



*AFFIDABILITÀ,
DIRITTI FONDAMENTALI,
CENTRALITÀ DELL'ESSERE
UMANO:
UNA STRATEGIA EUROPEA
PER L'INTELLIGENZA
ARTIFICIALE*

MARIA ZANICHELLI

i-lex

i-lex. Scienze Giuridiche, Scienze Cognitive e Intelligenza Artificiale
Rivista quadrimestrale on-line: www.i-lex.it

Dicembre 2019
Fascicolo 12, 1-3
ISSN 1825-1927

AFFIDABILITÀ, DIRITTI FONDAMENTALI, CENTRALITÀ DELL'ESSERE UMANO: UNA STRATEGIA EUROPEA PER L'INTELLIGENZA ARTIFICIALE

MARIA ZANICHELLI*

Abstract: European Union's vision of a trustworthy AI is based on ethical principles and requirements that are ultimately connected to fundamental rights and human rights. The impact of AI systems on human dignity, human autonomy, fundamental rights and human rights is not a purely ethical question, but it is increasingly becoming the very issue at stake concerning the relationship between law and new technologies.

Parole chiave: IA, Trustworthiness, Fundamental Rights, Human Being

L'idea di copiare il funzionamento del cervello e della mente umana attraverso l'IA al di fuori del contesto delle relazioni, dell'ambiente e della cultura in cui ciascuno è immerso ha condotto a delle semplificazioni e a degli errori di valutazione fuorvianti, che la più recente IA sta oggi, tuttavia, riducendo ed eliminando.

Remo Bodei, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, 2019

1. “Accettare il cambiamento sulla base dei valori”: l’approccio dell’Unione europea all’IA

Tra le molteplici questioni che lo sviluppo delle tecnologie informatiche pone oggi all'attenzione del giurista, un tema-chiave è costituito dalle interazioni tra diritto e intelligenza artificiale. L'impatto dell'IA in ambito giuridico si dispiega sul piano culturale prima ancora che su quello applicativo: essa infatti, oltre che oggetto - necessariamente - di regolazione giuridica, è anche un fattore di trasformazione degli istituti e come tale incide sulle logiche proprie della giuridicità. Per-

tanto, la capacità di ‘governare’ l’intelligenza artificiale (in altri termini, di controllarne e disciplinarne gli sviluppi, di sfruttarne al massimo le risorse, neutralizzandone il più possibile gli effetti collaterali) è uno dei principali elementi critici su cui si misura l’adeguatezza dei sistemi giuridici rispetto alle sfide attuali¹.

Sotto questo profilo riveste un indubbio interesse quella che può essere definita la ‘visione europea’ dell’IA, risultante da alcuni documenti ufficiali recentemente emanati o promossi dall’Unione europea. Pur proponendosi di affrontare prevalentemente le implicazioni tecnologiche, economiche, sociali ed etiche della materia, tali documenti toccano in alcuni casi anche profili giuridici. Tra gli altri si possono ricordare:

- due risoluzioni del Parlamento europeo: *Civil Law Rules on Robotics* del 16 febbraio 2017 (2015/2103(INL)), e *A Comprehensive*

*Università degli Studi di Parma

¹ Tra gli studi più recenti: J. Turner, *Robot Rules. Regulating Artificial Intelligence*, Palgrave MacMillan, 2019; A. Bensamoun, G. Loiseau, *Droit de l’intelligence artificielle*, LGDJ, 2019; M. Corrales Compagnucci, M. Fenwick, N. Forgó (eds.), *Robotics, AI, and the Future of Law*, Springer, 2019; R. Susskind, *L’avvocato di domani. Il futuro della professione legale tra rivoluzione tecnologica e intelligenza artificiale*, Guerini e Associati, 2019; S. Lohsse, R. Schulze, D. Staudenmayer (eds.), *Liability for Artificial Intelligence and the Internet of Things*. Münster Colloquia on EU Law and the Digital Economy, Nomos, 2019; J. Nieva-Fenoll, *Intelligenza artificiale e processo*, Giappichelli, 2019; G. Teubner, *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, ESI, 2019; W. Barfield, U. Pagallo (eds.), *Research Handbook on the Law of Artificial Intelligence*, Edward Elgar Publishing, 2018; F. Pizzetti (a cura di), *Intelligenza artificiale, protezione dei dati personali e regolazione*, Giappichelli, 2018; K.D. Ashley, *Artificial Intelligence and Legal Analytics. New Tools for Law Practice in the Digital Age*, Cambridge University Press, 2017; P. Casanovas, U. Pagallo, M. Palmirani, G. Sartor, *AI Approaches to the Complexity of Legal Systems*, Springer, 2014; U. Pagallo, *The Laws of Robots. Crimes, Contracts, and Torts*, Springer, 2013. Dal 1992 l’editore Springer pubblica una rivista specialistica sul tema: *Artificial Intelligence and Law*. Tra le varie iniziative italiane vanno segnalati due recenti focus monografici su riviste giuridiche: *Soggettività, responsabilità, normatività 4.0. Profili filosofico-giuridici dell’intelligenza artificiale*, in *Rivista di filosofia del diritto - Journal of Legal Philosophy*, 1/2019, pp. 81-142 (contributi di E. Ancona, A. Simoncini, S. Suweis, A.C. Amato Mangiameli, G. De Anna) e *BioLaw Journal - Rivista di BioDiritto*, 1/2019, pp. 3-266 (contributi di A. D’Aloia, D. Amoroso, G. Tamburrini, L.B. Solum, A. Simoncini, A. Celotto, M. Fasan, F. Pacini S. Quattrocchio, F. Faini, E.A. Ferioli, C. Casonato, A. Vallini, A. Perin, E. Pulice, G. Beggio).

European Industrial Policy on Artificial Intelligence and Robotics del 12 febbraio 2019 (2018/2088(INI));

- tre comunicazioni della Commissione: *Artificial Intelligence for Europe* del 25 aprile 2018 (COM (2018) 237); *Coordinated Plan on Artificial Intelligence “Made in Europe”* del 7 dicembre 2018 (COM (2018) 795), sviluppato di concerto con gli Stati membri, anche con l'adesione di Norvegia e Svizzera; *Building Trust in Human-Centric Artificial Intelligence* dell'8 aprile 2019 (COM(2019) 168);

- gli orientamenti elaborati dal gruppo indipendente di 52 esperti istituito dalla Commissione nel 2018 (High-Level Expert Group on Artificial Intelligence, AI HLEG): *A Definition of AI. Main Capabilities and Scientific Disciplines*, dell'8 aprile 2019; *Ethics Guidelines for Trustworthy AI*, dell'8 aprile 2019; *Policy and Investment Recommendations for Trustworthy AI*, del 26 giugno 2019, queste ultime presentate nel corso della prima assemblea dell'Alleanza europea per l'IA, piattaforma multilaterale che coinvolge tutti i portatori di interessi per la condivisione di buone pratiche e attività correlate allo sviluppo dell'IA;

- e infine il *Report on Liability for Artificial Intelligence and Other Emerging Technologies* del 21 novembre 2019, elaborato dall'Expert Group on Liability and New Technologies - New Technologies Formation².

