta sicura, allora anche la domanda nel contesto del mercato non sussisterebbe e dunque non vi sarebbero i necessari investimenti per mancanza di prospettive di guadagno. L'affidabilità non è dunque solamente uno strumento di tutela delle persone ma diventa, nell'ottica degli esperti, anche un fattore di protezione del capitale e della proprietà, nonché di un mercato competitivo¹⁸.

Affinchè un'IA sia affidabile, il documento individua tre profili necessari (che devono essere coesistenti, poiché in grado di rafforzarsi vicendevolmente), che devono però essere interpretati in senso olistico¹⁹:

1. legalità, ossia il rispetto della normativa, primaria e secondaria, europea e nazionale (profilo del quale gli Orientamenti Etici non si occupano);
2. eticità, l'IA deve aderire a principi e valori etici;
3. robustezza tecnica e sociale.

È interessante, ai fini del presente contributo, notare come il profilo dell'eticità prenda spunto da un'affermazione del Gruppo, che mette in guardia il lettore dal ritenere sufficiente che una IA debba solo osservare la legge: infatti, non solo la legge spesso non riesce a reggere il passo dell'evoluzione tecnologica²⁰ o con l'etica più in generale²¹ ma spesso non è nemmeno adatta. Pertanto, per superare i limiti appena descritti, si è ritenuto opportuno affermare che l'Intelligenza Artificiale deve anche essere

14. Si veda ad esempio il "Piano d'azione per l'istruzione digitale" presentato dalla Commissione nel 2018

15. COM(2018) 795 final

16. <https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance>

17. GRUPPO DI ESPERTI AD ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, *Orientamenti etici per un'IA affidabile*, Commissione europea, 8.4.2019

18. IASELLI M., *AI, il paradosso della legge come ostacolo o supporto: perché serve un nuovo approccio*, AgendaDigitale.eu, 4.3.2021

19. Ibidem, "L'impegno a favore di un'IA affidabile non riguarda solo l'affidabilità del sistema di IA stesso, ma richiede un approccio olistico e sistemico, che comporti l'affidabilità di tutti gli attori e di tutti i processi appartenenti al contesto sociotecnico del sistema durante l'intero ciclo di vita"

20. IASELLI M., *AI, il paradosso della legge come ostacolo o supporto: perché serve un nuovo approccio*, AgendaDigitale.eu, 4.3.2021

21. CARBONE L., *Etica morale e deontologia nell'esercizio delle libere professioni, La previdenza forense*, 3/2020

compatibile con valori e principi etici²². Si precisa, tuttavia, che la disciplina di riferimento per quanto concerne questo aspetto è quella specifica dell'Etica dell'Intelligenza Artificiale, ossia *“una branca dell'etica applicata che studia gli interrogativi etici posti dallo sviluppo, dalla distribuzione e dall'utilizzo dell'IA. Il suo interesse principale risiede nell'individuare come l'IA possa favorire o mettere a rischio la felicità degli individui, sia in termini di qualità della vita che di autonomia umana e libertà necessarie per una società democratica”*. In definitiva un'Intelligenza Artificiale antropocentrica²³.

L'approccio al tema del Gruppo di Esperti è quello di un'etica basata sui diritti fondamentali sanciti dai Trattati dell'unione Europea e dal diritto internazionale in materia di diritti umani. Dunque, un'etica che parzialmente si sovrappone con il profilo della legalità, ma che non si esaurisce meramente in questo, poiché affronta un'ulteriore questione di grande importanza: ciò che implicano i diritti fondamentali e i valori ad essi riferiti. Concretamente questo tipo di analisi viene svolto dal Gruppo mediante la ricerca di connessioni fra i diritti fondamentali.

L'epilogo dell'esame²⁴ condotto è la “scoperta” di quattro imperativi etici sottostanti ai diritti (giuridici) fondamentali:

1. il rispetto dell'autonomia dell'essere umano;
2. un'impostazione tesa alla prevenzione di danni;
3. l'equità;
4. l'esplicabilità.

Quanto al primo di questi, deve essere inteso come mantenimento della capacità dell'essere umano di effettiva autodeterminazione²⁵. Tale parola, in questo contesto, deve significare, contemporaneamente, sia assenza di costrizione o inganno o manipolazione sia miglioramento di capacità cognitive della singola persona. In tal guisa, l'Intelligenza Artificiale deve essere soggetta al principio della progettazione Antropocentrica che si traduce nell'attribuzione all'essere umano del potere di sorveglianza e il controllo dei processi operativi.

Il secondo attiene alla sicurezza nell'utilizzo dell'intelligenza artificiale, che deve essere garantita sia quanto a prevenzione di veri e propri danni alla persona o agli ambienti, sia a impedimento di manipolazioni. Elemento centrale in questo senso potrebbe essere il concetto di “security by design²⁶”, ossia la necessità di progettare sistemi informatici contemporaneamente funzionanti e sicuri nell'utilizzo, nella consapevolezza che la sicurezza è un processo, non un prodotto²⁷. L'imperativo in parola include anche un altro genere di sicurezza, ossia quella determinata dal rispetto dell'ambiente, dunque dalla sostenibilità nel design della tecnologia²⁸ e nell'utilizzo della tecnologia.

Il terzo, l'equità, attiene invece allo sviluppo, alla distribuzione e all'utilizzo dei sistemi di IA. Essa può essere a sua volta distinta in due modi: sostanziale²⁹ e procedurale³⁰.

22. Per una definizione di valori e principi potrebbe essere utile e maneggevole leggere il seguente contributo: ZAGRELBESKY G., *Valori e diritti nei conflitti della politica*, la Repubblica, 22.2.2008

23. PEYRON L., *Nell'IA “antropocentrica”, l'umano dov'è? Qualche risposta*, AgendaDigitale.eu, 11.2.2022

24. Per una diversa visione della questione si rinvia a FLORIDI L., *Etica dell'Intelligenza Artificiale*, Raffaello Cortina Editore, 2022

25. Si veda ad esempio quanto ampiamente trattato in: DE MENECH C., *Intelligenza artificiale e autodeterminazione in materia sanitaria*, BioLaw Journal – Rivista di BioDiritto, n. 1/2022

26. COSTA G., *Security-by-design: strumenti e metodologie per lo sviluppo sicuro del software*, Cybersecurity360, 12.10.2018

27. VALENTINI A., *Fuggetta: “la cyber security non è un prodotto da acquistare”*, Cybersecurity360, 29.6.2022

28. Basti pensare, per un paragone, al dibattito in corso nell'ambito della tecnologia Blockchain, in particolare alla c.d. Proof-of-work: DE CONNO A., *Criptovalute e impatto ambientale: verso una nuova regolamentazione europea. La UE monitora il consumo di energia generato dagli strumenti finanziari e mira alla sostenibilità*, Altalex, 30.08.2022

29. *“La dimensione sostanziale implica un impegno a garantire una distribuzione giusta ed equa di costi e di benefici e a garantire che gli individui e i gruppi siano liberi da distorsioni inique, discriminazioni e stigmatizzazioni. Riuscendo a evitare distorsioni inique, i sistemi di IA potrebbero persino aumentare l'equità sociale. Occorre inoltre promuovere le pari opportunità in termini di accesso all'istruzione, ai beni, ai servizi e alla tecnologia.[...] Inoltre, l'equità implica che gli operatori del settore dell'IA rispettino il principio di proporzionalità tra mezzi e fini, e valutino attentamente come bilanciare interessi e obiettivi concorrenti”*.

30. *“capacità di impugnare le decisioni elaborate dai sistemi di IA e dagli esseri umani che li gestiscono e la possibilità di presentare un ricorso efficace contro di esse. A tal fine, l'organismo responsabile della decisione deve essere identificabile e i processi decisionali devono essere spiegabili”*

Il quarto imperativo, l'esplicabilità, è certamente connesso alla dimensione procedurale dell'equità, ma deve essere inteso in senso più ampio come trasparenza³¹ dei metodi e delle logiche sottostanti all'Intelligenza Artificiale, le capacità e gli scopi di ciascun sistema di IA devono essere comunicati apertamente e le decisioni, per quanto possibile, devono poter essere spiegate agli interessati, anche ai fini di un'eventuale impugnazione.

Il documento prende atto che talvolta questi imperativi possono tra loro scontrarsi e per tale ragione il Gruppo esorta l'Unione Europea a trovare una metodologia democratica ed inclusiva per garantire la creazione di compromessi accettabili. Non sempre, tuttavia, ciò sarà possibile, ossia nel caso in cui sussistano dei diritti umani non compromettibili. In tali casi dovranno prevalere le esigenze di tutela di tali diritti fondamentali.

Gli Orientamenti, preso atto della sussistenza di tali imperativi e del loro rapporto, hanno provato a tradurli in 7 requisiti per un'Intelligenza Artificiale affidabile:

1. Intervento e sorveglianza umani
2. Robustezza tecnica e sicurezza
3. Riservatezza e governance dei dati
4. Trasparenza
5. Diversità, non discriminazione ed equità
6. Benessere sociale e ambientale
7. Accountability

Infine, dopo aver descritto approfonditamente le caratteristiche di ciascuno di questi requisiti, si è provato a creare – con la partecipazione di numerosissimi stakeholder del settore – una Check List utile agli sviluppatori per verificare la sussistenza di tali requisiti all'interno di una vera Intelligenza Artificiale.

Si vuole precisare che in nessuno di questi punti si è mai imposto in alcun modo l'utilizzo del "risk based method" e, anzi, come si è potuto notare, il Gruppo ha scelto di mettere in primo piano il principio di autodeterminazione dell'essere umano utilizzatore dell'Intelligenza Artificiale.

2.3 Libro bianco sull'intelligenza artificiale

Con il "White Paper on Artificial Intelligence: a European approach to excellence and trust"³² cambia notevolmente la prospettiva della Commissione, adeguandosi all'indirizzo politico tracciato dal Consiglio europeo del settembre 2017. In tale documento, infatti, viene presa una forte presa di posizione, anche sulla spinta di atti di soft law emanati da alcuni stati nazionali³³, per un approccio europeo basato sul rischio.

Il documento della commissione è incentrato su tre profili, tre esigenze: la tutela dei diritti fondamentali (tra cui protezione dei dati personali o la non discriminazione), la sicurezza e la responsabilità.

Quanto al primo di questi, vengono evidenziate due tipologie di cause di pericolo per i diritti fondamentali: un'errata progettazione³⁴ del software alla base dell'Intelligenza Artificiale e un'errata raccolta e selezione dei dati³⁵ sulla base dei quali si è svolto l'addestramento. Entrambe porterebbero a decisioni dell'IA discriminatorie e dunque lesive.

