

Verso la tutela giuridica dei sistemi intelligenti. Prospettive critiche della soggettività robotica

Giulia De Bona

Departamento de Ciencias Jurídicas Internacionales
e Históricas y Filosofía del Derecho
Universidad de Córdoba
z12bobog@uco.es

Abstract: The emergence of sophisticated self-learning technologies in artificial intelligence has led to a real paradigm shift, leading to the transcending of the boundaries within which law is accustomed to operate and giving rise to real regulatory gaps. The most debated issue is that of who can be identified as being liable for damages caused by the conduct of artificial intelligence-programmed systems. However, as an essential prerequisite to finding a solution to this question, it is necessary to understand what form of recognition and protection to grant to an intelligent system. This implies, upstream, the need to delve into the philosophical assumptions of artificial subjectivity and, first and foremost, to rethink the very idea of human subjectivity that connotes the construct of the entire current legal apparatus. In particular, in the context of the emerging robotic subjectivity, current reflections, cleansed of anthropocentric prejudices, call for the acceptance of a de-anthropomorphisation of the definition of autonomous agent subject, linking it not so much to self-consciousness and inner psychological states, but rather to the decision-making and communicative capacity of the intelligent entity, which would acquire an identity on the ethical and social plane.

Keywords: *personalità elettronica, soggettività robotica, soggettività funzionale, soggettività intenzionale, teoria attore-rete, coscienza artificiale.*

1. Considerazioni preliminari

«L'intelligenza artificiale (IA) non è fantascienza: fa già parte delle nostre vite»¹, così esordisce la comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato Economico e Sociale Europeo e al Comitato delle Regioni, intitolata *L'Intelligenza Artificiale per l'Europa*.

L'intervento di questa nuova scienza ha condotto all'affacciarsi di quella che viene comunemente definita *Quarta Rivoluzione*; si sta, infatti, assistendo ad un mutamento epocale, segnato da una forte presa di posizione da parte di strumenti e macchine programmati secondo l'Intelligenza Artificiale, o più semplicemente IA.

Di fronte a questo fenomeno, l'impressionante sforzo richiesto ai giuristi moderni è quello di guardare all'intelligenza artificiale in una prospettiva proattiva, tesa all'individuazione preventiva di forme di *governance*, senza attendere che l'evidenza dei fatti costringa ad adottare soluzioni di stampo consequenzialista. Il punto di partenza dovrebbe essere quello di cercare di rispondere all'esigenza di tutela dei soggetti rispetto alle condotte poste in essere dei sistemi intelligenti; infatti, l'intervento di tecnologie sofisticate, del calibro del *machine* e *deep learning*, hanno reso le macchine capaci di autoapprendimento, e quindi in grado di modificare se stesse in base all'esperienza mettendo duramente in crisi la distinzione *persona-res*, e riflettendosi sul regime di responsabilità loro ascrivibile. Alla luce di questa nuova realtà abitata da uomini e artefatti tecnologici, appiattirsi sulle tradizionali categorie normative e semantiche frutto della costruzione antropocentrica del diritto attuale, allontana dalla possibilità di guardare con lucidità alla soluzione prospettabile: l'alternativa tra il continuare a trattare i sistemi *AI-powered* come meri oggetti sotto il controllo umano, o elegerli a veri e propri soggetti di diritto al pari degli uomini è mal posta, e conduce ad un vicolo cieco in entrambi i casi; tuttavia, ciò fa comprendere che analizzare lo *status* giuridico dei sistemi intelligenti si presenti come il passaggio immediatamente prodromico al fine della identificazione della disciplina legale applicabile in termini di responsabilità. Ecco allora che nel presente studio si propone di esaminare, con particolare riguardo all'elaborazione di una cd. "soggettività robotica", le due principali possibilità aperte, ossia la strutturazione di una soggettività intenzionale e di una soggettività funzionale, con l'obiettivo precipuo di porre sotto i riflettori e prima d'altro la comprensione etica e giuridica della condizione umana, perché «il diritto appartiene alla struttura universale dell'esistenza umana; non si limita ad un ruolo di mera qualifi-

¹ Commissione Europea, 2018.

cazione o validazione e avvaloramento estrinseci del fenomeno diritto»². Il diritto è da intendersi proprio come una approfondita comprensione della condizione umano-esistenziale e della sua evoluzione; e se, *ubi societas ibi ius*, allora è proprio l'odierna società, questa nuova organizzazione abitata anche da sistemi programmati secondo l'intelligenza artificiale, che legittima e stimola lo studio che segue.