In un quadro internazionale in cui tutti i Paesi economicamente e tecnologicamente più avanzati riservano investimenti sempre più ingenti alla ricerca sull'IA e alle sue applicazioni, in quanto area strategica e altamente competitiva (si pensi per esempio alle politiche condotte in materia da Stati Uniti, Cina, Giappone, Canada, Israele), non stupisce che l'Unione europea abbia elaborato un proprio approccio al tema, che aspira a combinare l'implementazione dell'IA, quale fattore di innovazione, di profitto e di progresso a vari livelli, con un'attenzione programmatica ai rischi e alle criticità che essa implica, in particolare per il suo possibile impatto sulla dignità umana, sui diritti fondamentali, sull'ambiente, sulla democrazia e lo Stato di dirit-

² Qui saranno considerate in particolare le comunicazioni della Commissione e le linee guida etiche dell'AI HLEG. Le citazioni riportate successivamente nel testo sono tratte dalle traduzioni italiane ufficiali dei documenti, consultabili all'indirizzo <https://ec.europa.eu/digital-single-market/en/artificial-intelligence> (sito visitato il 30 novembre 2019).

to. Si va così profilando una narrazione prettamente europea dell'intelligenza artificiale, che si richiama espressamente ai valori e ai principi ispiratori sui quali l'Unione ha fondato storicamente la sua azione e in base ai quali tuttora delinea la propria identità sulla scena globale. Presentati come meritevoli di una specifica attenzione e tutela, a fronte di uno sviluppo tecnologico che sempre più appare metterli a repentaglio, quei principi e quei valori sono dunque esibiti dall'Unione come i tratti qualificanti di una prospettiva europea sull'intelligenza artificiale, i parametri valutativi per vagliarne l'accettabilità, per tracciarne i confini, per governarne gli esiti.

Per l'Europa si tratta, cioè, di potenziare gli stanziamenti mirati allo sviluppo dei sistemi di IA, in modo da non farsi distanziare in misura irrecuperabile dagli altri attori internazionali e da reggere una competizione che si farà sempre più serrata; e al tempo stesso di qualificarsi, proprio nella competizione con tali attori, per l'adozione di un approccio spiccatamente 'etico' a questo sviluppo, cioè un approccio che inquadri le tecnologie in un *frame* chiaro e riconoscibile di valori, finalità, parametri critici, opzioni culturali di riferimento (in termini di sostenibilità, di responsabilità, di diritti delle persone, di vantaggi per la società, di attendibilità e trasparenza dei processi decisionali, e così via).

Come ha sintetizzato la Commissione in due passaggi della comunicazione *Artificial Intelligence for Europe* del 2018,

“L’approccio sostenibile dell’UE alle tecnologie crea un vantaggio competitivo, accettando il cambiamento sulla base dei valori dell’Unione. Come ogni tecnologia trasformativa, alcune applicazioni dell’IA possono far sorgere nuovi interrogativi etici e giuridici, che riguardano per esempio la responsabilità o processi decisionali potenzialmente inficiati da condizionamenti. L’UE deve pertanto assicurare che l’IA sia sviluppata e applicata in un quadro adeguato che promuova l’innovazione e rispetti i valori dell’Unione e i diritti fondamentali, oltre ai principi etici come la responsabilità e la trasparenza. L’UE si trova in ottima posizione per condurre questo dibattito a livello mondiale. In questo senso l’UE può fare la differenza e propugnare un approccio all’IA che apporti benefici alle persone e alla società nell’insieme” [par. 1].

“Sono presenti tutti gli elementi principali affinché l’UE sia all’avanguardia nella rivoluzione dell’IA, secondo le proprie modalità e i propri valori. L’approccio all’IA descritto nel presente documento mostra la strada da seguire ed evidenzia la necessità di unire le forze a livello europeo, per assicurare che tutti i cittadini europei partecipino alla trasformazione digitale, che risorse adeguate siano dedicate all’IA e che i valori e i diritti fondamentali dell’Unione siano in primo piano nel contesto dell’IA” [Conclusioni].

La strategia europea per l’IA si caratterizza, dunque, per una declinazione ‘normativa’ della materia in questione: come ogni altro aspetto delle politiche dell’Unione, anche le conquiste tecnologiche, proprio per l’intensità del loro impatto effettivo e potenziale sulla società e sulle persone, vanno valutate e regolate alla luce di quanto stabilito dall’art. 2 TUE: *“L’Unione si fonda sui valori del rispetto della dignità umana, della libertà, della democrazia, dell’uguaglianza, dello Stato di diritto e del rispetto dei diritti umani, compresi i diritti delle persone appartenenti a minoranze. Questi valori sono comuni agli Stati membri in una società caratterizzata dal pluralismo, dalla non discriminazione, dalla tolleranza, dalla giustizia, dalla solidarietà e dalla parità tra donne e uomini”*. Non a caso tra gli obiettivi posti espressamente dalla Commissione vi è quello di *“assicurare un quadro etico e giuridico adeguato, basato sui valori dell’Unione e coerente con la Carta dei diritti fondamentali dell’UE”*.

Un primo dato significativo emergente dai documenti citati è il modo stesso in cui l’IA è definita. Sempre nella comunicazione *Artificial Intelligence for Europe* la Commissione ha affermato:

“‘Intelligenza artificiale’ (IA) indica sistemi che mostrano un comportamento intelligente analizzando il proprio ambiente e compiendo azioni, con un certo grado di autonomia, per raggiungere specifici obiettivi. I sistemi basati sull’IA possono consistere solo in software che agiscono nel mondo virtuale (ad esempio assistenti vocali, software per l’analisi delle immagini, motori di ricerca, sistemi di riconoscimento vocale e facciale), oppure incorporare l’IA in dispositivi hardware (per esempio in robot avanzati, auto a guida autonoma, droni o applicazioni dell’Internet delle cose)” [par. 1].

Successivamente il gruppo di esperti AI HLEG, allo scopo di chiarire alcuni profili dell'IA in quanto disciplina scientifica e in quanto tecnologia, ha elaborato una definizione ampliata e aggiornata, nel documento *A Definition of AI. Main Capabilities and Scientific Disciplines* del 2019:

“I sistemi di intelligenza artificiale (IA) sono sistemi software (ed eventualmente hardware) progettati dall'uomo [gli esseri umani progettano direttamente i sistemi di IA, ma possono anche utilizzare tecniche di IA per ottimizzarne la progettazione] che, dato un obiettivo complesso, agiscono nella dimensione fisica o digitale percependo il proprio ambiente attraverso l'acquisizione di dati, interpretando i dati strutturati o non strutturati raccolti, ragionando sulle conoscenze, o elaborando le informazioni derivate da questi dati e decidendo le migliori azioni da intraprendere per raggiungere l'obiettivo dato. I sistemi di IA possono usare regole simboliche o apprendere un modello numerico, e possono anche adattare il loro comportamento analizzando come l'ambiente è influenzato dalle loro azioni precedenti. Come disciplina scientifica, l'IA include diversi approcci e diverse tecniche, come l'apprendimento automatico (di cui l'apprendimento profondo e l'apprendimento per rinforzo sono esempi specifici), il ragionamento meccanico (che include la pianificazione, la programmazione, la rappresentazione delle conoscenze e il ragionamento, la ricerca e l'ottimizzazione), e la robotica (che include il controllo, la percezione, i sensori e gli attuatori e l'integrazione di tutte le altre tecniche nei sistemi ciberfisici)” [par. 4].