31. Sotto il profilo della trasparenza, nell'ordinamento giuridico italiano, negli ultimi anni sta nascendo un interessante orientamento giurisprudenziale sulla c.d. trasparenza algoritmica: PURIFICATI J., *La Corte di Cassazione sulla trasparenza degli algoritmi e la validità del consenso*, e-lex, 7.6.2021

32. COM(2020) 65 final

33. Tra gli altri, si veda quanto scritto in DER BUNDESBEAUFTRAGTE FÜR DEN DATENSCHUTZ UND DIE INFORMATIONSFREIHEIT (BFDI), *Tätigkeitsbericht zum Datenschutz 2019*

34. FALCHI M. C., *Intelligenza Artificiale: se l'algoritmo è discriminatorio*, Ius in itinere, 2020-2022

35. MARTORANA M. SAVELLA R., *Algoritmi razzisti, ecco le cause del problema e le soluzioni in campo*, AgendaDigitale.eu, 9.7.2020

Il tutto risulterebbe aggravato dall'opacità dei meccanismi decisionali, fattore che renderebbe sostanzialmente impossibile la verifica del rispetto della normativa (europea ma non solamente) applicabile al caso concreto, con il rischio di giungere a decisioni lesive dei diritti fondamentali dell'essere umano.

Vi è poi da considerare un fenomeno ad esso connesso, ossia la c.d. fuga dalla percezione diretta³⁶, ossia dal fatto che le decisioni prese non appaiano all'utente come l'esito di un percorso motivazionale basato sul confronto e sulla valutazione di diversi argomenti, quanto piuttosto un mero dato di fatto, quasi come se giungesse all'inconscio dell'utente e questi non potesse che aderire in maniera automatica.

In ultima istanza, tutte queste caratteristiche ben potrebbero impedire a chi subisca una decisione illegittima, discriminatoria o semplicemente ingiusta, un'effettiva tutela giurisdizionale, (tema peraltro in linea con una più ampia trattazione attinente alla trasparenza delle decisioni delle pubbliche amministrazioni europee, da molto tempo presente nella dottrina e nella giurisprudenza³⁷). Un contributo recente ha sintetizzato molto efficacemente la questione anche in rapporto alla Intelligenza Artificiale: *“Per contestare efficacemente le decisioni basate sull'utilizzo dell'IA le persone devono essere a conoscenza dell'utilizzo dell'IA, e delle modalità e delle sedi in cui presentare un reclamo. Le organizzazioni che si avvalgono dell'IA devono saper spiegare il loro sistema e le decisioni basate sull'IA”*³⁸.

Un ulteriore profilo interessante del White paper attiene alla ratio dietro la scelta del metodo basato sul rischio, cioè garantire una proporzionalità³⁹ dell'intervento legislativo europeo, per poter dare certezza giuridica e nel contempo tutela adeguata alle persone. Inoltre, per non porre un peso eccessivo sulle spalle degli attori del mercato, in particolare le piccole e medie imprese, vera ossatura dell'apparato produttivo europeo, si è deciso di ridurre al minimo gli oneri documentali, di analisi e design del prodotto Intelligenza Artificiale. Si è pertanto coniata la formula “Intelligenza Artificiale ad Alto rischio⁴⁰”, alla quale si dovrebbero applicare tali adempimenti assai impegnativi: ciò sulla base sia del settore di utilizzo, sia sul modo di utilizzo all'interno di tale settore. Le prescrizioni che i produttori di queste tipologie di Intelligenze Artificiali dovrebbero seguire, a loro volta, dovrebbero essere influenzate dalle determinazioni del Gruppo di Esperti ad Alto livello poc'anzi individuate.

La commissione modifica *de facto* la funzione dell'elenco individuato dal Gruppo degli esperti. Si potrebbe affermare come tali determinazioni siano state trasformate in un mezzo di attenuazione del rischio, quando in origine avevano in realtà lo scopo di (in)formare la tecnologia di riferimento. Da requisiti a adempimenti.

2.4 Lineamenti generali della Proposta di Regolamento del 2021

Senza entrare nel merito della discussione relativa alla tipologia di definizione (peraltro ormai superata dall'evoluzione dei testi normativi) utilizzata dal legislatore europeo, che non ha convinto parte della dottrina⁴¹, la Proposta di Regolamento c.d. Ai Act⁴² conferma comunque l'approccio dettato dal Libro bianco con riguardo al metodo basato sul rischio, anche alla luce del consenso ricevuto dalla maggioranza partecipante alla Consultazione Pubblica che si è svolta. Assieme alla proposta vi è anche un Relazione preliminare molto utile alla ricostruzione della ratio delle norme e che verrà utilizzata quale guida alla presente analisi.

Preliminarmente alla redazione della Proposta in questione, sono state valutate alcune opzioni strategiche con diversi gradi a livello normativo:

1. Sistema volontario di etichettatura;

36. BUCCI F., Algoritmi e diritti umani: qual è il punto d'incontro? Ius in itinere, 2021-2022

37. TOMMASI C., Il diritto di accesso nell'ordinamento dell'unione europea: Trasparenza o opacità amministrativa? Seminario annuale con i dottorandi in materie gius-pubblicistiche

Università degli studi roma tre – dipartimento di giurisprudenza 21 settembre 2018

38. AGENZIA DELL'UNIONE EUROPEA PER I DIRITTI FONDAMENTALI, Preparare un giusto futuro l'Intelligenza Artificiale e i diritti fondamentali – Sintesi, 2021

39. Glossario - Politica regionale - Commissione europea (europa.eu)

40. Per un approfondimento sul conetto di Intelligenza Artificiale ad alto rischio si veda DIMALTA D., Intelligenza artificiale “ad alto rischio”, tutti gli ambiti su cui vigilare per la UE, Agentadigitale.eu, 5.5.2021

41. CONTISSA G., GALLI F., GODANO F., SARTOR G., *Il regolamento europeo sull'intelligenza artificiale*, i-lex, 2021, fasc. 2

42. COM/2021/206 final

2. Approccio settoriale specifico;
3. Normativa orizzontale con approccio basato sul rischio;
4. Normativa orizzontale con approccio basato sul rischio (per le IA ad alto rischio) e codici di condotta per IA non ad alto rischio;
5. Normativa orizzontale dell'UE che stabilisce requisiti obbligatori per tutti i sistemi di IA a prescindere dal rischio.

Tra queste possibilità si è scelto di optare per la numero 4, che si è ritenuta essere la migliore sotto il profilo dell'efficienza e sotto quello dell'efficacia dell'intervento legislativo a seguito dell'indagine compiuta in sede di Valutazione di Impatto⁴³.

Sotto il profilo tecnico, la Proposta del 2021 prevede essenzialmente due tipologie di interventi legislativi: la prima di queste ha carattere obbligatorio e la seconda di applicazione volontaria (si fa riferimento ai codici di condotta).

Per quanto concerne la tipologia di applicazione obbligatoria, essa si sviluppa a sua volta in ulteriori rami. Il primo attiene alla disciplina delle Intelligenze Artificiali che comportano un rischio ritenuto inaccettabile perché espressamente contrario ai valori tutelati e promossi dal Diritto dell'Unione Europea, le quali sono espressamente vietate. Tra queste si ritrovano, ad esempio, Intelligenze Artificiali che sono in grado di manipolare le persone e quindi ledere il principio di autodeterminazione (si veda in proposito quanto scritto precedentemente sul punto), oggi talmente rilevante e determinante nel contesto del mercato, in particolare quello pubblicitario⁴⁴, che una dottrina ha ritenuto che fosse il momento di introdurre un nuovo Principio, quello dell'*Habeas Mentem*⁴⁵, che dovrebbe diventare informatore nell'ambito della legislazione delle neurotecnologie. Altra tipologia di Intelligenza Artificiale vietata sarebbe quella relativa al c.d. Reputation scoring⁴⁶, già utilizzato dai sistemi politici di alcuni stati extra-europei, seppur ancora ad uno stato embrionale.

Un ulteriore ramo è quello relativo alle Intelligenze Artificiali ad alto rischio. Tra queste rientrano quelle di cui allegato III. O anche quelle destinate a essere utilizzate come componente di sicurezza di un prodotto, o siano esse stesse un prodotto, disciplinato dalla normativa di armonizzazione dell'Unione elencata nell'allegato II e che siano prodotti, il cui componente di sicurezza è il sistema di IA, o il sistema di IA stesso in quanto prodotto sia soggetto a una valutazione della conformità da parte di terzi ai fini dell'immissione sul mercato o della messa in servizio ai sensi della normativa di armonizzazione dell'Unione elencata nell'allegato II. Ciò a prescindere dal fatto che sia immesso sul mercato o messo in servizio in modo indipendente rispetto ai prodotti citati poc'anzi.

Come prima osservazione, una dottrina⁴⁷ ha avuto modo di evidenziare come *“le aree connesse al rischio sono certamente sensibili: tuttavia, non si individua un metodo chiaro e generale per determinare le specifiche caratteristiche che darebbero luogo al rischio di un sistema”*. Ciò comporterebbe un eccessivo ampliamento della sfera di applicazione delle norme che verranno riassunte nei punti essenziali qui avanti. Molte Intelligenze Artificiali, ancorchè operanti in ambiti teoricamente coperti dalle discipline di cui all'Allegato II, infatti, potrebbero non comportare effettivamente alcun rischio. Una tale difficoltà interpretativa, tuttavia, in virtù della successiva evoluzione della Proposta di Regolamento, potrebbe non più verificarsi.

Per tali tipologie di Intelligenze Artificiali la proposta prevede l'obbligo di porre numerose azioni positive tese⁴⁸ a limitare i rischi. In particolare, i rischi considerati dal Regolamento sono i seguenti:

- 1) Alla salute. Per un esempio di questo genere di rischi si pensi al contesto lavorativo, nel quale potrebbero sussistere problemi nel rapporto⁴⁹ fra lavoratori e robot che potrebbero provocare impatti negativi sulla salute mentale dei lavoratori:

43. SWD(2021)84

44. Da una logica think-choose a una feel-choose. Per un approfondimento si rinvia a PERON E., *Neuromarketing: come il cervello guida il consumatore*, thismarketerslife.it, 22.01.2016

45. RESTA F., *IA e neuroscienze, la nostra “privacy mentale” è a rischio? I paletti che servono*, AgendaDigitale.eu, 17.6.2021

46. DI CARPEGNA BRIVIO E., *Il Reputation scoring e la quantificazione del valore sociale*, Federalismi.it, 13.7.2022

47. CONTISSA G., GALLI F., GODANO F., SARTOR G., *Il regolamento europeo sull'intelligenza artificiale*, i-lex, 2021, fasc. 2

48. Si precisa che l'art. 40 della Proposta di Regolamento stabilisce quanto segue: *“i sistemi di IA ad alto rischio che sono conformi alle norme armonizzate o a parti di esse i cui riferimenti sono stati pubblicati nella Gazzetta ufficiale dell'Unione europea si presumono conformi ai requisiti essenziali, di cui al capo 2 del presente titolo, nella misura in cui tali requisiti sono contemplati da tali norme”*.