Allo stato dell'arte, un'integrazione giuridica, etica e sociale delle macchine intelligenti in un sistema che, seppur aggiornato, non è ancora pronto ad accogliere simili innovazioni e richiede pertanto tempestive azioni di reinterpretazione e di adeguamento, appare, chiaramente, molto complessa. A tal proposito, nel tentativo di garantire un congruo bilanciamento tra i vantaggi e gli svantaggi correlati all'impiego dell'intelligenza artificiale, nell'aprile 2021, è stata avanzata la proposta di Regolamento del Parlamento Europeo e del Consiglio con la finalità dichiarata «di definire le opzioni strategiche su come conseguire il duplice obiettivo di promuovere l'adozione dell'IA e affrontare i rischi associati a determinati utilizzi di tale tecnologia», basando una simile ricerca sui valori e sui diritti fondamentali dell'Unione Europea, al fine di trasformare l'IA in «uno strumento per le persone e un fattore positivo per la società [...]»³.

Si tenga presente che gli strumenti giuridici in senso stretto e la legislazione non sono sufficienti al raggiungimento dello scopo. Infatti, la recente storia insegna che le conseguenze della ricerca scientifica, tecnologica ed informatica sembrano mostrare l'esigenza di una regolamentazione dettagliata per la corretta convivenza di questi nuovi protagonisti intelligenti con gli esseri umani. È nata così la roboetica⁴, con l'obiettivo primario di ricercare, e poi di definire i valori e i limiti comportamentali da imprimere alle macchine intelligenti, in modo da realizzare una pacifica coesistenza tra le stesse e le persone. Difatti, dal momento che «la trasformazione sociale in cui la rivoluzione informatica si sostanzia produce una modificazione significativa dell'esperienza etica e della comprensione delle dinamiche morali della collettività che a sua volta: 1) incide nella determinazione del contenuto dei valori fondamentali richiamati giuridicamente, 2) guida la riformulazione normativa e, agendo sul senso di obbligatorietà, 3) incide sull'efficacia delle regole così come determinate fino a quel momento⁵», appare irrinunciabile un appropriato studio anche delle implicazioni e degli sviluppi etici e morali sottesi alla nascita relazione uomo-macchina⁶.

² Montanari, 2009, 1.

³ Commissione Europea, 2021.

⁴ Veruggio, 2005, 1-4.

⁵ Sarra, 2021, 28.

⁶ Daly, Hagendorff, Hui, Mann, Marda, Wagner, Wei Wang, 2022, 182-201.

Nell'affrontare questo percorso di regolamentazione e limitazione del fenomeno dell'intelligenza artificiale è opportuno scongiurare il rischio del superamento di quella che Benanti chiama «condizione biologica dell'uomo»⁷; ed al proposito lapidario è stato il Parlamento Europeo, che al punto 3 della Risoluzione del 16 febbraio 2017 recante raccomandazioni alla Commissione, precisa: «Lo sviluppo della tecnologia robotica dovrebbe mirare a integrare le capacità umane e non a sostituirle»; inoltre, prosegue il Parlamento, «[è] fondamentale, nello sviluppo della robotica e dell'intelligenza artificiale, garantire che gli uomini mantengano in qualsiasi momento il controllo sulle macchine intelligenti [...]»⁸. Data questa premessa, i lavori portati a compimento da un ristretto gruppo di esperti nella redazione di un documento intitolato *An Ethical Framework for a Good AI Society* sono fecondi; questi propongono l'elencazione di cinque principi etici – *beneficence, non maleficence, autonomy, justice, explicability* - sui quali fondare lo sviluppo e l'adozione dell'Intelligenza Artificiale, e di venti Raccomandazioni per la valutazione, l'affermazione, lo sviluppo e il supporto della IA⁹; più di recente la AI-HLEG ha recuperato questi contenuti inserendoli ed estendendoli nell'*Assessment List for Trustworthy AI*¹⁰. Viene, così, fornita una cornice che funge da apparato architettonico a partire dal quale costruire leggi, regole e standard tecnici per un'industria e una giurisdizione attuale.