È evidente dalla combinazione di queste due definizioni la scelta di dare risalto all'elemento specifico che contraddistingue prioritariamente i sistemi intelligenti: il loro essere contemporaneamente prodotto di una progettazione dell'essere umano e capaci di prestazioni prettamente umane (nella terminologia delle definizioni riportate sopra: 'percepire' e 'analizzare' il proprio ambiente, 'agire', 'raggiungere obiettivi', 'acquisire dati', 'elaborare informazioni', 'decidere le migliori azioni da intraprendere', 'usare regole', 'apprendere', 'adattare il proprio comportamento', 'ragionare', 'pianificare', 'programmare', 'rappresentare conoscenze', etc.).

In quest'ottica, diviene essenziale stabilire quale grado di autonomia e soggettività sia riconoscibile a sistemi che presentano simulta-

neamente lo *status* di creazioni artificiali e quello di agenti, e conseguentemente adottare misure per preservare il ruolo e lo statuto in-fungibile dell'essere umano nell'interazione con tali sistemi. Sulle differenti prerogative proprie dei sistemi intelligenti e degli esseri umani, proprio il documento dell'AI HLEG appena citato, dedicato alla definizione di IA, contiene alcune importanti puntualizzazioni, per esempio con riferimento alla distinzione tra decidere l'obiettivo da raggiungere e individuare l'azione adeguata da intraprendere in vista di quell'obiettivo, e al differente grado di autonomia che le due operazioni rispettivamente implicano:

“In generale, il sistema di IA deve essere dotato di sensori adeguati che consentano al sistema di percepire nell'ambiente i dati pertinenti all'obiettivo assegnatogli dall'essere umano che lo ha progettato” [par. 1].

“Si osservi che il termine “decisione” va inteso in un'accezione ampia, come atto con cui si seleziona l'azione da intraprendere, e non sta necessariamente a indicare che i sistemi di IA sono completamente autonomi. Una decisione può essere anche la selezione di una raccomandazione da fornire a un essere umano, a cui spetta la decisione finale” [par. 1].

“Un sistema di IA generale è progettato per essere in grado di eseguire tutte o quasi le attività svolte dall'uomo. Un sistema di IA ristretta è invece in grado di svolgere soltanto uno o alcuni compiti specifici. I sistemi di IA attualmente in uso sono esempi di IA ristretta” [par. 3].

“Gli attuali sistemi di IA sono orientati agli obiettivi: ricevono dall'uomo i dettagli dell'obiettivo da raggiungere e si avvalgono di alcune tecniche per realizzarlo. Sono sistemi che non decidono autonomamente i propri obiettivi. Tuttavia, alcuni sistemi di IA (come quelli basati su determinate tecniche di apprendimento autonomo) godono di maggiore libertà nel decidere la via da percorrere per raggiungere l'obiettivo dato” [par. 3].

Già da queste prime indicazioni risulta evidente che, nel resoconto europeo, la definizione e la qualificazione del rapporto tra i sistemi intelligenti e gli esseri umani che li progettano, li realizzano, ne orien-

tano e sfruttano l'azione non è un aspetto marginale, relegabile sullo sfondo, ma è uno dei punti critici decisivi della questione: ne rappresenta la vera posta in gioco sul piano normativo³.

2. Una visione dell'IA centrata sull'umano

L'insistenza sull'interazione artificiale-umano è in effetti il *fil rouge* che connota i documenti dell'Unione in tema di IA, e può essere considerato il tratto distintivo della strategia europea in materia.

Non si può dimenticare, del resto, che il rapporto macchina-uomo è un tema classico di riflessione fin dagli albori delle tecnologie informatiche. Nel noto articolo *Computing Machinery and Intelligence*, pubblicato su *Mind* nel 1950, il matematico Alan M. Turing, ponendosi la domanda “*Can a machine think?*”, immaginava una macchina capace di tenere comportamenti non distinguibili da quelli di un essere umano, e inaugurava così l'idea di una simulazione artificiale dei processi cognitivi dell'uomo. Il modello di Turing fu poi proseguito da John Von Neumann nelle *Silliman Lectures* commissionategli dall'Università di Yale nel 1955, pubblicate postume nel 1958 con il titolo *The Computer and the Brain*, in cui poneva a confronto le capacità computazionali della macchina e quelle dell'uomo. A Turing si è poi contrapposto John Searle, con l'esperimento mentale della *Chinese Room* (nell'articolo *Minds, Brains and Programs*, pubblicato nel 1980 nella rivista *Behavioral and Brain Sciences*), con cui dimostrava che anche una macchina capace di superare il test di Turing non è

³ Il rapporto tra sistemi intelligenti ed esseri umani è uno dei nodi teorici e normativi centrali di questa materia. Si veda tra gli altri M. Hildebrandt, J. Gaakeer (eds.), *Human Law and Computer Law: Comparative Perspectives*, Springer, 2013; L. Floridi (ed.), *The Onlife Manifesto. Being Human in a Hyperconnected Era*, SpringerOpen, 2015; M. Negrotti (a cura di), *Uomini e macchine*, Armando, 2017; P. Benanti, *Le macchine sapienti. Intelligenze artificiali e decisioni umane*, Marietti, 2018; M. Tegmark, *Vita 3.0. Essere umani nell'era dell'intelligenza artificiale*, Raffaello Cortina Editore, 2018; P. Dumouchel, L. Damiano, *Vivere con i robot. Saggio sull'empatia artificiale*, Raffaello Cortina Editore, 2019; É. Sadin, *Critica della ragione artificiale. Una difesa dell'umanità*, LUISS University Press, 2019; E. Di Mauro, *La mente umana e la mente artificiale*, Asterios, 2019; P. Malavasi, *Educare robot? Pedagogia dell'intelligenza artificiale*, Vita & Pensiero, 2019; R. Bodei, *Dominio e sottomissione. Schiavi, animali, macchine, Intelligenza Artificiale*, il Mulino, 2019.

comunque in grado di 'capire' alla pari di un cervello umano, mancando di una mente e di una coscienza (a differenza di quanto ipotizzato nel paradigma dell'"*intelligenza artificiale forte*"). La dialettica *robot - human being* è al centro anche della letteratura fantascientifica, un vero *leitmotiv* per esempio nelle opere di Isaac Asimov, che nel racconto *Runaround* del 1942 la codificò nelle tre celebri "*leggi della robotica*":

“1. *A robot may not injure a human being or, through inaction, allow a human being to come to harm.*

2. *A robot must obey orders given it by human beings except where such orders would conflict with the First Law.*

3. *A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.”*

Questo retroterra di interrogativi, ipotesi scientifiche e modellizzazioni teoriche risalenti al secolo scorso è ben presente nei recenti tentativi dell'Unione europea di inquadrare normativamente l'impatto dell'intelligenza artificiale sulla società. E proprio il riferimento letterario appena ricordato è citato, non a caso, nella risoluzione del Parlamento del 2017 *Norme di diritto civile sulla robotica*, in cui la sezione dedicata ai principi generali si apre con l'affermazione che "*le leggi di Asimov devono essere considerate come rivolte ai progettisti, ai fabbricanti e agli utilizzatori di robot, compresi i robot con capacità di autonomia e di autoapprendimento integrate*" (considerando T).