49. EUROPEAN AGENCY FOR SAFETY AND HEALTH AT WORK, *Impact of artificial intelligence on occupational safety and health*, 2021

“Workers having to keep up with a cobot’s pace and level of work may be put under pressure to achieve the same level of productivity. Increased working with robots may also significantly reduce contact with human peers and social support. This may have negative impacts on workers’ safety and health, particularly their mental health”.

- 2) Alla sicurezza, come ad esempio nel caso di utilizzo di queste tecnologie per attacchi informatici⁵⁰.
- 3) Ai diritti fondamentali. Questi, in linea di continuità con quanto scritto all’interno del White Paper, costituiscono un lungo elenco non esauribile nel presente contributo, peraltro strettamente connesso con i primi due punti di questo elenco. Tanto che si potrebbe affermare come il terzo punto di fatto inglobi anche i primi due. Le tecnologie quali quelle in esame ben potrebbero causare notevoli lesioni ai diritti fondamentali degli esseri umani, come si è potuto osservare nel capitolo “Introduzione” del presente contributo.

La Proposta di Regolamento, conferma tale definizione di rischio all’art. 65, che fa espresso richiamo dell’art. 3 punto 19 del Regolamento UE 2019/1020, che definisce il concetto di *“prodotto che presenta un rischio”* come un prodotto che in ipotesi potrebbe provocare un pregiudizio alla salute e alla sicurezza delle persone, anche sul luogo di lavoro, la protezione dei consumatori, l’ambiente e la pubblica sicurezza, nonché altri interessi pubblici tutelati dalla normativa di armonizzazione dell’Unione applicabile.

Ebbene, la prima applicazione sostanziale, forse più immediata, di tale metodo basato sul rischio è costituita innanzitutto dal Sistema di gestione dei rischi di cui all’art. 9 della proposta, ossia un processo continuo eseguito nel corso dell’intero ciclo di vita dell’Intelligenza Artificiale ad alto rischio che si suddivide in plurime fasi: identificazione e analisi dei rischi noti e prevedibili, stima e valutazione dei rischi (da utilizzi conformi o impropri ragionevolmente prevedibili⁵¹), valutazione di rischi raccolti da un monitoraggio successivo dell’operatività del sistema e la creazione di misure di gestione dei suddetti rischi, le quali devono tenere conto degli effetti delle interazioni con gli altri istituti di cui al capo 2 del regolamento, nonché dello stato della tecnologia presente in un determinato momento. Le misure di gestione dei rischi sono poi individuate mediante la sottoposizione dell’Intelligenza Artificiale ad una prova di funzionamento preliminare alla immissione sul mercato o della messa in servizio del sistema tecnologico, nell’ottica (limitata) di conseguire la finalità prevista.

La proposta regola poi molti altri aspetti, tra i quali: i criteri di qualità che devono essere rispettati da coloro che realizzano set di dati di addestramento, convalida e prova, tra le quali rientrano corrette pratiche di governance dei dati e caratteristiche intrinseche dei dati che rispondano a criteri predeterminati.

Il principio dell’accountability informa poi la normativa europea, poiché viene imposta la conservazione della Documentazione tecnica dell’IA, in modo che sia possibile verificarne la rispondenza alla Proposta di regolamento.

Profilo tecnico e giuridico essenziale è quello disciplinato all’art. 12, il dovere di conservazione dei c.d. Log, cioè la registrazione automatica di tutti gli eventi che si verificano nel funzionamento dell’Intelligenza Artificiale e in particolari di alcuni dati e metadati da essa generati.

Altri profili, attinenti al concetto già descritto di human-in-command, sono quelli della trasparenza del funzionamento dell’Intelligenza Artificiale, in modo che permetta agli utenti una facile comprensione, interpretazione ed utilizzo dell’output del sistema e più in generale la progettazione del sistema in modo tale da consentire la sottoposizione dello stesso alla sorveglianza umana.

In ultima istanza, vengono posti dei principi che hanno lo scopo di raggiungere una migliore accuratezza, robustezza e cibersicurezza di ciascuna Intelligenza Artificiale ad alto rischio.

Il successivo capo della Proposta di regolamento, invece, ha lo scopo di applicare concretamente questi requisiti appena individuati all’intera filiera di progettazione, produzione, distribuzione, importazione ed utilizzo dei prodotti in questione. Ciò viene fatto attribuendo a ciascun soggetto partecipante di questa catena di valore numerosi obblighi e correlate responsabilità, collegate ai requisiti che devono essere rispettati dalle IA ad alto rischio. Ad esempio, si prevedono obblighi di redigere il sistema di gestione della qualità, conservazione di log, redazione della documentazione tecnica pertinente e così via in capo al fornitore. Di notevole interesse è l’art. 28, che stabilisce quello che si potrebbe chiamare “principio di conversione”, in virtù del quale

50. HEINEMEYER M. *L’intelligenza artificiale negli attacchi informatici: tecniche offensive e strategie di difesa*, Cybersecurity360, 1.06.2020

51. Prospettiva confermata dal rimando di cui all’art. 65 della Proposta al Regolamento UE 2019/1020, in virtù del quale il livello di rischio considerato arriva *“oltre quanto ritenuto ragionevole ed accettabile in relazione all’uso previsto del prodotto o nelle condizioni d’uso normali o ragionevolmente prevedibili, incluse la durata di utilizzo e, se del caso, i requisiti relativi alla messa in servizio, all’installazione e alla manutenzione”*.

un qualsiasi distributore, importatore, utente o altro terzo, alle condizioni ivi descritte (es. apposizione del proprio marchio), verrebbe assimilato ad un fornitore quanto a doveri.

I capi successivi della Proposta, invece, disciplinano la governance dei profili tecnici e dei requisiti sopracitati, stabilendo un doppio livello di funzionamento: il primo livello è quello delle Autorità di Notifica, che hanno lo scopo di predisporre e di eseguire le procedure necessarie per la valutazione, la designazione e la notifica degli organismi di valutazione della conformità e per il loro monitoraggio. Il secondo livello è quello relativo agli Organismi Notificati, che hanno il compito di verificare la conformità del sistema di Intelligenza Artificiale ad alto rischio secondo le procedure di valutazione della conformità previste in altro capo del regolamento (eccezionalmente e temporaneamente derogabili per ragioni di pubblica sicurezza o di protezione della vita e della salute, dell'ambiente e dei principali beni industriali e infrastrutturali ai sensi dell'art. 47).

Il terzo ramo relativo alla disciplina obbligatoriamente applicabile è quello disciplinato al Titolo IV, cioè relativo ai sistemi di IA:

- 1) destinati a interagire con le persone fisiche;
- 2) utilizzati per rilevare emozioni o stabilire un'associazione con categorie (sociali) sulla base di dati biometrici;
- 3) in grado di generare o manipolare contenuti ("deep fake").

Per queste specifiche tipologie sono previsti numerosi obblighi di trasparenza, in particolare gli utilizzatori dovranno essere perfettamente a conoscenza del fatto che stanno interagendo con un'Intelligenza Artificiale o che il contenuto che stanno consultando è stato generato da un'IA. Ciò salve alcune eccezioni particolari per finalità legittime, quali ad esempio il contrasto dei reati⁵².

Le norme successive disciplinano altri istituti con lo scopo di applicare concretamente i principi individuati precedentemente. Tra questi si citano la disciplina relativa alle istituzioni coinvolte nella gestione del fenomeno "Intelligenza Artificiale" nell'UE, le attività di monitoraggio successivo all'immissione sul mercato di un'IA ad alto rischio e l'attività di vigilanza più generale eseguita dalle Autorità di Vigilanza, sino alla definizione delle sanzioni.

2.5 Emendamento del settembre 2022

La Proposta della Commissione, in seguito alla sua pubblicazione, è poi giunta all'esame del Consiglio dell'Unione Europea e del Parlamento⁵³ dell'Unione Europea. In particolare, il Consiglio ha formulato due importanti emendamenti nel corso della propria attività di mediazione alla ricerca di un compromesso, dapprima trovato in una prima versione sotto la presidenza slovena e, in seguito, sotto la presidenza ceca in una seconda⁵⁴.

Descrivendo molto brevemente alcuni dei profili più rilevanti di questo nuovo testo normativo, si nota sin dall'inizio come sia radicalmente cambiata la definizione⁵⁵ di Intelligenza Artificiale impiegata dal regolamento. Senza anche in questo caso fare riferimento ai dibattiti sul tema, occorre portare alla luce un diverso approccio: la prima e più evidente modifica è certamente l'eliminazione dell'Allegato II contenente l'elenco delle tecnologie (modificabile con atti delegati dalla commissione) che possono essere trattate come Intelligenze Artificiali soggette alla disciplina della Proposta. Ciò sta a significare che non si adotta più un metodo a "lista chiusa", caratterizzato peraltro dall'assenza di un principio cardine di gran parte della normativa europea,

52. Per i quali la proposta prevede notevoli obblighi. Per un approfondimento di questi strumenti si legga: SARANITI F., AI: l'uso delle chatbot nel contrasto ai reati pedopornografici, AI4business, 29.1.2021

53. Si noti come il Parlamento abbia, in particolare, già istituito nel 2020 una commissione parlamentare ad hoc, denominata AIDA (Artificial Intelligence in a Digital Age) che ha redatto numerosissimi documenti di studio molto interessanti, rinvenibili al seguente Link: <https://www.europarl.europa.eu/committees/en/aida/home/publications?tabCode=working-papers>

54. Si precisa che nel novembre del 2022 è stata pubblicata una Versione finale di proposta di Regolamento da parte del Consiglio dell'Unione Europea, la quale non ha apportato significative modifiche rispetto alla versione di settembre del 2022 (la terza). Si ritiene sia opportuno segnalare in particolare l'importantissima esclusione dal novero dell'applicazione della normativa relativa gli utilizzi, da parte di utenti persone fisiche, non professionali di un'intelligenza artificiale, fatta eccezione per gli obblighi di trasparenza

55. ART. 3 N.1 Proposta di Regolamento 2021/0106(COD) 'artificial intelligence system' (AI system) means a system that is designed to operate with a certain level of autonomy and that, based on machine and/or human-provided data and inputs, infers how to achieve a given set of human-defined objectives using machine learning and/or logic- and knowledge based approaches, and produces system-generated outputs such as content (generative AI systems), predictions, recommendations or decisions, influencing the environments with which the AI system interacts;

ossia quello della neutralità tecnologica⁵⁶, ma al contrario un metodo a “lista aperta”. Esso viene improntato ad uno schema di pensiero più descrittivo delle logiche di funzionamento della tecnologia che non al contenuto tecnico in sé e per sé considerato. Nello specifico vengono indicate le due grandi famiglie di tecnologie: machine learning based e logic-and knowledge based, le quali, sulla base di certi input offrono, con una certa autonomia, alcuni risultati quali per esempio contenuti, raccomandazioni, previsioni, ecc.