Di fronte ad un simile quadro etico e giuridico di riferimento, il quale offre una sintesi coerente per la comprensione dei principi e delle politiche sulle quali fondare l'inserimento dell'intelligenza artificiale nella società contemporanea, si evidenzia quale presupposto necessario il fatto che l'IA debba essere utilizzata dall'uomo per supportarne e migliorarne le decisioni, sempre nel rispetto della sua piena autonomia; in altre parole, l'essere umano deve poter mantenere uno sguardo di supervisione rispetto all'automatismo, un legame indissolubile tra uomo e macchina che determina una secondarietà della stessa, quasi a ridurla a mero strumento.

2. Etica e diritto di fronte all'intelligenza artificiale

Come accennato nel paragrafo precedente, la questione più ardua che sta non solo animando il dibattito attuale, ma sta anche mettendo a dura prova

⁷ Benanti, 2016, 339.

⁸ Parlamento Europeo, 2017.

⁹ Floridi, Cows, Beltrametti, Chatila, Chazerand, Dignum, Luetge, Medeline, Pagallo, Rossi, Schafer, Valcke, Vayena, 2018, 689-707.

¹⁰ Nello specifico sette sono i punti citati nel documento: Human Agency and Oversight; Technical Robustness and Safety; Privacy and Data Governance; Transparency; Diversity, Non-discrimination and Fairness; Societal and Environmental Well-being; Accountability.

gli studiosi è quella che si propone di indagare come possa fare ingresso nei sistemi giuridici nazionali ed europei un ente programmato secondo l'intelligenza artificiale: lo si deve continuare a considerare mera *res*, un mero strumento, nelle mani dei soggetti "controllori" umani, o potrà ricevere riconoscimento e tutela grazie all'individuazione di una forma di personalità elettronica? La soluzione di un simile quesito è di non trascurabile rilevanza, questo perché comprendere il grado di riconoscimento adattabile al sistema intelligente è fondamentale al fine della successiva strutturazione di un regime di responsabilità per i danni cagionati da tali macchine.

Nel febbraio 2020 la Commissione Europea presenta al Parlamento Europeo la *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*¹¹, con la quale si propone di fornire una panoramica chiara e affidabile del regime giuridico capace di competere con la polivalenza dello sviluppo tecnologico ed informatico.

La difficoltà riscontrata è quella di capire a chi imputare la responsabilità in caso di danni cagionati da una macchina programmata secondo l'intelligenza artificiale. Numerose sono le opinioni in continua elaborazione, e certamente interessante è quella proposta dalla Commissione Europea, la quale ha ottenuto grande accoglimento.

Da un lato ci si avvale delle Direttive Comunitarie in tema di responsabilità da prodotto difettoso, la quale completa la disciplina nazionale individuando una responsabilità oggettiva in capo al produttore. Rispetto a questa soluzione parziale, tuttavia, si aggiunge la difficoltà di individuare la figura del produttore tra le tante che si susseguono nel processo di realizzazione di un sistema intelligente, quali il fornitore, il programmatore del *software*, l'assemblatore dell'apparecchio fisico. La Commissione Europea cerca di dare una soluzione a monte, prevedendo che la responsabilità del produttore, con riferimento alle macchine intelligenti il cui comportamento futuro possa essere determinato anticipatamente, risieda nel fornire le istruzioni e le informazioni per una adeguata loro utilizzazione, le quali dovranno tener presente una attenta valutazione del rischio legato all'utilizzo della macchina¹².

¹¹ Commissione Europea, 2020.

¹² Cfr. *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità* della Commissione al Parlamento Europeo, al Consiglio e al Comitato Economico e Sociale Europeo, 19 febbraio 2020. «Nella misura in cui il "comportamento" futuro dei prodotti basati sull'intelligenza artificiale può essere determinato in anticipo mediante la valutazione del rischio (...), il quadro dell'Unione in materia di sicurezza dei prodotti fissa già l'obbligo per i produttori di tener conto, nella valutazione dei rischi, dell' "uso" dei prodotti per tutto il loro ciclo di vita. Stabilisce inoltre che i fabbricanti devono fornire agli utilizzatori istruzioni e informazioni sulla sicurezza o avvertenze.»