Se la riproducibilità artificiale di alcune competenze proprie dell'essere umano mette in questione lo statuto dell'umanità, si rende tanto più necessario tutelare la specificità e l'incommensurabilità di tale statuto, cioè preservare il confine tra *status* umano e *status* artificiale. Il nodo del rapporto tra intelligenza artificiale e intelligenza umana (e conseguentemente del rapporto tra sistemi intelligenti e dignità, libertà, autonomia, diritti dell'essere umano) resta dunque in primo piano, anche senza dare credito a chi prospetta il rischio della 'singolarità', cioè l'avvento di forme di intelligenza artificiale più potenti di quella umana, che finirebbero per decretare il superamento della mente umana che le ha ideate e costruite⁴.

⁴ Cfr. R. Kurzweil, *La singolarità è vicina*, Apogeo, 2010; J. Barrat, *La nostra invenzione finale. L'intelligenza artificiale e la fine dell'età dell'uomo*, Nutrimenti, 2019.

Nella visione risultante dalle dichiarazioni europee, è proprio lo sviluppo di sistemi ‘intelligenti’, progettati per riprodurre abilità e prestazioni tipicamente umane, che rende necessario rimarcare la radicale alterità uomo-macchina e difendere il primato e la centralità dell’umanità. Questa è una pre-condizione perché i vantaggi di vario tipo offerti o prospettati dai sistemi di IA possano conquistare la fiducia delle società, delle istituzioni e delle persone. Non sarebbe sufficiente, viceversa, una sorta di ‘legittimazione dell’*output*’, che avallesse come accettabile di per sé qualunque tecnologia, esclusivamente in termini di efficacia dei risultati, abbattimento dei costi, razionalizzazione dei processi, e così via.

D’altra parte, preservare il primato e la centralità dell’umano in rapporto agli sviluppi tecnologici significa, nella prospettiva adottata dall’Unione europea, non soltanto tracciare confini certi e rassicuranti tra prerogative dell’uomo e prerogative dei sistemi intelligenti, ma anche eleggere l’essere umano a destinatario dei vantaggi offerti dall’IA; e dunque selezionare in modo chiaro, nell’ambito delle svariate applicazioni che la tecnologia potenzialmente rende disponibili, soltanto quelle che sono compatibili con questo fine, escludendo quelle che invece danneggerebbero l’umanità. “*Insieme, potremo mettere la forza dell’IA al servizio del progresso umano*”: è l’auspicio con cui la Commissione conclude la comunicazione *Artificial Intelligence for Europe*, un documento in cui la centralità dell’umano è continuamente ribadita tramite affermazioni che esprimono una concezione dell’IA decisamente orientata sull’essere umano. Per esempio, a proposito del sostegno che la Commissione si impegna a dare alla ricerca [par. 3.1]:

“Il principio guida di tutto il sostegno alla ricerca collegata all’IA sarà lo sviluppo dell’IA responsabile’ incentrata sugli esseri umani [...]”

Oppure a proposito dell’impatto sull’organizzazione del lavoro [par. 3.2]:

“Queste tecnologie possono rendere la vita dei lavoratori più facile. Possono ad esempio aiutarli nell’esecuzione di compiti ripetitivi, faticosi e perfino pericolosi (ad esempio nella pulizia di elementi il cui accesso è difficile o pericoloso quali le condotte industriali).

Possono anche facilitare la formulazione di riassunti a partire da grandi volumi di dati, fornire informazioni più accurate e suggerire decisioni, ad esempio mediante il ricorso all'IA per aiutare i medici nell'attività diagnostica. Esse servono in ultima analisi a migliorare le abilità degli esseri umani.”

[...]

“Per gestire la trasformazione dell'IA, i lavoratori il cui impiego sta cambiando o può scomparire a causa dell'automazione devono avere tutte le opportunità di acquisire le competenze e le conoscenze necessarie per padroneggiare la nuova tecnologia e ricevere sostegno durante la transizione nel mercato del lavoro. Questo atteggiamento proattivo e incentrato sull'investire nelle persone è il fondamento di un approccio all'IA inclusivo e antropocentrico, che richiederà un investimento notevole. [...] La Commissione continuerà [...] a sostenere la ricerca nell'interazione e nella cooperazione tra gli esseri umani e l'IA.”

[...]

“Affinché le persone accettino l'IA sarà decisivo garantire che i lavoratori abbiano la possibilità di adattarsi e di accedere alle nuove opportunità. Come per qualsiasi altra tecnologia, l'IA non viene semplicemente imposta alla società. È compito dei governi, dialogando con i partner sociali e la società civile, guidare collettivamente il processo in modo da assicurare che i suoi vantaggi siano ampiamente condivisi, che tutti i cittadini siano adeguatamente preparati per avvalersi appieno di questa tecnologia e che abbia luogo una riflessione più ampia su cambiamenti sociali che potrebbero avere portata ancora maggiore.”

O ancora a proposito dell'assicurazione di un adeguato quadro etico e giuridico [par. 3.3]:

“Per rafforzare ulteriormente la fiducia è anche necessario che le persone comprendano come funziona la tecnologia, il che spiega l'importanza della ricerca nella spiegabilità dei sistemi di IA. In effetti, allo scopo di aumentare la trasparenza e minimizzare il rischio di condizionamenti o errori, i sistemi di IA dovrebbero essere sviluppati in modo da permettere agli esseri umani di comprendere le loro azioni e la logica sottostante.”

[...]

“L'utilizzo su larga scala di strumenti capaci di usare l'IA nelle transazioni tra imprese e consumatori deve essere equo, trasparente

e conforme alla legislazione sulla tutela dei consumatori. I consumatori devono ricevere informazioni chiare sull'utilizzo, le caratteristiche e le proprietà dei prodotti capaci di usare l'IA. I singoli individui dovrebbero poter controllare i dati generati mediante l'utilizzo di questi strumenti e dovrebbero sapere se stanno comunicando con una macchina o un altro essere umano. In particolare, durante l'interazione con un sistema automatizzato, sarebbe necessario definire in qual momento gli utenti dovrebbero essere informati su come contattare un essere umano e come garantire che le decisioni di un sistema possano essere controllate o corrette.”