Una seconda modifica al testo normativo molto rilevante è l’aggiunta di un Titolo iniziale ad oggetto le Intelligenze Artificiali con scopo generale⁵⁷. Si potrebbe ritenere che la ragione di questa modifica sia stata la pubblicazione di un contributo da parte di una dottrina⁵⁸, nel quale si afferma come sia essenziale regolare anche queste tipologie di intelligenza artificiale. Un successivo documento⁵⁹ ha provato a chiarire i termini di tale necessità: quella di evitare cioè che ogni responsabilità di utilizzo di tali sistemi ricada sugli utenti degli stessi e non sugli sviluppatori, con ciò frustrando le esigenze del mercato europeo che richiedono un prodotto affidabile.

Tale titolo estende a tali tipologie di tecnologie l’applicazione della disciplina già vista per le intelligenze artificiali ad alto rischio, delegando tuttavia la Commissione a redigere atti delegati che abbiano lo scopo di rendere tale disciplina coerente e compatibile con le caratteristiche strutturali di un’Intelligenza Artificiale con scopo generale, diverse da quelle c.d. a scopo specifico. Le norme specificano che l’espressione “*is intended by the provider to perform generally applicable functions*” presente nella definizione, deve essere interpretata anche quale “*possible use of the general purpose AI systems*” nell’applicazione delle norme (adattate) del titolo III. Tale puntualizzazione sta a significare che, nell’analisi del rischio di siffatta intelligenza artificiale, è necessario tenere conto di un ampio numero di fattori, ciò in ragione della grande flessibilità che questa tipologia di tecnologia offre, la quale permette molteplici utilizzi. Tale ragionamento potrebbe apparire molto simile alla disciplina poc’anzi descritta in merito alla nozione di rischio, molto ampia e assai simile alle norme relative ai prodotti difettosi e alle normative di settore (ad esempio quella relativa all’immissione sul mercato o la messa a disposizione nel territorio dell’UE di dispositivi medici).

Vengono poi posti diversi obblighi in capo agli sviluppatori, quali ad esempio l’obbligo di tenuta per 10 anni di documentazione tecnica di cui all’art. 18 della Proposta, nonché quello di collaborare con altri produttori di Intelligenza Artificiale a scopo generale.

La nuova disciplina stabilisce poi un’eccezione a quanto appena riportato, che sussiste allorché il produttore dichiara (in buona fede) che l’IA non è ad alto rischio, nemmeno in caso di utilizzo improprio.

Senza la pretesa di esaustività, ulteriori modifiche molto rilevanti attengono a una parte rilevante della regolamentazione delle Intelligenze Artificiali ad Alto rischio. Ad esempio, le condizioni di applicazione di tali norme sono state specificate e ristrette per quanto concerne le IA riferite all’Allegato III⁶⁰, le quali saranno qualificate ad Alto rischio nel caso in cui il risultato del sistema non sia accessorio all’azione o decisione conseguente e, contestualmente, comporti un elevato grado di rischio (nella già più volte citata accezione). Anche i doveri di tenuta della documentazione tecnica sono stati oggetto di modifiche. Sono state razionalizzate le norme relative agli obblighi apposti ai vari soggetti appartenenti alla catena di valore dell’Intelligenza Artificiale (produttori, importatori, ecc.) e specificate le condizioni di applicazione della normativa a soggetti diversi dai fornitori (art. 23a), tra le quali si ricorda l’apposizione del marchio, una rilevante modifica, una modifica dello scopo previsto che trasformi un’IA in alto rischio o la connessione ad altri IA ad altro rischio. Altre modifiche molto rilevanti sono quelle in tema di organismi notificati, con innovazioni sia sul piano dei requisiti (stabilimento presunzioni conformità per organismi di notifica che rispettino certi standard ovvero normative armonizzate), nonché su quello delle procedure ivi stabilite. A proposito di standard, anche l’art. 40 è stato specificato con l’intento di descrivere in maniera specifica quali caratteristiche debbano avere gli standard applicabili all’Intelligenza Artificiale. Anche le misure in materia di sostegno all’innovazione (ossia le regulatory sandboxes) sono state radicalmente riscritte, includendo anche gli obiettivi che tali normative dovrebbero perseguire. Sotto il profilo della governance nuovi articoli stabiliscono Commissione UE (art.58a) debba adottare, su richiesta degli stati membri o del Comitato

56. “Neutralità tecnologica: la libertà delle persone e delle organizzazioni di scegliere la tecnologia più adeguata ai loro bisogni. I prodotti, servizi o quadri normativi che tengono conto del principio della neutralità tecnologica non impongono né introducono discriminazioni a favore dell’impiego di un tipo particolare di tecnologia” <https://eur-lex.europa.eu/IT/legal-content/summary/supporting-telecommunications-networks-and-digital-service-infrastructures-across-europe.html#:~:text=Neutralità%20tecnologica%3A%20la%20libertà%20delle,più%20adeguata%20ai%20loro%20bisogni>.

57. Per una definizione di Intelligenza Artificiale a scopo generale si veda l’art. 3 par. 1 b della Proposta come emendata nel settembre 2022

58. MULLER C. E ALTRE, AIA in-depth #1 Objective Scope Definition Articles 1 - 4 & ANNEX I, ALLAI

59. FUTURE FOR LIFE INSTITUTE, General Purpose AI and the AI Act, 5.2022

60. Art. 6 Proposta di Regolamento: “the output of the system not purely accessory in respect of the relevant action or decision to be taken, and may therefore lead to a significant risk to the health, safety or fundamental rights”

sull'Intelligenza Artificiale, Linee guida per l'applicazione concreta del Regolamento, in particolare con riguardo alle norme relative ai requisiti e alle IA vietate e altre ancora. In particolare, con riguardo alla relazione del Regolamento con altre normative dell'Unione Europea.

2.6 Conclusione parziale

L'analisi appena condotta può già portare a una prima considerazione e ad un suo (possibile) rimedio. Come si è avuto modo di esprimere nel capitolo introduttivo e nel prosieguo (in particolare nel punto ove si è trattato della natura dei rischi), l'Intelligenza Artificiale ha avuto e avrà un notevole impatto sulla vita delle persone per molte ragioni e in diversi modi, incidendo fortemente sull'essere umano in quanto tale, non solo da un punto di vista economico o sociale ma anche più propriamente sotto il profilo psicologico o di essente.

Una dottrina⁶¹ ha pertanto ritenuto, alla luce di tali caratteristiche, la Proposta di Regolamento criticabile: la Persona come essere umano, in questo ampio scenario normativo, che costituisce il soggetto passivo di una tecnologia così incisiva, sarebbe del tutto assente⁶². Altra dottrina ha evidenziato come la mancanza di modi di impugnazione delle decisioni di un'Intelligenza Artificiale renda sostanzialmente inesistente la protezione dei diritti fondamentali delle persone⁶³.

In ogni caso, pare comunque (e quantomeno) allo scrivente come se le premesse contenute all'interno degli Orientamenti etici, ma non solamente in tale documento, fossero state disattese. La prospettiva non è più quella delle origini di una disciplina *human-centred* ma quella di una normativa *tech-centred*. Questo (anche) a causa del risk-based method impiegato dal legislatore europeo, che pone al centro la tecnologia⁶⁴ e non il suo impatto sulla vita delle Persone.

Alla luce del percorso ricostruttivo appena compiuto, ecco un (possibile) rimedio: per recuperare la centralità dell'essere umano in modo non incompatibile con il testo normativo potrebbe essere opportuno orientare l'interpretazione di alcune norme del Regolamento (in particolare gli artt. 14 e 69, ossia la disciplina dei codici di condotta) nel senso di una maggiore presenza dell'essere umano nel discorso, recuperando altresì allo stesso tempo anche la centralità della Carta di Nizza.

3 II I Codici di condotta di cui all'art. 69 della Proposta di Regolamento

3.1 Della natura dei codici di condotta

In via di premessa è opportuno evidenziare come l'emersione dei Codici di condotta nel panorama normativo europeo e nazionale non costituisca una novità specifica della regolazione dell'Intelligenza Artificiale, ma al contrario rientra all'interno di un fenomeno di portata molto più ampia. Si inserisce, in particolare, all'interno della distinzione dogmatica⁶⁵ fra fonti formali e fonti materiali del diritto e si collega al concetto di autoregolazione.

I codici di condotta potrebbero essere descritti⁶⁶ come atti di autonomia privata che possono concorrere, a determinate condizioni laddove previsto (o imposto) dalla legge o più in generale dalla normativa di riferimento, a integrare le norme giuridiche in senso formale. Si è appena citato il concetto di "descrivere", perché è molto complesso rintracciare un elemento unitario in tutti questi codici di condotta, in virtù del fatto che ciascuno ha una propria peculiarità. In particolare, è possibile trovare notevoli

61. CONTISSA G., GALLI F., GODANO F., SARTOR G., Il regolamento europeo sull'intelligenza artificiale, i-lex, 2021, fasc. 2

62. Si ritiene opportuno specificare come tale contributo si riferisca alla versione della Proposta del 2021, ma è possibile sostenere, alla luce del fatto che i cambiamenti apportati non incidono su questa dimensione, che tali considerazioni siano valide anche per quanto concerne quella del settembre 2022.

63. EBERS M. E ALTRI, *The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)*, Multidisciplinary scientific Journal, 2021 "Although the regulation is intended to protect fundamental rights, it lacks remedies by which individuals can seek redress for a breach of the regulation. In particular, the draft does not foresee any mechanism to facilitate individuals' recourse against AI-driven decision-making. This may be in line with the product safety regulatory approach, but it may fundamentally challenge that approach. In any case, individual rights should be well coordinated with existing entitlements"

64. Peraltro, si consideri come, nelle proposte appena analizzate, manchi una disciplina specifica ed approfondita relativa alle Intelligenze Artificiali generative di altre Intelligenze Artificiali.

65. TORRENTE A., SCHLESINGER P., Manuale di diritto privato, Giuffrè, 2021

66. Codici di condotta, Digesto Civile, Aggiornamento, Utet, Torino, 2011, 163

distinzioni fra gli effetti giuridici di tali atti, che possono talora meramente orientare il comportamento dei soggetti cui sono rivolti mediante meccanismi di soft-law e quindi legati alla forte autorevolezza del soggetto autore di tali documenti; oppure in alcuni ambiti possono comportare, nel caso di loro violazione, vere e proprie sanzioni (di vario tipo) o comunque una intrinseca “giustiziabilità” dei diritti in essi previsti. Talvolta, poi, l’adozione di codici di condotta può quantomeno ridurre, se non escludere del tutto, una certa responsabilità in capo ad una persona fisica o giuridica per un certo fatto.