Le norme interne ed europee riescono a dare una soluzione al problema della responsabilità e del relativo risarcimento, fino, però, al momento in cui il *robot* viene ridotto a strumento usato dall'uomo. Infatti, la questione si acuisce nel momento in cui entrano in campo gli algoritmi di autoapprendimento, i quali donano un soffio di vita pulsante e forgiano la *mentalità* del sistema¹³ e possono condurre a comportamenti non preventivati dal programmatore: tanto più autonomi saranno i *robot* e tanto più complesso sarà continuare a considerarli meri strumenti nelle mani degli esseri umani. Da questa prospettiva, si comprende come i sistemi normativi attuali non siano assolutamente sufficienti per affrontare una simile rivoluzione informatica; è sorta così l'idea di introdurre una personalità *ad hoc* per il sistema programmato secondo l'intelligenza artificiale. Riconoscere una personalità ad un sistema informatico conduce a compiere una considerazione degna di rilievo, ossia che coloro che sono dotati di personalità giuridica sono anche titolari di un complesso di situazioni giuridiche attive o passive, in termini di diritti ed obblighi: è questo il caso anche dell'intelligenza artificiale? In altre parole, spingersi a riconoscere la macchina come soggetto di diritto significa porre sotto i riflettori la sua natura giuridica.

Ma la questione della responsabilità, che forse appare come quella che più concretamente abbisogna di approfondimento e ricerca costante di soluzione, non si pone solo sotto il punto di vista meramente giuridico; infatti, un grande peso ha anche il profilo etico e sociale.

Si pensi ad una situazione eticamente e giuridicamente complessa come quella dell'inserimento nelle strade dei veicoli autonomi¹⁴: normalmente questo tipo di veicoli sono dotati di sistemi di localizzazione GPS, e in aggiunta sono connessi a dai sensori disposti sulle strade; quindi, viaggiano in «ambienti cooperativi che favoriscono la prevedibilità e dunque la controllabilità dei loro comportamenti»¹⁵. Simili considerazioni contribuiranno ad una diminuzione esponenziale degli incidenti stradali dovuti principalmente a fattori umani, come inesperienza, errori di percezione, lentezza di reazione; e allo stesso tempo renderanno più accessibili gli spostamenti a quelle categorie di persone non totalmente idonee alla guida. Queste argomentazioni offrono ottimi elementi a favore dei veicoli autonomi, dal punto

¹³ Ruffolo, 2020, 98.

¹⁴ La Society of Automotive Engineers International (SAE International), a proposito di autonomia operativa di macchine programmate secondo l'intelligenza artificiale, ha redatto un elenco gerarchico di livelli di autonomia (L0-L5): L0 è il livello più basso ed indica una assoluta assenza di automazione, mentre L5 è il livello di completa autonomia. Attualmente in commercio vi sono veicoli con un livello di autonomia L3, ossia dotati del sistema di *autopilot*. Cfr., SAE International 2018.

¹⁵ Amoroso, Tamburrini, 2019, 43.

di vista dell'«etica delle conseguenze»¹⁶, per la quale si dovrebbe fondare la valutazione di ogni decisione solo ed esclusivamente sulle conseguenze negative o positive scaturenti dalla azione stessa, ed in questo specifico caso si potrà notare una massimizzazione del benessere della collettività. Dal punto di vista dell'etica deontologica, ove si parla più precisamente di «etica dei doveri»¹⁷, la quale si concentra sui doveri morali e sui diritti delle persone che non sono derogabili in base alla valutazione delle conseguenze attese, si dovrebbe affidare all'uomo, e non alla macchina, il compimento di certe categorie di attività riconosciute come eticamente sensibili.

Lo scenario diviene più complesso quando si passa a discutere il dilemma delle collisioni inevitabili¹⁸, ponendo sotto i riflettori la questione del giudizio morale sulle azioni dei veicoli automatici: a causa della scarsa visibilità, un veicolo autonomo si trova nella situazione di dover effettuare una manovra di emergenza per evitare un pedone sulle strisce pedonali; tuttavia, l'unica manovra possibile è suscettibile di investire due ciclisti che in quel momento si trovano a bordo strada. Questo esempio fa emergere una grave tensione tra le due tipologie di etiche normative: mentre quella consequenzialista sarà portata a scegliere moralmente, in un'ottica di minimizzazione dei danni, per il sacrificio di una sola vittima, anziché due; la posizione deontologica non riterrà mai moralmente accettabile una simile azione, in quanto ravvedrebbe una strumentalizzazione della vita, in violazione dei principi della dignità umana.