Su questi intenti la Commissione ha posto ulteriore enfasi con una successiva comunicazione del 2019 dal titolo eloquente: *Building Trust in Human-Centric Artificial Intelligence*, che promuove espressamente una concezione dell'intelligenza artificiale che meriti fiducia in quanto “*centrata sull'umano*” (così sembra corretto rendere in italiano l'espressione *human-centric, centré sur l'humain, menschenzentrierte, centrado en la persona* adottata rispettivamente nelle versioni inglese, francese, tedesca, spagnola del documento)⁵. Due passaggi del testo in particolare, sono paradigmatici rispetto a questa finalità:

“la fiducia è una condizione indispensabile per assicurare un approccio antropocentrico all'IA: l'intelligenza artificiale non è fine a sé stessa, ma è uno strumento a servizio delle persone che ha come fine ultimo quello di migliorare il benessere degli esseri umani. Per questo occorre garantire l'affidabilità dell'IA. I valori su cui si basano le nostre società devono essere pienamente integrati nelle modalità di sviluppo dell'IA. [...] L'UE dispone di un solido quadro normativo destinato a diventare lo standard internazionale per un'IA antropocentrica. [...] L'IA pone tuttavia nuove sfide perché

⁵ Non del tutto corrispondente agli intenti del documento in questione è, invece, la traduzione ‘antropocentrico’ scelta nella versione italiana ufficiale della comunicazione, sia nel titolo sia nelle varie occorrenze della nozione all'interno del testo (in divergenza dalle soluzioni adottate nelle altre lingue citate, nessuna delle quali fa riferimento all'antropocentrismo). L'antropocentrismo rappresenta, come è noto, un'assunzione teorica ben precisa e filosoficamente connotata; mentre la prospettiva che l'Unione assume programmaticamente per qualificare l'approccio europeo all'AI è “*centrata sull'umano*” in un senso più ampio e generale, cioè in quanto si propone di tutelare *in primis* l'umanità, ossia quello *status* di esseri umani che rimarrà sempre ontologicamente incommensurabile rispetto alle macchine, per quanto intelligenti.

consente alle macchine di "imparare", prendere decisioni ed eseguirle senza l'intervento umano. In breve tempo questa modalità di funzionamento diventerà la norma per molti tipi di beni e servizi, dagli smartphone alle automobili automatizzate, ai robot e alle applicazioni online. Le decisioni prese dagli algoritmi potrebbero però basarsi su dati incompleti e quindi non affidabili, manomessi a seguito di attacchi informatici, inficiati da condizionamenti o semplicemente errati. Applicare acriticamente la tecnologia man mano che questa viene sviluppata porterebbe pertanto a risultati problematici e a una certa riluttanza da parte dei cittadini ad accettarla o a utilizzarla. La tecnologia dell'IA dovrebbe invece essere sviluppata in modo da porre al centro l'essere umano e permetterle di conquistare la fiducia del pubblico [...]. Le applicazioni di IA dovrebbero dare potere alle persone e rispettarne diritti fondamentali; dovrebbero puntare a rafforzare le capacità dei cittadini, non a sostituirsi a loro, e consentire l'accesso anche alle persone con disabilità" [par. 2].

"Puntando a un'IA antropocentrica e basata sulla fiducia salvaguardiamo il rispetto dei nostri valori sociali fondamentali e consentiamo all'Europa e alla sua industria di distinguersi chiaramente in quanto leader di un'IA all'avanguardia affidabile in tutto il mondo" [Conclusioni].

L'approccio "centrato sull'umano" riemerge ripetutamente anche nelle linee guida etiche elaborate dall'AI HLEG (*Ethics Guidelines for Trustworthy AI*, d'ora in poi "linee guida"), rese pubbliche sempre l'8 aprile 2019, contestualmente alla comunicazione appena ricordata, e basate sul lavoro svolto dal Gruppo europeo sull'etica nelle scienze e nelle nuove tecnologie e dall'Agenzia dell'Unione europea per i diritti fondamentali. A proposito degli interrogativi etici suscitati dall'interazione uomo-macchina si possono citare due passaggi emblematici, uno riguardante la comunicazione tra sistemi intelligenti e utenti umani, in quanto specificazione del principio di trasparenza, e l'altro riguardante le preoccupazioni suscitate dai "sistemi di IA nascosti":

"(78) I sistemi di IA non devono presentarsi agli utenti come esseri umani e gli esseri umani hanno il diritto di essere a conoscenza del fatto che stanno interagendo con un sistema di IA. Ciò implica che i

sistemi di IA debbano essere identificabili come tali. Inoltre, per garantire il rispetto dei diritti fondamentali, dovrebbe essere prevista ove necessario la possibilità di preferire l'interazione umana a quella con il sistema di IA. Oltre a ciò, dovrebbero essere comunicate agli operatori del settore dell'IA o agli utenti finali le capacità e le limitazioni del sistema in maniera consona al caso d'uso in questione. Ciò potrebbe comprendere la comunicazione del livello di precisione del sistema di IA e dei suoi limiti.

(131) Gli esseri umani dovrebbero sempre sapere se stanno interagendo direttamente con un altro essere umano o con una macchina e spetta agli operatori del settore dell'IA la responsabilità di comunicarlo in modo affidabile. Gli operatori del settore dell'IA dovrebbero pertanto garantire (ad esempio tramite clausole di esclusione della responsabilità chiare e trasparenti) che le persone siano consapevoli del fatto che stanno interagendo con un sistema di IA o che possano chiedere informazioni in merito e approvare tale interazione. Va notato che esistono casi limite che complicano la questione (ad esempio una voce umana filtrata da un sistema di IA). Occorre tenere presente che la confusione tra esseri umani e macchine potrebbe avere molteplici conseguenze quali attaccamento, influenza o svilimento dell'essere umano. Lo sviluppo di androidi dovrebbe pertanto essere oggetto di un'attenta valutazione etica.”

Nelle linee guida, inoltre, è sottolineata più volte la centralità dell'umano come parametro normativo di valutazione dell'accettabilità degli sviluppi tecnologici. A titolo di esempio:

“(10) A tal fine si devono realizzare sistemi di IA antropocentrici, tenendo fede all'impegno di metterli sempre al servizio dell'umanità e del bene comune, con l'obiettivo di migliorare il benessere e la libertà degli esseri umani.

(38) I trattati dell'UE e la Carta dell'UE sanciscono una serie di diritti fondamentali che gli Stati membri e le istituzioni dell'UE sono giuridicamente tenuti a rispettare quando attuano il diritto unionale. Tali diritti sono descritti nella Carta dei diritti fondamentali dell'Unione europea con riferimento alla dignità, alle libertà, all'uguaglianza, alla solidarietà, ai diritti dei cittadini e alla giustizia. Il fondamento che accomuna questi diritti può essere inteso come radicato nel rispetto della dignità umana, riflettendo così

quello che definiamo un “approccio antropocentrico” in cui l'essere umano gode di uno status morale unico e inalienabile di primato in campo civile, politico, economico e sociale (Occorre notare che l'impegno per un'IA antropocentrica e il suo nesso con i diritti fondamentali necessita di basi collettive e costituzionali in cui la libertà individuale e il rispetto della dignità umana siano praticabili e significative, anziché implicare un indebito tornaconto individualistico dell'essere umano).