Il Comitato economico e sociale europeo⁶⁷ ha studiato il tema molto approfonditamente, giungendo a emanare un Parere nel 2015. All’interno di tale documento viene definita l’autoregolamentazione (ovviamente nel seppur ampio ma limitato contesto della normativa europea) nel seguente modo: “*si designa genericamente, quando ci si riferisce al comportamento economico, l’adozione da parte degli attori economici di certe regole di condotta nelle relazioni reciproche oppure nei confronti di terzi sul mercato e nella società, regole il cui rispetto è frutto di un accordo tra gli stessi attori, senza meccanismi coercitivi esterni*”. Il parere individua anche ciò che non ritiene possa essere definito come autoregolamentazione, ossia “soft law” o il “diritto non vincolante”, come gli atti preparatori o informativi (ad esempio i white paper o i green paper), gli atti interpretativi (ad esempio, le comunicazioni della commissione, le dichiarazioni, le risoluzioni del parlamento UE e, infine, le raccomandazioni e i pareri di cui all’articolo 288 del TFUE. Peraltro, anche tale fonte puntualizza che l’autoregolamentazione, in assenza di una base giuridica solidamente ancorata alle fonti europee, non può essere considerata alternativa alle norme in senso formale ma meramente integrativa⁶⁸ e solo laddove essa sia prevista. Il Comitato si propone di promuovere l’adozione di una riforma dell’Accordo interistituzionale AII attuativo del programma “Legiferare meglio” che includa anche una disciplina dell’autoregolazione (e per essere precisi anche la coregolamentazione).

Nel parere si ritrovano le caratteristiche che le norme di autoregolamentazione dovrebbero possedere per essere coerenti con i principi e i valori fondativi del diritto dell’Unione Europea. Tra questi si vogliono citare quello della conformità al diritto UE e al diritto internazionale, la trasparenza e la pubblicità, la rappresentatività delle parti interessate e la preliminare consultazione di queste e altre ancora. Su di un principio conviene soffermarsi nello specifico, in quanto costituisce un punto cruciale per il presente contributo, in virtù del fatto che porrebbe dei rigorosi limiti di applicabilità di atti di autoregolazione nel caso in cui “*si tratti di definire diritti fondamentali*”. Si chiarisce pertanto che il contenuto dei codici di condotta non può – ovviamente – spingersi sino a definire i diritti fondamentali della persona, ciò in piena applicazione del principio per il quale non è possibile disporre dei diritti fondamentali, in quanto intangibili. Ebbene, il comitato individua tre tipologie di distinzioni dell’autoregolamentazione: quella pubblica, perché sorvegliata da un organismo pubblico, quella giuridica (consuetudinaria o giurisprudenziale) e infine quella nazionale o transnazionale.

Spesso la nascita e l’evoluzione dei Codici di condotta è stata dettata⁶⁹ dall’emersione di esigenze di uniformazione del mercato, per una migliore circolazione dei beni giuridici in un mondo sempre più connesso a causa della globalizzazione, mediante alcuni standard e la nascita della c.d. *nuova lex mercatoria*⁷⁰; mentre in altri casi la causa è da rintracciare nell’elevata tecnicità dei temi oggetto dell’autoregolamentazione privata, di cui i privati sono certamente più esperti, mentre gli stati, autori delle leggi in senso formale, non sono spesso in grado di produrre norme efficaci perché deficitarie di quelle competenze essenziali allo scopo (si parla da più parti e da molto tempo ormai di abdicazione della funzione normativa da parte degli stati verso la direzione di un pluralismo delle fonti giuridiche⁷¹). Ulteriore circostanza che amplifica l’effetto di tali fonti del diritto è certamente il notevole grado di autorevolezza esercitato dagli autori di queste norme giuridiche private, spesso associazioni o raggruppamenti di operatori economici molto rilevanti nei rispettivi campi di azione.

67. Parere del Comitato economico e sociale europeo sul tema “Autoregolamentazione e coregolamentazione nel quadro legislativo dell’UE”, del 4 settembre 2015 (2015/C 291/05)

68. Ibidem: “*quel che manca è una norma abilitante che autorizzi queste istituzioni a rinunciare al loro potere legislativo e a delegarlo ai suddetti meccanismi, quale alternativa legittima agli strumenti di regolamentazione dell’UE definiti nei Trattati*”.

69. FONDAZIONE ITALIANA DEL NOTARIATO, *Autonomia privata e fonti del diritto*, 2019

70. Seppur sia difficile individuare una definizione di tale fenomeno giuridico, una dottrina (LUCIANI A.M., *La nuova lex mercatoria*, in Osservatorio sulle fonti, n. 1/2021) ha individuato nella spontaneità un elemento comune a tutte le opinioni: “*La Lex mercatoria è da tutti riconosciuta come un insieme di regole giuridiche create spontaneamente dagli stessi attori del commercio internazionale e finalizzate a disciplinare, nello specifico, le relazioni commerciali internazionali*”.

71. CARETTI P., *La “crisi” della legge parlamentare*, Osservatorio sulle fonti, 2010

3.2 Breve analisi dei codici di condotta di cui al GDPR

Un ambito molto affine a quello relativo all'Intelligenza Artificiale è sicuramente quello del c.d. GDPR, il Regolamento UE 679/2016, pertanto può essere utile effettuare preliminarmente una rapida analisi dei codici di condotta ivi contemplati. Costituiscono, secondo lo European Data Protection Board⁷², “*strumenti di responsabilizzazione volontari che stabiliscono specifiche norme di protezione dei dati per categorie di titolari e di responsabili del trattamento*”.

All'art. 40 del Regolamento viene imposto ad alcuni enti ed istituzioni l'obbligo di incoraggiare le associazioni o gli organismi rappresentanti categorie di titolari di trattamento nell'elaborare questi documenti per contribuire alla corretta applicazione del Regolamento, tenendo conto dei vari settori di trattamento e delle esigenze di ogni tipologia di impresa. Ai sensi dell'art. 24 paragrafo 3 del GDPR, infatti, l'adesione a questi strumenti di autoregolazione può costituire un mezzo per dimostrare, nell'ottica del principio *accountability* informatore della disciplina, l'applicazione del Regolamento europeo da parte del Titolare del trattamento dei dati personali. La loro funzione⁷³, tuttavia, potrebbe non essere limitata solamente a questo, ma potrebbe avere anche una notevole *vis expansiva* per un coordinamento di settore ovvero intersettoriale di molti Titolari di trattamento di un certo tipo o di più tipi, oltre che per colmare alcune differenze rilevanti nell'applicazione del Regolamento fra i paesi membri. Il tutto “*tenuto conto del potenziale rischio del trattamento per i diritti e le libertà delle persone fisiche*” (Considerando 98 del GDPR).

Nell'ottica di una più efficace, economicamente sostenibile e pratica applicazione di tale complesso normativo (spesso assai grave per le Piccole e medie imprese) la norma esemplifica alcuni possibili settori di intervento di tali codici, tra i quali potrebbe rientrare la disciplina dell'utilizzo della tecnica di pseudonimizzazione e i legittimi interessi perseguiti dal responsabile del trattamento in contesti specifici.

In applicazione dell'importante principio di trasparenza viene istituito un procedimento per l'approvazione del Codice di condotta, che parte da un'iniziativa privata, ossia un progetto di codice,⁷⁴ e passa per un parere di un ente pubblico quale l'autorità di controllo: condizione per l'approvazione è sussistenza di sufficienti garanzie adeguate. In seguito, in caso di approvazione, la medesima autorità registra e pubblica il codice. Nel caso in cui il codice disciplini attività di trattamento in vari Stati membri il par. 7 integra la procedura con un parere del comitato e con l'approvazione della commissione nonché la successiva pubblicazione. Anche il comitato ha l'obbligo di raccogliere in un registro tutti i codici di condotta e renderli pubblici con mezzi appropriati.

In ultima analisi, è opportuno segnalare che l'adozione di Codici di condotta in tale ambito influisce anche sul regime sanzionatorio in caso di violazione di una delle norme del Regolamento. Infatti, in tale evenienza, dovrebbe essere tenuta in considerazione, in merito alla quantificazione della sanzione amministrativa pecuniaria, anche l'adozione di un Codice di condotta da parte del Titolare del trattamento.

3.3 Disciplina della Proposta di Regolamento in materia di Codici di condotta

Il Titolo IX della Proposta di Regolamento fatta dalla Commissione nel 2021 in materia di Intelligenza Artificiale contempla un'unica norma, ossia l'art. 69, che disciplina i Codici di condotta.

In essa viene stabilito come la Commissione e gli Stati membri debbano promuovere la volontaria redazione – da parte di soggetti privati fornitori ovvero anche delle loro organizzazioni rappresentative, anche con la partecipazione dei portatori di interessi o delle loro organizzazioni rappresentative – di codici di condotta con lo scopo di applicare ai sistemi di IA non ad alto rischio i requisiti di cui al titolo III, capo 2, ossia tutte quelle norme già viste in materia di azioni positive obbligatorie che i fornitori di IA di tale tipo devono porre in essere. Ciò dovrà essere fatto sulla base di tecniche e soluzioni, caratterizzate dall'attributo della adeguatezza, per garantirne la conformità in vista della finalità prevista dei sistemi.

Inoltre, la norma fa riferimento ad alcuni requisiti supplementari che potranno essere oggetto di Codici di condotta: “*la sostenibilità ambientale, l'accessibilità per le persone con disabilità, alla partecipazione dei portatori di interessi alla progettazione e allo sviluppo dei sistemi di IA e alla diversità dei gruppi che si occupano dello sviluppo sulla base di obiettivi chiari e indicatori chiave di prestazione volti a misurare il conseguimento di tali obiettivi*”.

72. Linee guida 1/2019 sui codici di condotta e sugli organismi di monitoraggio a norma del regolamento (UE) 2016/679 del 4.6.2019

73. Ibidem

74. Preliminarmente, è necessaria una valutazione di ammissibilità del progetto. A tal fine le Linee guida offrono un'interessante Lista di controllo per la presentazione (Appendice 3)

La norma prevede poi che i codici di condotta possano avere un'applicazione orizzontale con riferimento a uno o più sistemi di IA, purchè sia tenuta in conto la similarità delle finalità da queste previste. Infine, fornitori di piccole dimensioni e delle start-up dovranno essere tenuti particolarmente in conto nelle azioni di predisposte dalla Commissione per incoraggiare l'adozione dei Codici di condotta.