Per ora l'unica soluzione è stata offerta dal Rapporto della Commissione Etica su *Guida automatica e connessa*, nominata dal Ministero dei Trasporti e delle Infrastrutture Digitali tedesco, la quale alla regola numero 9 afferma:

*in the event of unavoidable accident situations, any distinction based on personal features (age, gender, physical or mental constitution) is strictly prohibited. It is also prohibited to offset victims against one other. General programming to reduce number of personal injuries may be justifiable. Those parties involved in the generation of mobility risks must not sacrifice non-involved parties*¹⁹.

¹⁶ Tamburrini, 2020, 17.

¹⁷ Tamburrini, 2020, 28.

¹⁸ Questo dilemma di stampo tecnico-scientifico, incentrato sull'autonomia dei sistemi intelligenti, si pone come sviluppo logico del «dilemma del carrello», introdotto dalla filosofa Philippa Foot nel 1967. Si tratta di un esperimento di matrice etico-filosofica, la cui versione originale prevede: un tram sta per investire un gruppo di cinque operai che si trovano sulla sua traiettoria, un osservatore assiste alla scena ed ha la possibilità di azionare la leva per modificare il tragitto del tram, facendolo confluire su un altro binario nel quale vi è un'unica persona al lavoro. L'osservatore potrà scegliere se sacrificare una persona, salvando i cinque operai o viceversa.

¹⁹ Ethik-Kommission, 2017.

Viene quindi rimarcato il divieto di operare ogni genere di discriminazione, in accordo con la lettura deontologica; ma, viene consentito, conformemente alla posizione consequenzialista, di programmare i veicoli autonomi in modo da limitare i danni, e di ripartire gli stessi solo tra le persone coinvolte nell'incidente.

In questo contesto, si evidenzia una ricerca di regole etiche e morali in base alle quali definire le forme del controllo umano sui sistemi autonomi, per la creazione di un sistema giuridico adeguato. Per la realizzazione di un simile obiettivo è opportuno considerare il controllo umano come un «meccanismo di salvaguardia»²⁰ volto ad impedire che la macchina provochi danni evitabili. Inoltre, il controllo umano diverrebbe utile a livello di imputazione della responsabilità: in questi termini, infatti, sarebbe più semplice individuare il soggetto responsabile per eventuali danni. Ma soprattutto, il controllo umano consentirebbe di ravvedere, dietro il sistema automatico, un «agente morale»; in tal senso, la macchina diverrebbe un mero «delegato morale»²¹.

A questo punto allora ci si potrebbe domandare: se si volesse riconoscere, come anticipato in apertura di questo paragrafo, una personalità giuridica dalla quale deriverebbero diritti e doveri per la macchina, sarebbe possibile affidare, nell'ambito di un'etica delle macchine e "per le macchine", simili decisioni al sistema intelligente? Tentare di dare una risposta ad una simile domanda non fa altro che ampliare la vasta gamma di questioni di natura ontologica, epistemologica e pratica già pendenti²², ma allo stesso tempo stimola la necessità di ritornare ai concetti ed alle categorie fondamentali del diritto e dell'etica.

3. Coscienza umana e intelligenza artificiale: la questione della soggettività robotica

Ancor prima di prendere seriamente in considerazione la questione relativa alla configurazione di una forma di responsabilità per i danni cagionati da condotte ascrivibili ai soggetti intelligenti, non banale e di assoluta priorità è la questione di quale possa essere la forma di soggettività attribuibile a questi nuovi protagonisti della realtà etica, giuridica, filosofica e sociale. La problematica, in realtà, si pone come molto più complessa rispetto a quello che ci si possa aspettare perché richiede di porre, a monte, un'altra questione, ossia quale sia il significato di soggettività umana, che appare un modello costantemente imitato nella programmazione di un sistema di intelligenza artificia-

²⁰ Amoroso, Tamburrini, 2019, 51.

²¹ Amoroso, Tamburrini, 2019, 51.

²² Wallach, Allen, 2009.

le. Comprendere se un ente intelligente possa o meno essere dotato di una soggettività significa innanzitutto ripensare all'idea stessa della soggettività che ha connotato la costruzione dell'intera struttura giuridica attuale che, in gran parte, è debitrice della cultura moderna.