(50) I diritti fondamentali su cui si fonda l'Unione europea sono volti a garantire il rispetto della libertà e dell'autonomia degli esseri umani. Gli esseri umani che interagiscono con i sistemi di IA devono poter mantenere la propria piena ed effettiva autodeterminazione e devono poter essere partecipi del processo democratico. I sistemi di IA non devono subordinare, costringere, ingannare, manipolare, condizionare o aggregare in modo ingiustificato gli esseri umani. Al contrario, devono essere progettati per aumentare, integrare e potenziare le abilità cognitive, sociali e culturali umane. La distribuzione delle funzioni tra esseri umani e sistemi di IA dovrebbe seguire i principi di progettazione antropocentrica e lasciare ampie opportunità di scelta all'essere umano. Ciò significa garantire la sorveglianza e il controllo dei processi operativi nei sistemi di IA da parte di esseri umani. I sistemi di IA possono anche cambiare radicalmente il mondo del lavoro, sostenendo l'uomo nell'ambiente lavorativo al fine di generare un lavoro significativo.

(153) L'approccio antropocentrico all'IA è volto a garantire che i valori umani rivestano un ruolo centrale nelle modalità di sviluppo, distribuzione, utilizzo e monitoraggio dei sistemi di IA, garantendo il rispetto dei diritti fondamentali, tra cui quelli sanciti nei trattati dell'Unione europea e nella Carta dei diritti fondamentali dell'Unione europea, accomunati dal riferimento a un fondamento condiviso radicato nel rispetto della dignità umana, nei quali l'essere umano gode di uno status morale unico e inalienabile. Ciò implica anche il rispetto dell'ambiente naturale e di altri esseri viventi che fanno parte dell'ecosistema umano e un approccio sostenibile che consenta alle generazioni future di prosperare.”

Come si vede, la visione di un'IA “centrata sull'umano” auspicata dall'AI HLEG si ispira dichiaratamente a una precisa grammatica normativa: la Carta dei diritti fondamentali dell'Unione, con i sei

principi su cui è imperniata (dignità, libertà, uguaglianza, solidarietà, cittadinanza, giustizia). È dunque individuabile un ulteriore requisito che contribuisce a identificare la prospettiva europea sull'intelligenza artificiale: il richiamo ai 'diritti fondamentali', nella loro versione tipicamente europea, cioè basata sul primato della dignità umana, sull'idea di un radicamento dei diritti nelle istituzioni, e su una concezione non individualistica delle libertà dei singoli.

È da notare che l'incidenza dell'IA sui diritti (con specifico riferimento ai 'diritti umani') è un tema al centro dell'attenzione anche del Consiglio d'Europa, organismo distinto - come è noto - dall'Unione europea, che va moltiplicando le iniziative istituzionali su questo tema, sulla base degli stessi valori e principi di riferimento condivisi dall'Unione. Per esempio, nel febbraio 2019 si è tenuta a Helsinki la Conferenza di alto livello *Governing the Game Changer - Impacts of Artificial Intelligence Development on Human Rights, Democracy and the Rule of Law*. Inoltre nel maggio 2019 la Commissaria per i diritti umani, Dunja Mijatović, ha pubblicato la raccomandazione *Unboxing Artificial Intelligence: 10 Steps to Protect Human Rights*, indicante una lista di azioni da intraprendere per massimizzare il potenziale dei sistemi di IA e prevenire o attenuare il loro impatto negativo sulla vita e sui diritti delle persone. E recentemente, dal 18 al 20 novembre 2019, si è tenuta a Strasburgo la prima riunione del Comitato ad hoc sull'intelligenza artificiale (CAHAI), organo intergovernativo incaricato di effettuare uno studio di fattibilità sui principali obiettivi concernenti l'IA (stabilire una definizione comune e un quadro giuridico vincolante, e realizzare una mappatura dei rischi e delle opportunità che l'IA può generare, in particolare il suo impatto sui diritti umani, lo Stato di diritto e la democrazia)⁶.

Il rischio che le nuove tecnologie incidano negativamente sui diritti è oggi ovviamente un tema cruciale⁷. Tanto più significativo appare

⁶ Sempre nell'ambito del Consiglio d'Europa va ricordata anche la *Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari* emanata il 3 dicembre 2018 dalla Commissione Europea per l'Efficienza della Giustizia (CEPEJ(2018)14).

⁷ Sul tema cfr. B. Wagner, M.C. Kettemann, K. Vieth (eds.), *Research Handbook on Human Rights and Digital Technology. Global Politics, Law and International Relations*, Edward Elgar Publishing, 2019; D. Ruggiu, *Human Rights and Emerging Technologies. Analysis and Perspectives in Europe*, Pan Stanford Publishing, 2018.

lo sforzo dell'Unione europea di farsene carico, con riferimento specifico all'IA, e di declinare tale attenzione con un proprio linguaggio riconoscibile, dando così un apporto che va a integrare l'azione svolta da enti e organismi istituzionalmente impegnati nella promozione e nella tutela dei diritti umani.

3. IA, diritti fondamentali e diritti umani

Alla luce di quanto osservato finora, si può ipotizzare che, se è individuabile una via specificamente europea nella regolazione etico-giuridica dell'intelligenza artificiale, essa consista proprio nell'approccio *rights-based* adottato nelle linee guida dell'AI HLEG. A riguardo in esse si afferma programmaticamente:

“(21) I presenti orientamenti definiscono un quadro di riferimento per ottenere un'IA affidabile basata sui diritti fondamentali sanciti dalla Carta dei diritti fondamentali dell'Unione europea e dal pertinente diritto internazionale in materia di diritti umani.

(37) Crediamo in un approccio all'etica dell'IA basato sui diritti fondamentali sanciti dai trattati UE, dalla Carta dei diritti fondamentali dell'Unione europea e dal diritto internazionale in materia di diritti umani. Il rispetto dei diritti fondamentali, nel quadro della democrazia e dello Stato di diritto, costituisce la base più incoraggiante per individuare principi e valori etici astratti che possono essere resi operativi nei sistemi di IA.”

I diritti, in questa loro duplice declinazione (Carta UE dei diritti fondamentali e norme internazionali sui diritti umani), sono dunque il faro alla luce del quale l'Europa persegue l'obiettivo di “*un'IA affidabile*”, esemplificando tanto le “*opportunità vantaggiose*” da cogliere tramite l'IA quanto le “*serie preoccupazioni che i sistemi di IA suscitano e che dovrebbero essere attentamente considerate*”.

Se però si va in cerca di un'opzione teorica alla base delle linee guida europee, il dato che colpisce è la scelta fatta dal gruppo di esperti di considerare espressamente i diritti fondamentali come istanze appartenenti all'ambito etico più che a quello giuridico:

“(39) [...] Intesi come diritti giuridicamente applicabili, i diritti fondamentali rientrano [...] nella prima componente di un’IA affidabile (legalità dell’IA), che salvaguarda il rispetto della legge. Intesi come diritti di ciascuno, radicati nello status morale intrinseco degli esseri umani, essi sono anche alla base della seconda componente dell’IA affidabile (eticità dell’IA), che si occupa di norme etiche che non sono necessariamente vincolanti dal punto di vista giuridico ma cruciali per garantire l’affidabilità. Poiché il presente documento non intende offrire indicazioni sulla prima componente, ai fini dei presenti orientamenti non vincolanti, i riferimenti ai diritti fondamentali riflettono la seconda componente.”