La versione della Proposta di Regolamento, come emendata dal Consiglio dell'Unione nel settembre del 2022, contempla alcune modifiche di non eccessivo rilievo al testo normativo. La natura dei codici di condotta non cambia, ma viene ridotta l'enfasi attorno a questi documenti, rendendo più ragionevole il loro impiego: in particolare l'obiettivo di applicare le norme per le IA ad alto rischio viene ridotto nel senso di consentire che anche solo uno dei requisiti possa essere riprodotto entro un limite che è quello della massima estensione possibile e tenendo conto dei limiti tecnici di queste applicazioni. Pare come se il legislatore europeo abbia deciso di imporre "meno" in modo da facilitare l'adozione di prassi virtuose ad una più ampia parte del tessuto economico. Ulteriore modifica attiene all'inclusione, all'interno dell'elenco di esempi relativi a specifici requisiti di cui al par. II, anche quello di una programmazione efficiente sotto il profilo energetico, nonché con riguardo a codici di condotta applicabili agli utenti di Intelligenza Artificiale. Ancora, viene inclusa la possibilità di coinvolgere utenti nella redazione dei codici di condotta con riguardo alle loro obbligazioni.

4 III L'applicazione dello "Human Rights Approach" ai codici di condotta per l'IA

4.1 In cosa consiste lo "Human Rights Based Approach" o HRBA

La fonte-guida nella descrizione di cosa sia lo Human Rights Approach, ai fini del presente paragrafo, sarà il documento "UN Statement of Common Understanding on Human Rights-Based Approaches to Development Cooperation and Programming" adottata dallo United Nations Development Group (UNDG) nel 2003.

Di seguito ne verranno riassunti i contenuti principali, ma preliminarmente conviene subito precisare che tale documento deve essere riadattato alle specifiche caratteristiche dell'Intelligenza Artificiale; ciò in quanto il suo scopo (dichiarato espressamente) non era altro che dare alle Agenzie delle Nazioni Unite un framework uniforme mediante il quale applicare, nella loro azione, i diritti fondamentali. Sino a quel momento, infatti, tali enti hanno effettivamente applicato i diritti fondamentali come metodo di lavoro⁷⁵, ma ciascuno secondo una propria interpretazione e questo ha portato difficoltà di coordinamento. Tuttavia, poter considerare l'Intelligenza Artificiale alla stregua di una vera e propria Istituzione pubblica quanto a disciplina appare, allo scrivente, come una soluzione ottimale, alla luce del potenziale enorme effetto che, come si è potuto vedere, tale famiglia di tecnologie avrà in futuro. Ciò con gli opportuni limiti della ragionevolezza (e tenendo bene in considerazione come il presente contributo sia limitato a proporre un Codice di condotta volontario ai sensi dell'art. 69 della Proposta di Regolamento).

Le fonti dei diritti fondamentali citate espressamente dallo *Statement* sono la Dichiarazione Universale dei Diritti Umani approvata e proclamata il 10 dicembre 1948 dall'Assemblea Generale delle Nazioni Unite e la Dichiarazione di Vienna e Programma d'azione adottati dalla Seconda Conferenza Mondiale delle Nazioni Unite sui diritti umani il 25 giugno 1993. Pur tuttavia esse non possono considerarsi esclusive, in quanto in tutto il documento sono presenti punti di apertura a ulteriori diritti fondamentali ivi non disciplinati direttamente.

L'approccio proposto dalle Nazioni Unite è caratterizzato da tre pilastri che vengono riportati tradotti per comodità di lettura:

- 1) "Tutti i programmi di cooperazione allo sviluppo, le politiche e l'assistenza tecnica dovrebbero promuovere la realizzazione dei diritti umani come stabilito nella Dichiarazione universale dei diritti umani e in altri strumenti internazionali sui diritti umani".

75. Si precisa che sino al 1997 le Agenzie ONU applicavano il "Basic needs approach" ben descritto da RICE K. E GIRVIN H., *Human Rights-Based Approach to Short-Term Study Abroad*, Springer Cham, 2021 come in grado di introdurre "greater objectivity into the process of selecting who is helped and how by using evidence to demonstrate need and introducing effective and efficient interventions to improve the lot of the needy and society as a whole". Tuttavia, ciò che viene realizzato riflette l'ideologia delle classi dominanti, non quella delle classi in difficoltà e ciò può portare frustrazione.

- 2) *“Gli standard sui diritti umani contenuti nella Dichiarazione universale dei diritti umani e altri strumenti internazionali sui diritti umani e i principi derivati da essa guidano tutta la cooperazione allo sviluppo e la programmazione in tutti i settori e in tutte le fasi del processo di programmazione”.*
- 3) *“I programmi di cooperazione allo sviluppo contribuiscono allo sviluppo delle capacità dei portatori di doveri di adempiere ai propri obblighi e dei “titolari di diritti” a rivendicare i propri diritti”*

Analizzati questi pilastri, lo Statement propone alcune caratteristiche necessarie del Metodo. Il primo è la necessità di utilizzare buone pratiche di programmazione. Come si è visto, sin dall’inizio della programmazione delle attività delle Agenzie è necessario effettuare una valutazione ed una analisi per l’identificazione dei diritti umani rivendicati dai titolari e gli obblighi corrispondenti dei c.d. portatori di doveri. Inoltre, devono essere individuate anche le ragioni dirette, sottostanti e strutturali della non applicazione dei diritti fondamentali. Devono inoltre essere analizzate (e sviluppate) le capacità dei titolari di diritti di far valere i propri diritti e dei correlativi titolari di doveri di adempiere ai propri obblighi. Deve essere previsto un monitoraggio e una valutazione dei risultati delle azioni delle Agenzie e devono essere tenute in conto concretamente le raccomandazioni degli organismi internazionali per i diritti umani.

Alcuni degli altri profili molto importanti sono: il riconoscimento delle persone come attori chiave nel proprio sviluppo, la partecipazione come mezzo e contemporaneamente come scopo, il focus sui gruppi marginalizzati, svantaggiati ed esclusi e altro ancora. Anche in tale documento viene sottolineata l’importanza del principio dell’Accountability.

Una prima applicazione interessante del metodo descritto, benchè non esattamente comparabile al tema oggetto del presente contributo, può essere quella che è stata fatta dall’Alto Commissario per i diritti Umani nell’ambito della lotta alla povertà⁷⁶ nel 2006. È interessante in quanto permette di chiarire la vera utilità di un siffatto approccio, ossia la capacità di osservare un fenomeno sociale con una prospettiva multidimensionale⁷⁷. La povertà, infatti, non sarebbe mera carenza di beni necessari, ma sarebbe qualcosa di molto più complesso e variabile in base alle diverse situazioni concrete. Altro profilo evidenziato di grande utilità sarebbe la preesistenza di un apparato normativo già completo, ossia le Dichiarazioni internazionali o regionali ampiamente note, sulle quali basare la predisposizione di politiche attive⁷⁸ di eliminazione della povertà. Dunque, l’impegno profuso nel lavoro sarebbe inferiore. È facile constatare come tali affermazioni siano facilmente applicabili anche fuori dal contesto specifico trattato dal documento.

Una dottrina⁷⁹ sottolinea ancora che l’approccio basato sui diritti fondamentali ha il grande merito di essere olistico, perché include numerosi aspetti della persona che non sarebbero contemplati nella prospettiva dei bisogni o in quella “caritatevole”.

Anche la c.d. Agenda UN 2030⁸⁰ adottata nel 2015, che propone 17 Sustainable Development Goals (SDG) e 169 target da raggiungere entro il 2030 per l’eradicazione della povertà tra gli esseri umani e per mettere in sicurezza il pianeta Terra pone grande attenzione al metodo. Si propone di raggiungere l’obiettivo dello Sviluppo Sostenibile, che si articola su tre fronti: economico, sociale e ambientale. Ciascuno di questi termini del discorso è fortemente influenzato dai diritti umani, perché una società sostenibile è una società che rispetta i diritti fondamentali: *“The new Agenda recognizes the need to build peaceful, just and inclusive societies that provide equal access to justice and that are based on respect for human rights”.*

76. OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR HUMAN RIGHTS, Principles and guidelines for a human rights approach to poverty reduction strategies, HR/PUB/06/12

77. *“The human rights approach underlines the multidimensional nature of poverty, describing poverty in terms of a range of interrelated and mutually reinforcing deprivations, and drawing attention to the stigma, discrimination, insecurity and social exclusion associated with poverty. The deprivation and indignity of poverty stem from various sources, such as the lack of an adequate standard of living, including food, clothing and housing, and the fact that poor people tend to be marginalized and socially excluded. The commitment to ensure respect for human rights will act as a force against all these forms of deprivation”.*

78. *“policies and institutions for poverty reduction should be based explicitly on the norms and values set out in international human rights law”*

79. RICE K. E GIRVIN H., Human Rights-Based Approach to Short-Term Study Abroad, Springer Cham, 2021 *“A rights-based approach brings a holistic perspective with regards to civil, political, social, economic, and cultural roles we hold as human beings and a more holistic understanding of well-being that goes beyond the meeting of material needs. Our understanding of human rights is always evolving, and our methods, practices, research, interventions, and processes should evolve as our understanding deepens”.*

80. Adottata il 25.9.2015 dall’Assemblea Generale delle nazioni unite con la Risoluzione *“Transforming our world: the 2030 Agenda for Sustainable Development”* A/RES/70/1

Lo “United Nations sustainable development cooperation framework”⁸¹ pubblicato il 3.6.2019 esplicita molto chiaramente questo aspetto, allorchè evidenzia come il metodo permetta di individuare chiaramente chi sono i beneficiari di politiche tese allo sviluppo sostenibile e i soggetti viceversa obbligati a realizzarle. Un passo del documento è particolarmente interessante, laddove viene espressamente indicata la necessità di applicare lo HRBA anche al trattamento dei dati personali⁸². Il framework offre poi un “Outline of the UN Common Country Analysis”⁸³ Report for the Cooperation Framework (Allegato 2)”. Un ruolo essenziale è dato dalla fase nella quale “Identify human rights standards relevant to development challenges, map relevant international obligations and commitments made by the country in terms of human rights, LNOB, gender and the environment and connect recommendations from human rights mechanisms with the analysis on challenges and opportunities for SDG achievement”.

4.2 L’utilizzo del HRBA nell’ambito dei Codici di condotta di cui all’art. 69 della Proposta di Regolamento UE sull’Intelligenza Artificiale

Ad un primo sguardo, alla luce dello scopo dato dal Regolamento ai Codici di condotta di applicare uno o alcuni dei requisiti stabiliti per le IA ad alto rischio al di fuori di questa categoria, non pare ci sia spazio per un’applicazione in un senso differente da quello offerto dalla norma in questione.

È purtuttavia vero che la definizione di Codici di condotta presente proprio nel contesto europeo (nel già visto Parere del 2015) lascia ampio spazio all’autonomia delle parti. Certamente questa indicazione di matrice europea non è vincolante, ma dà conto di un sostrato dottrinale e culturale sul punto ribadito anche dal visto confronto con altri codici di condotta di altre discipline. Confronto che ha peraltro permesso di evidenziare l’ampiezza della libertà offerta alle parti, i cui unici limiti sono spesso rappresentati – ovviamente – dai diritti fondamentali.