Oramai emerge con evidenza il fatto che le macchine intelligenti siano in grado non solo, come in origine, di risolvere problemi definiti complessi, ma anche di riprodurre condotte e comportamenti frutto di stimolazioni esterne. Da questo punto di vista, possono certamente «essere paragonate a soggetti morali e possono essere considerate come centri di autodeterminazione volitiva e d'imputazione di responsabilità etica e giuridica [...]»²³. Pertanto, per capire davvero se un sistema intelligente possa essere dotato di soggettività piena è opportuno indagare la natura della soggettività.

Secondo la ricostruzione della soggettività umana offerta da Platone, questa si presenta in una struttura tripartita, la quale consta di: ragione, definita come l'anima di un soggetto intelligente; volontà, ossia quella parte dell'anima che realizza il comportamento; e sentimento, il *pathos* aristotelico²⁴. Al proposito si può osservare che per quanto esistano sistemi intelligenti capaci di simulare i processi di ragionamento umano sino a realizzare una forma di intelligenza potenziata, ad immagine e somiglianza di quella umana, tanto da costituirne un'estensione, appunto, artificiale, ancora non si dispone di macchine capaci di compiere attività totalmente connaturate al pensiero umano, inteso come massimo contenuto della coscienza. Gli aspetti filosofici della questione possono essere sintetizzati nei seguenti interrogativi: un mero processo di calcolo che avviene all'interno di un sistema programmato secondo l'intelligenza artificiale può essere considerato simile al pensiero? O ancora, il pensiero umano può essere semplicisticamente considerato un processo di calcolo?²⁵ Per ora non ci sono ancora risposte convincenti, ciò che può essere offerto è una disamina interdisciplinare che coinvolga il settore filosofico, antropologico, informatico e neuroscientifico.

Innanzitutto, si tenga presente che le attività mentali non possono essere ridotte solo ad attività cognitive, intese come elaborazioni meccaniche e razionali di simboli, ordini e dati. L'esperienza empirica insegna, infatti, che sono diventati sempre più irrinunciabili «gli aspetti legati alla capacità del cervello di determinare il soggetto unitario, capace di fare esperienza di sé stesso e del mondo circostante, capace di genuine e primarie capacità semantiche intenzionali nei confronti di eventi esterni, passati e futuri; in grado di

²³ Moro, 2021, 55.

²⁴ Moro, 2021, 62.

²⁵ Buttazzo, 2002, 16.

produrre motivazioni e fini, valori soggettivi e sensazioni. In poche parole, la coscienza»²⁶.

Nel compiere un'azione, all'essere umano è riconosciuta una consapevolezza della scelta effettuata rispetto alle altre possibilità presentate, però è vero anche che vi sono delle azioni completamente autonome e meccaniche rispetto alle quali non si può parlare puramente di scelta, essendo esse compiute in base ad una spontaneità involontaria data dal riproporsi di scenari già più volte vissuti. Si pensi, ad esempio, alla guida: inizialmente si dovrà prestare molta attenzione nell'abbassare ordinatamente la frizione prima di cambiare marcia; una volta acquisito il modello procedimentale il livello di attenzione richiesto si abbasserà notevolmente, e l'atto assumerà le fattezze di un automatismo. Quando ci si riferisce ai sistemi intelligenti, si parla del loro agire *meccanicamente* proprio per fare riferimento a questa mancanza della coscienza di un sé capace di eleggere, fra le tante, la possibilità di azione più appropriata.

Inoltre, come emerge dall'analisi appena compiuta, sin dagli esordi dell'IA si è sempre messa in dubbio la circostanza che per essa si potesse parlare di "coscienza di sé", basandosi la sua operatività sul pedissequo rispetto di un procedimento di logica computazionale. Tuttavia, partendo dalla considerazione che i più sofisticati sistemi di intelligenza artificiale sono capaci di migliorare loro stessi imparando dagli "errori" commessi e ricalibrando la loro azione²⁷, non risulta essere questa espressione di una coscienza di sé²⁸?

Peraltro, per indagare la possibilità di avere dei sistemi intelligenti coscienzi è necessario ripercorrere secoli di storia e di evoluzione culturale filosofica, scientifica e psicologica. Nel fare ciò, si riscontra che l'elemento che rende unico l'uomo in quanto essere vivente sia la coscienza, per la quale, però, è molto difficile individuare una definizione unitariamente accettata. Contestualizzando il problema definitorio con riferimento all'argomento qui trattato ci si chiede: come si può costruire una macchina pensante e cosciente, se ancora non è chiaro come la coscienza sia manifestata dagli esseri umani?