Per inquadrare adeguatamente questo aspetto, è utile dare uno sguardo all’architettura complessiva e alle finalità generali di questo documento. Anzitutto, si tratta di orientamenti il cui obiettivo non è sostituire azioni politiche o regolative, bensì porsi come punto di partenza per una discussione sul tema, aperta a revisioni e aggiornamenti. Il diritto non è certo ignorato: l’affidabilità dell’IA è in effetti articolata in termini di ‘legalità’, ‘eticità’ e ‘robustezza’ (intese rispettivamente come ottemperanza a leggi e regolamenti, adesione a principi e valori etici, capacità di evitare che i sistemi provochino danni); tale definizione interdisciplinare presuppone dunque la connessione, assunta oggi abitualmente nell’approccio all’IA, tra la dimensione giuridica, quella etica e quella tecnologica⁸. Tuttavia il documento sottolinea più volte la scelta di limitarsi al secondo e al terzo profilo, mentre la dimensione giuridica è volutamente lasciata sullo sfondo (almeno negli intenti): l’oggetto principale sono le implicazioni etiche nella loro autonomia e distinzione da quelle giuridiche, pur nel riconoscimento di occasionali consonanze tra i due ambiti⁹. Nel docu-

⁸ Un esempio recente è il *focus* interdisciplinare sull’intelligenza artificiale contenuto in *BioLaw Journal - Rivista di BioDiritto*, vol. 3, 2019, pp. 179-254 (presentazione di L. Violante e A. Pajno; contributi di S. Quintarelli, F. Corea, F. Fossa, A. Loreggia, S. Sapienza sui profili etici; di A. Pajno, M. Bassini, G. De Gregorio, M. Macchia, F.P. Patti, O. Pollicino, S. Quattrocolo, D. Simeoli, P. Sirena sui profili giuridici; di M.C. Carrozza, C. Oddo, S. Orvieto, A. di Minin, G. Montemagni sui profili tecnologici).

⁹ Già nella nota 1 della sintesi iniziale si precisa: *“Tutte le affermazioni normative del presente documento intendono rispecchiare le indicazioni per il raggiungimento della seconda e terza componente di un’IA affidabile (eticità e robustezza). Pertanto, tali affermazioni non intendono fornire né pareri giuridici né indicazioni*

mento si dà per acquisito che i sistemi di IA non operano in un *vacuum* normativo, e anzi si ribadisce come ovvio il dovere di ognuno di rispettare le leggi vigenti sull'IA, oltre che i diritti e gli obblighi giuridici inderogabili in materia. Il *focus* tuttavia resta il profilo etico, non quello giuridico:

“(25) Nulla di quanto contenuto nel presente documento deve essere inteso o interpretato come consulenza legale o indicazioni su come conformarsi alle norme e alle disposizioni giuridiche vigenti. Nulla di quanto contenuto nel presente documento crea diritti giuridici o impone obblighi giuridici nei confronti di terzi [...]”.

“(26) Per ottenere un'IA affidabile non è sufficiente il rispetto della legge, che è solo una delle tre componenti. Il diritto non è sempre al passo con gli sviluppi tecnologici, e a volte non lo è nemmeno con le norme etiche o semplicemente non è adatto ad affrontare determinate questioni. Affinché i sistemi di IA siano affidabili, essi dovrebbero quindi essere anche etici garantendo la compatibilità con le norme etiche.”

È dunque sul piano etico che sono individuati i principi da osservare nello sviluppo, nella distribuzione e nell'utilizzo dei sistemi di IA: rispetto dell'autonomia umana, prevenzione dei danni, equità, esplicabilità (capitolo I, “Basi di un'IA affidabile”). L'obiettivo è “*un'IA affidabile basata sui diritti fondamentali*”, e i relativi principi sono intesi come imperativi etici “*radicati nei diritti fondamentali*”. Pur non negando che tali imperativi possano essere già presenti nel diritto vigente, e dunque in disposizioni giuridiche vincolanti (e come tali rientrano anche nella componente di legalità dell'IA), le linee guida rivendicano decisamente la loro ascrizione alla sfera etica: “*anche se molti obblighi giuridici riflettono principi etici, l'adesione ai principi etici va oltre il rispetto formale del diritto vigente*” (punto 49). Si raccomanda inoltre un'attenzione speciale per i soggetti più vulnerabili (bambini, persone con disabilità, altri gruppi storicamente svantaggiati o a rischio di esclusione, situazioni caratterizzate da asimmetrie di potere o di informazione, ad esempio tra datori di lavoro e lavoratori o tra imprese e consumatori). Infine, si auspica l'adozione di provve-

su come ottemperare alle leggi vigenti, sebbene vada riconosciuto che molte di tali affermazioni sono in qualche misura già presenti nelle leggi vigenti”.

dimenti adeguati ad attenuare i possibili effetti negativi dell'IA “*sulla democrazia, sullo Stato di diritto, sulla giustizia distributiva o sulla stessa mente umana*”.

I principi etici hanno poi una traduzione concreta in sette requisiti che i sistemi intelligenti devono soddisfare per garantire un'IA affidabile: intervento e sorveglianza umani; robustezza tecnica e sicurezza; riservatezza e *governance* dei dati; trasparenza; diversità, non discriminazione ed equità; benessere sociale e ambientale; *accountability* (capitolo II, “*Realizzare un'IA affidabile*”). Qui emerge di nuovo il nesso con i diritti: si afferma infatti che l'elenco dei requisiti fa riferimento ai principi e ai diritti enunciati nella Carta dei diritti fondamentali, di cui intende riprodurre l'ordine di successione, senza alcuna valenza gerarchica. Allo scopo di rendere operativi i sette requisiti, è proposta una lista di controllo per valutare concretamente l'affidabilità di un sistema di IA (capitolo III, “*Valutazione di un'IA affidabile*”). La lista non è esaustiva né definitiva (si tratta infatti di una fase pilota, al termine della quale nel 2020 sarà presentata alla Commissione una versione aggiornata sulla base dei riscontri pervenuti da tutti gli *stakeholder* che hanno partecipato alla sperimentazione). Anche a proposito della lista le linee guida dichiarano di volersi porre ad un livello distinto da quello giuridico: si specifica infatti che l'osservanza della lista non è una prova di conformità giuridica, e che essa mira a concretizzare i requisiti di eticità e robustezza di un sistema di IA, mentre non dà indicazioni su come garantire la conformità dell'IA alle leggi vigenti (punti 112 e 119). E a proposito dell'eventuale violazione di una delle componenti dell'IA affidabile, si afferma che l'ottemperanza alle norme giuridiche in materia non è risolutiva: “*resta il fatto che tali norme potrebbero non risolvere le preoccupazioni etiche che possono insorgere*” (punto 129)¹⁰.

¹⁰ Che l'intelligenza artificiale sollevi interrogativi e problemi di tenore etico e richieda un approccio anche eticamente orientato è oggi un dato acquisito. Non a caso nell'ambito delle etiche applicate l'etica dell'intelligenza artificiale va assumendo un proprio statuto autonomo (*IA Ethics, Roboethics, Machine Ethics* si definiscono come branche specifiche della *Computer Ethics* e della *Digital Ethics*). Cfr. tra gli altri G. Tamburrini, *Etica delle macchine*, Carocci, *forthcoming*; M. Coeckelbergh, *AI Ethics*, The MIT Press, *forthcoming*; N. Weidenfeld, J. Nida-Rümelin, *Umanesimo digitale. Un'etica per l'epoca dell'Intelligenza Artificiale*, FrancoAngeli, 2019; P. Lin, K. Abney, G.A. Bekey (eds), *Robot Ethics. The Ethical and Social Implications of Robotics*, The MIT Press, 2011.