Proprio la libertà e l’autonomia di cui le parti firmatarie godono potrebbe permettere un’interpretazione estensiva di alcune norme della Proposta di Regolamento. La norma del Titolo III che potrebbe fungere da strumento interpretativo in tal senso potrebbe essere l’art. 14 in materia di sorveglianza umana. Questa offre subito un obiettivo: “Human oversight shall aim at preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse, in particular when such risks persist notwithstanding the application of other requirements set out in this Chapter”.

Per una corretta implementazione dello scopo, ai sistemi di sorveglianza umana sono necessarie diverse caratteristiche, come stabilito dalla norma:

- a) La comprensione delle capacità e dei limiti dell’Intelligenza Artificiale;
- b) Evitare il c.d. automation bias⁸⁴;
- c) Una corretta interpretazione dell’output;
- d) La possibilità di non usare l’IA;
- e) Interrompere le operazioni in qualunque momento.

81. Par. 19 “The Human Rights-Based Approach to Development is a conceptual framework for the process of sustainable development that is normatively based on international human rights standards and principles and operationally directed to promoting and protecting human rights. Under the HRBA, the plans, policies and processes of development are anchored in a system of rights and corresponding obligations established by international law, including all civil, cultural, economic, political and social rights, and the right to development. HRBA requires human rights principles (equality and non-discrimination, participation, accountability) to guide UN development cooperation, and focus on capacity development of both ‘duty-bearers’ to meet their obligations and ‘rights-holders’ to claim their rights”.

82. Par. 36 “The UN CCA will analyse existing data and data gaps for national SDG indicators. It should go beyond official national statistics to use new sources of data and diagnostic tools, including but not limited to big data, national surveys and assessments, targeted surveys using mobile technology and others. This should be done in accordance with the human rights-based approach to data, international data protection standards and the UN Principles on Personal Data Protection and Privacy”.

83. Par. 33 “The UN CCA is the UN system’s independent, impartial and collective assessment (i.e., a description of a country situation) and analysis (i.e., a description of causes and their implications) of a country situation for its internal use in developing the Cooperation Framework. It examines progress, gaps, opportunities and bottlenecks vis-à-vis a country’s commitment to achieving the 2030 Agenda, UN norms and standards, and the principles of the UN Charter, including as reflected in the Cooperation Framework Guiding Principles”

84. CASCELLA L., *Artificial Intelligence Risks: Automation Bias*, MedProGroup

In un'ottica interpretativa che faccia leva sulla ratio quantomeno degli atti introduttivi alla Proposta di Regolamento, già ampiamente descritta nel primo capitolo come tesa alla tutela dei diritti fondamentali delle persone, i Codici di condotta potrebbero essere utilizzati (anche) al fine di implementare la sorveglianza umana mediante un diverso approccio, poc'anzi descritto come *Human Rights Approach*.

Quest'ultimo rappresenterebbe, in questa prospettiva, uno dei possibili modi per ottenere queste cinque caratteristiche e, in particolare, la prima: per una corretta ed effettiva comprensione dell'Intelligenza Artificiale, infatti, un possibile strumento potrebbe essere quello di un'analisi approfondita della portata dell'impatto⁸⁵ della singola Intelligenza Artificiale e delle sue specifiche funzioni su ciascun diritto fondamentale previsto dalle fonti internazionali e, soprattutto in un'ottica europea, anche continentale. Allo scopo di evitare il c.d. automation bias, dovrebbe essere analizzato l'effettivo e concreto impatto dell'Intelligenza Artificiale sulla libertà dell'utente, il che presuppone anche un'effettiva conoscenza (giuridica) di tale diritto. Per garantire una corretta interpretazione dell'output dovrebbe essere analizzato il modo con il quale tale risultato viene posto all'utente, in quale contesto e in quale momento dell'assunzione della relativa decisione, con ciò applicando – se necessario – i principi di motivazione dei provvedimenti amministrativi o giurisdizionali e predisponendo dei corretti mezzi di impugnazione delle decisioni algoritmiche. Il non uso e l'interruzione dell'uso dovrebbero essere calibrati in base ai concreti impatti che questi eventi potrebbero avere sull'utente.

A conferma del percorso interpretativo appena esposto, l'art. 69, nell'elencare i requisiti supplementari, cita il seguente: lo “sviluppo sulla base di obiettivi chiari e indicatori chiave di prestazione volti a misurare il conseguimento di tali obiettivi”. Gli obiettivi chiari potrebbero essere rappresentati dalla tutela dei diritti fondamentali della persona in modo da riportare la persona al centro della disciplina e gli indicatori chiave di prestazione, invece, potrebbero essere i livelli di impatto (da quantificare e qualificare sulla base di parametri oggettivi) della tecnologia su questi diritti.

Per ottenere una vera visione olistica del fenomeno, dovrebbero essere inclusi nel metodo non solo aspetti che coinvolgono direttamente il rischio dell'Intelligenza Artificiale, ma anche le opportunità che questa offre: in sintesi, per una migliore comprensione⁸⁶ dell'intelligenza artificiale non dovrebbe essere guardato solamente il profilo dei rischi che questa comporta, ma anche quello delle opportunità.

4.3 I vantaggi di un utilizzo dello HRBA nell'ambito dei Codici di condotta per l'Intelligenza Artificiale

L'utilizzo del HRBA quale base di partenza per l'elaborazione di un Codice di condotta potrebbe offrire numerosi vantaggi a tutte le persone coinvolte dalle decisioni di un'Intelligenza artificiale.

Il primo vero vantaggio sarebbe quello di riportare l'essere umano (in una prospettiva multidimensionale) al centro di un progetto tecnologico, creando allo stesso tempo un movimento culturale che potrebbe vedere effettivamente l'Europa quale viatico per lo sviluppo dei diritti fondamentali.

In secondo luogo, inciderebbe in maniera molto efficace sul design dell'IA, perché richiederebbe da un lato la cooperazione alla programmazione di competenze umanistiche quali quelle del giurista, dello psicologo e/o del filosofo esperto di diritti fondamentali e dall'altro una progressiva emersione di una maggiore sensibilità ai temi dei diritti fondamentali da parte delle competenze STEM (Science, Technology, Engineer, Maths). Ciò potrebbe portare ad un produttivo dialogo fra diverse persone in grado peraltro arricchire il panorama culturale e creare maggiore coesione sociale⁸⁷. Condizione essenziale perché questo scenario possa verificarsi, tuttavia, sarebbe la nascita di corsi interdisciplinari post-universitari per fornire a tutti coloro che fossero interessati una base di conoscenze comune. Per una vera collaborazione, tuttavia, gli studenti di tali corsi dovrebbero essere contemporaneamente insegnanti verso i propri colleghi con riferimento alle materie di specializzazione. Così un giurista esperto di diritto civile potrebbe insegnare la differenza fra diritti disponibili e diritti indisponibili (per una parte del tempo del corso) ad un fisico che si occupa di statistica e quest'ultimo potrebbe insegnare al primo la portata delle Leggi di Boltzmann.

85. CONSIGLIO D'EUROPA, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, 2019 propone uno “Human rights impact assessments” o (HRIA): “This self-assessment should evaluate the potential impact of the AI system on human rights taking into account the nature, context, scope, and purpose of the system”

86. DONATI F., *Diritti fondamentali e algoritmi nella proposta di regolamento sull'intelligenza artificiale*, in *La rivoluzione dell'intelligenza artificiale: profili giuridici*, a cura di F. DONATI, A. PAJNO E A. PERRUCCI

87. SANSAVINI S., *Il difficile dialogo fra le culture umanistica e scientifica*, *Il carrobbio*, 2014-2015

A tutela di Intelligenze Artificiali realizzate secondo questi criteri potrebbe essere realizzata un'apposita Certificazione (in applicazione della Regulatory sandboxes), unita a fondi europei destinati alle Piccole e Medie Imprese o alle Start-Up per il sostegno e finanziamento di progetti così caratterizzati.

Un ulteriore vantaggio sarebbe quello di ottenere una migliore analisi dell'ambiente socioeconomico di utilizzo dell'Intelligenza Artificiale, andando alla ricerca delle conseguenze dell'applicazione di un'Intelligenza Artificiale nelle sfere dei diritti dei soggetti direttamente ed indirettamente rilevanti per gli output, in particolare:

- a) I portatori di diritti, coloro che subiscono le decisioni di un'IA;
- b) I portatori di doveri, coloro agiscono mediante le decisioni di un'IA o che agiscono con le IA;
- c) Le situazioni intermedie, nelle quali un soggetto risulti essere sia un portatore di diritti sia un portatore di doveri;
- d) I soggetti apparentemente esterni al rapporto che si instaura fra i soggetti dei punti a), b) e c) ma che subiscono indirettamente l'influenza delle decisioni, come ad esempio soggetti concorrenti degli utenti.

L'impiego di tale metodologia sarebbe molto coerente con gli scopi iniziali individuati dall'Unione Europea, in particolare potrebbe facilitare l'adozione di una prospettiva etica all'Intelligenza Artificiale e questa potrebbe anche diventare – da rischio – uno strumento di sviluppo dei diritti fondamentali.

4.4 Modello di Codice di Condotta

A questo punto della trattazione si intende provare ad offrire un modello per un Codice di condotta Human-rights based method, tenendo come punto di riferimento il framework poc'anzi citato⁸⁸ e in ogni caso entro i limiti della norma di cui all'art. 69 della Proposta, pur secondo la particolare interpretazione esposta nel presente contributo.

Il Codice di condotta qui ipotizzato avrebbe un'applicazione generalizzata a qualunque tipologia di Intelligenza Artificiale, senza distinzioni di settore. Per tale ragione si ritiene che possa essere utile un approccio procedurale più che sostanziale, nel quale cioè sia stabilito non tanto il modo concreto di funzionamento di un'Intelligenza Artificiale, quanto piuttosto quello di ispirare un vero e proprio processo di realizzazione di un'IA, lasciando all'autonomia scientifica di chi effettivamente partecipi ai progetti il compito di prendere determinate scelte in base al settore specifico.

La natura procedurale appena esposta ha una diretta conseguenza: un'applicazione della disciplina solo a posteriori renderebbe le previsioni vane. Perciò, per garantire l'efficacia delle previsioni contenute nel Codice, i partecipanti al progetto dovrebbero adottarlo sin dall'inizio della prima progettazione dell'Intelligenza Artificiale. Ciò potrebbe avvenire per il tramite della sottoscrizione di clausole contrattuali *ad hoc* (nel caso di contratti di collaborazione fra lavoratori autonomi e/o imprese), ovvero dell'inclusione del Codice di condotta all'interno del mansionario nel caso di lavoratori subordinati. In quest'ultimo caso il datore di lavoro dovrebbe provvedere altresì alla formazione (interdisciplinare) del proprio personale su questi temi.