In realtà però tra gli orientamenti etici e tecnologici sull'intelligenza artificiale, che *in primis* il documento si propone di fornire, e il livello della regolazione giuridica, che invece non è specificamente affrontato, vi è un'area di intersezione evidente e paradigmatica, costituita proprio dai diritti:

“(40) Nell’esauriente corpus di diritti indivisibili previsti dal diritto internazionale in materia di diritti umani, dai trattati UE e dalla Carta dei diritti fondamentali dell’Unione europea, le seguenti famiglie di diritti fondamentali sono particolarmente pertinenti per quanto riguarda i sistemi di IA. Molti di questi diritti, in determinate circostanze, sono giuridicamente applicabili nell’UE, pertanto il rispetto dei loro termini è giuridicamente vincolante. Ma anche una volta rispettati i diritti fondamentali giuridicamente applicabili, la riflessione etica può aiutarci a comprendere come lo sviluppo, la distribuzione e l’utilizzo dell’IA possano implicare i diritti fondamentali e i valori soggiacenti, e può contribuire a fornire un orientamento più dettagliato nel percorso di ricerca di ciò che dobbiamo fare piuttosto che ciò che (attualmente) possiamo fare con la tecnologia.”

Ed è proprio questa la sezione del documento che offre ampia materia di riflessione ai giuristi, indicando le aree di una possibile incidenza dell'IA sui diritti, e dunque i parametri normativi da rispettare nel suo sviluppo e nelle sue applicazioni (punti 41-45): rispetto della dignità umana, libertà individuale, rispetto della democrazia, della giustizia e dello Stato di diritto, uguaglianza, non discriminazione e solidarietà (compresi i diritti delle persone a rischio di esclusione), diritti dei cittadini. È evidente la concezione tipicamente 'europea' retrostante a questa lista, che declina i diritti intorno ad alcuni nuclei normativamente forti, che sono gli stessi su cui è imperniata la Carta dei diritti fondamentali, ma rapportati ai rischi di violazioni nuove e inedite prodotte dall'IA. Per esempio, stando al dettato delle linee guida, rispettare la dignità, in riferimento all'IA, significa che il “*valore intrinseco*” di ogni essere umano non può essere svilito dalle nuove tecnologie; rispettare la libertà comporta “*un impegno affinché gli individui possano esercitare un controllo addirittura maggiore sulla propria vita*”; rispettare la democrazia significa che i sistemi di IA non devono compromettere “*i processi democratici, la decisione*

umana o i sistemi di voto democratico”, né “*gli impegni di base su cui si fonda lo Stato di diritto*”¹¹.

Assumendo dunque che la visione europea dell’IA affidabile ha il suo fulcro nei diritti fondamentali e nei diritti umani, e che proprio questa interazione IA - diritti è destinata a diventare un fondamentale nodo di riflessione in tema di rapporti tra giuridicità e nuove tecnologie, una considerazione conclusiva può essere fatta sulla definizione del soggetto di tali diritti. Consideriamo solo tre esempi tratti in questo caso dal testo inglese delle linee guida.

A proposito del rispetto dell’autonomia umana, si afferma che va garantita un’effettiva autodeterminazione degli “*humans interacting with AI systems*”, e che nella ripartizione di funzioni “*between humans and AI systems*” va riservato spazio adeguato alla scelta umana (punto 50).

Ai fini dell’affidabilità dell’IA, il principio di trasparenza nelle interazioni uomo - IA comporta che “*AI systems should not represent themselves as humans to users*”, e che “*humans have the right to be informed that they are interacting with an AI system*” (punto 78).

Infine, per contrastare la possibilità di “sistemi di IA nascosti”, si afferma che “*human beings should always know if they are directly interacting with another human being or a machine*”; si sottolineano inoltre i rischi di “*confusion between humans and machines*” e le cautele necessarie nel caso di “*human-like robots*” (punto 131).

Come si può notare, proprio nell’era tecnologica, quando affiorano ipotesi di riconoscere ai sistemi di IA soggettività e personalità giuridica, o persino qualità morale e titolarità di diritti¹², sono gli “*umani*” (*human beings* o anche semplicemente *humans* nel testo delle linee guida) che tornano a emergere in primo piano, come perno di una visione etica dell’IA e come soggetti meritevoli di tutela

¹¹ Gli sviluppi dell’IA consentono forme di controllo, monitoraggio e sorveglianza massiva sulla popolazione che portano con sé il rischio di quello che è stato definito “*autoritarismo digitale*”. Sul tema cfr. N. Wright, *How Artificial Intelligence Will Reshape the Global Order. The Coming Competition between Digital Authoritarianism and Liberal Democracy*, in *Foreign Affairs*, 10 July 2018 <https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order> (sito visitato il 30 novembre 2019).

¹² D.J. Gunkel, *Robot Rights*, The MIT Press, 2018.

nell'interazione con i sistemi intelligenti. Se nella vicenda concettuale della soggettività giuridica abbiamo assistito storicamente ad una progressiva specificazione di nozioni diverse (“*soggetto*”, “*individuo*”, “*cittadino*”, “*persona*”)¹³ e successivamente ad un moltiplicarsi dei soggetti riconosciuti come titolari di diritti (i popoli, le culture, le generazioni future, gli animali, e così via), oggi in realtà il progresso scientifico-tecnologico, portando all'attenzione dell'etica e del diritto i temi della robotica e dell'intelligenza artificiale, ha l'effetto di restituire pregnanza a una versione per così dire molto più basilica ed essenziale della soggettività: quella che fa riferimento direttamente alla condizione di “*umanità*”. Parlare in senso normativo di *humans* in contrapposizione a *IA*, *machines*, *bots* è un uso linguistico che è anche spia della necessità di un nuovo modo di pensare, designare e tutelare la soggettività¹⁴.

¹³ Cfr. per tutti S. Rodotà, *Dal soggetto alla persona*, Editoriale Scientifica, 2007.

¹⁴ Questo uso linguistico attestato nelle linee guida europee rispecchia una tendenza ormai invalsa socialmente a vari livelli. A conferma si può considerare un dettaglio indicativo. La procedura di compilazione di *form online* prevede, come è noto, il CAPTCHA (“*Completely Automated Public Turing Test to tell Computers and Humans Apart*”); solitamente consiste nell'invito all'utente ad eseguire semplici operazioni logiche o nella richiesta di selezionare l'affermazione “*I am not a robot*”, mentre nei siti di alcune istituzioni contiene espressamente una richiesta all'utente di dichiarare la propria umanità; è previsto infatti uno step intitolato “*Confirm Humanity*”, in cui l'utente, per poter perfezionare l'iscrizione, viene invitato a confermare il suo *status* di essere umano: “*Before we subscribe you, we need to confirm you are a human*”.