Il focus iniziale della disciplina del Codice, nell'ottica di una chiara procedura e nell'ottica di una *Human Oversight*, dovrebbe essere posto sulle persone che partecipano al progetto: in particolare dovrebbe essere prevista una chiara e specifica indicazione di tutti i soggetti - persone fisiche - che concretamente apportano un contributo diretto al prodotto, ciò a prescindere dalla nozione di autore di opera creativa presente all'interno della Legge sul diritto d'autore. Il codice dovrebbe offrire indicazioni sulle concrete modalità di composizione e documentazione del Team che esegue i lavori e sulle caratteristiche professionali dei membri per favorire l'interdisciplinarietà.

Il processo creativo dovrebbe partire dal concetto di funzione – o funzioni – previste della specifica Intelligenza Artificiale e, sulla base di queste, effettuare una Valutazione di Impatto sui diritti fondamentali previsti dalle Carte dei diritti fondamentali. Sulla base di questa, definire strategie tecniche ed organizzative (identificando cioè i soggetti portatori di diritti, i soggetti portatori di doveri, le situazioni intermedie e i soggetti esterni ma coinvolti indirettamente e come questi debbano rapportarsi) in grado di soddisfare i requisiti⁸⁹ stabiliti dalle fonti internazionali dei diritti fondamentali.

88. *UN Statement of Common Understanding on Human Rights-Based Approaches to Development Cooperation and Programming* (2003)

89. *Ibidem*. Si vuole specificare che le misure avrebbero lo scopo di garantire non i diritti fondamentali in sé, quanto i seguenti requisiti dei diritti fondamentali, i quali rappresentano pertanto un prerequisito: “*universality and inalienability; indivisibility; interdependence and inter-relatedness; non-discrimination and equality; participation and inclusion; accountability and the rule of law*”

Esaurita questa prima analisi ad oggetto le funzioni previste, sarebbe opportuno proseguire con un esame delle funzioni prevedibili mediante un metodo controfattuale basato sulle ipotesi di funzioni e sulla definizione ipotetiche strategie tecniche ed organizzative per l'implementazione dei medesimi requisiti relativi ai diritti fondamentali.

Inoltre, una valutazione di impatto comparativa, coordinata ed aggregata di tutte le funzioni dell'Intelligenza Artificiale per introdurre altrettante misure per la tutela dei requisiti dei diritti fondamentali. Inoltre, converrebbe esplicitare quali funzioni non sia possibile realizzare e cosa non sia opportuno che venga realizzato, nonché delle ragioni di queste valutazioni. Tale analisi dovrebbe tenere conto di tutti i profili tecnici, organizzativi, giuridici ed etici attinenti al tema e al settore.

Terminata la fase di progettazione ed eseguita (nonché documentata) la realizzazione, dovrebbe essere prevista altresì una procedura di Test specifica, che non si limiti solamente alla verifica di funzionamento dell'Intelligenza Artificiale, ma che la sottoponga ad un vero e proprio *Human-rights Stress Test*⁹⁰: in particolare sottoponendo all'IA input-informazioni provenienti da categorie di persone discriminate, marginalizzate, svantaggiate o gruppi esclusi ed esaminando gli output prodotti.

Nel caso in cui tali esami non portassero a risultati soddisfacenti, si dovrebbe ripetere la procedura dall'inizio in modo da apportare quelle modifiche che si rendessero necessarie alla tutela dei diritti fondamentali.

Al termine di questa fase dovrebbe essere redatto un documento (denominabile *Notifica di Esecuzione del Codice*) che racchiuda, con modalità da definire secondo standard da disciplinare, tutto quanto sino a questo momento effettuato ovvero deciso, ossia una trasposizione di tutte le scelte tecniche, organizzative, giuridiche ed etiche.

In ultima analisi, dovrebbe essere predisposto un sistema di monitoraggio, valutazione (quantitativa e qualitativa) e aggiornamento permanente, alla stregua di quanto previsto dalla Proposta di Regolamento per le Intelligenze Artificiali ad alto rischio. Inoltre, dovrebbe essere realizzato uno stabile sistema adatto a garantire l'impugnazione avanti a un terzo imparziale delle decisioni dell'Intelligenza Artificiale, purché entro gli stretti limiti della ragionevolezza della decisione medesima, secondo criteri predeterminati dal diritto antidiscriminatorio.

Bibliografia

AGENZIA DELL'UNIONE EUROPEA PER I DIRITTI FONDAMENTALI, *Preparare un giusto futuro l'Intelligenza Artificiale e i diritti fondamentali – Sintesi*, 2021

BERNABE' F., TREMONTI G. e IANNINI A., *La nuova economia fondata sui dati*, in "Big Data e Privacy. La nuova geografia dei poteri", Garante per la protezione dei dati personali, 2017

BUCCI F., *Algoritmi e diritti umani: qual è il punto d'incontro?* Ius in itinere, 2021-2022

CARBONE L., *Etica morale e deontologia nell'esercizio delle libere professioni*, La previdenza forense, 3/2020

CARR D., Giving Viewers What They Want, The New York Times, 24.2.2013.

CASCELLA L., *Artificial Intelligence Risks: Automation Bias*, MedProGroup

CONSIGLIO D'EUROPA, *Unboxing Artificial Intelligence: 10 steps to protect Human Rights*, 2019

COSTA G., *Security-by-design: strumenti e metodologie per lo sviluppo sicuro del software*, Cybersecurity360, 12.10.2018

90. Concretamente, tale fase potrebbe svolgersi utilizzando il Processo proposto da Z-Inspection® (in ZICARI R. E ALTRI, *How to Assess Trustworthy AI in Practice*, 28.6.2022), il quale si articola in tre momenti:

Set Up phase, incentrata sulla verifica della sussistenza delle precondizioni oggettive e soggettive e dell'ambiente di svolgimento del Test;

Assess phase, che identifica i problemi e poi li analizza e categorizza;

Resolve phase, nella quale si identificano invece le opportune soluzioni.

Tale processo potrebbe essere particolarmente idoneo allo scopo alla luce del fatto che ha un approccio olistico al tema in questione, analizzando l'intelligenza artificiale nel suo complesso quale sistema sociotecnico ed in modo dinamico (includendo peraltro un monitoraggio continuo dell'Intelligenza Artificiale), con l'intervento di una pluralità di competenze diversificate. Peraltro, l'analisi relativa agli scenari di utilizzo appare molto coerente con quanto poc'anzi riferito in merito al concetto di design delle funzionalità.

- CONTISSA G., GALLI F., GODANO F., SARTOR G., *Il regolamento europeo sull'intelligenza artificiale*, i-lex, 2021, fasc. 2
- DE CONNO A., *Criptovalute e impatto ambientale: verso una nuova regolamentazione europea. La UE monitora il consumo di energia generato dagli strumenti finanziari e mira alla sostenibilità*, Altalex, 30.08.2022
- DE MENECH C., *Intelligenza artificiale e autodeterminazione in materia sanitaria*, BioLaw Journal – Rivista di BioDiritto, n. 1/2022
- DI CARPEGNA BRIVIO E., *Il Reputation scoring e la quantificazione del valore sociale*, Federalismi.it, 13.7.2022
- DIMALTA D., *Intelligenza artificiale “ad alto rischio”, tutti gli ambiti su cui vigilare per la UE*, Agentadigitale.eu, 5.5.2021
- DONATI F., *Diritti fondamentali e algoritmi nella proposta di regolamento sull'intelligenza artificiale*, in *La rivoluzione dell'intelligenza artificiale: profili giuridici*, a cura di F. DONATI, A. PAJNO E A. PERRUCCI
- EBERS M. E ALTRI, *The European Commission's Proposal for an Artificial Intelligence Act—A Critical Assessment by Members of the Robotics and AI Law Society (RAILS)*, Multidisciplinary scientific Journal, 2021
- EUROPEAN AGENCY FOR SAFETY AND HEALTH AT WORK, *Impact of artificial intelligence on occupational safety and health*, 2021
- FALCHI M. C., *Intelligenza Artificiale: se l'algoritmo è discriminatorio*, Ius in itinere, 2020-2022
- FONDAZIONE ITALIANA DEL NOTARIATO, *Autonomia privata e fonti del diritto*, 2019
- FUTURE FOR LIFE INSTITUTE, *General Purpose AI and the AI Act*, 5.2022
- GRUPPO DI ESPERTI AD ALTO LIVELLO SULL'INTELLIGENZA ARTIFICIALE, *Orientamenti etici per un'IA affidabile*, Commissione europea, 8.4.2019
- HEINEMEYER M. *L'intelligenza artificiale negli attacchi informatici: tecniche offensive e strategie di difesa*, Cybersecurity360, 1.06.2020
- IASELLI M., *AI, il paradosso della legge come ostacolo o supporto: perché serve un nuovo approccio*, AgendaDigitale.eu, 4.3.2021
- LUCIANI A.M., *La nuova lex mercatoria*, in Osservatorio sulle fonti, n. 1/2021
- MARTORANA M. SAVELLA R., *Algoritmi razzisti, ecco le cause del problema e le soluzioni in campo*, AgendaDigitale.eu, 9.7.2020
- KANT I., *Fondazione della metafisica dei costumi*, 1785
- OFFICE OF THE UNITED NATIONS HIGH COMMISSIONER FOR HUMAN RIGHTS, *Principles and guidelines for a human rights approach to poverty reduction strategies*, HR/PUB/06/12
- PERON E., *Neuromarketing: come il cervello guida il consumatore*, thismarketerslife.it, 22.01.2016
- PURIFICATI J., *La Corte di Cassazione sulla trasparenza degli algoritmi e la validità del consenso*, e-lex, 7.6.2021
- RESTA F., *IA e neuroscienze, la nostra “privacy mentale” è a rischio? I paletti che servono*, AgendaDigitale.eu, 17.6.2021
- RICE K. E GIRVIN H., *Human Rights-Based Approach to Short-Term Study Abroad*, Springer Cham, 2021
- SANSAVINI S., *Il difficile dialogo fra le culture umanistica e scientifica*, Il carrobbio, 2014-2015
- SCACCIA G., *Proporzionalità e bilanciamento tra diritti nella giurisprudenza delle corti europee*, Rivista AIC, 26.9.2017, fasc. 3
- TONOLA C., *La differenza tra le pratiche commerciali ingannevoli e aggressive*, NJus, 1.10.2021
- TREZZA R., *Artificial Intelligence Act. Giudizio “ciclico” di meritevolezza e accountability intelligenti*, Universitalia, Roma, 2021
- VALENTINI A., *Fuggetta: “la cyber security non è un prodotto da acquistare”*, Cybersecurity360, 29.6.2022
- VON DER LEYEN U., *Un'Unione più ambiziosa. Il mio programma per l'Europa*, 16.7.2019
- ZAGRELBESKY G., *Valori e diritti nei conflitti della politica*, la Repubblica, 22.2.2008
- ZICARI R. E ALTRI, *How to Assess Trustworthy AI in Practice*, 28.6.2022