Con la parola coscienza si intende quella forma di consapevolezza che ciascun soggetto ha di sé stesso e del mondo esterno con cui si rapporta, della propria identità e del complesso delle proprie attività interiori. Per la filosofia moderna il significato di coscienza è legato alla sfera dell'interiorità umana, e viene espresso come la risultante della relazione tra il proprio "io" interiore e la realtà quotidiana con la quale si viene in contatto. Si tratta di una com-

²⁶ Manzotti, Tagliasco, 2002, 92.

²⁷ Turing, 1950, 433-460.

²⁸ L'idea che si possano creare dei sistemi intelligenti capaci di avere coscienza di sé è la protagonista degli studi sulla *Strong AI* o intelligenza artificiale forte.

ponente che permette di prendere consapevolezza del mondo circostante e della presenza in ciascun soggetto di stati e fenomeni rispetto ai quali non può aversi una separazione; la coscienza incarna, pertanto, la capacità di un soggetto di fare esperienza dei propri pensieri, dei propri sentimenti, di sé stesso e della realtà dei fatti. Per enfatizzare il concetto, Balconi sostiene che «la consapevolezza corporea è legata alla rappresentazione propriocettiva, in particolare, all’esperienza conscia della localizzazione di specifiche parti del corpo nello spazio, una funzione che permette di definire i confini corporei, di esplorare gli oggetti e di guidare i movimenti finalizzati»²⁹. A partire da questa ultima considerazione si può azzardare un paragone con un sistema intelligente dotato di sensori di movimento o sonori: questi strumenti, essenziali al funzionamento del sistema stesso, gli permettono di avere una visione a tutto tondo dell’ambiente circostante, ma prima ancora della sua posizione all’interno dello spazio considerato, per calibrare e programmare l’azione finale. Senza forzare eccessivamente il contenuto definitorio qui considerato, non si sta forse parlando anche in questo secondo caso di una forma di consapevolezza?³⁰

Si aggiunga, inoltre, che non può sperimentarsi la presenza di una coscienza in un altro cervello, in quanto si configura come una proprietà interna dello stesso, osservabile solo dal suo possessore e non misurabile esternamente. Così dicendo, tuttavia, si giunge a concludere che non si potrà mai essere certi della presenza della coscienza in un altro essere, sia esso umano o non umano³¹.

Un approccio pragmatico suggerirebbe, anche in ossequio all’esperimento messo a punto da Turing, di considerare autocosciente una macchina che, superando delle prove scientifiche, sia in grado di convincere l’essere umano della sua capacità. Per di più, si consideri anche che tra gli esseri umani il riconoscimento dell’altro come cosciente avviene grazie ad un’applicazione automatica della similitudine: «poiché tutti noi siamo fatti allo stesso modo è ragionevole credere che la persona che ci sta davanti sia anch’essa autocosciente»³². Appare chiaro che questa sorta di principio non possa applicarsi per analogia anche al sistema robotico; difatti l’obiezione comune che viene

²⁹ Balconi, 2014, 83.

³⁰ Di assoluta attualità è la notizia trapelata da un gruppo di ricercatori dell’Università Columbia a New York, i quali hanno messo a punto un braccio robotico che, privo di dati sulla sua forma, è riuscito autonomamente ad avere “coscienza” della propria posizione e della propria struttura iniziando, di conseguenza, a rapportarsi con il mondo naturale; ciò è stato possibile solo grazie a delle telecamere che hanno permesso al braccio robotico di compiere l’attività di “auto-modellazione”, che può essere davvero definitiva come una primitiva forma di coscienza. Chen, Kwiatkowski, Vondrick, Lipson, 2022, 1-13.

³¹ Hofstadter, Dennett, 2001.

³² Buttazzo, 2002, 19.

sollevata è proprio che la coscienza artificiale consisterebbe in una serie di impulsi elettronici che operano in modo deterministico e programmato, ed il loro non poter provare emozioni e sentimenti li rende totalmente *altro* rispetto all'essere umano. Tuttavia, questo ragionamento presenta una debolezza, ossia il fatto che potrebbe essere applicato anche al cervello umano, la cui attività è regolata da specifiche leggi scientifiche che spiegano i processi fisico-biologici degli impulsi che muovono l'apparato neurale.

A questo punto, se si ritiene che la coscienza in quanto capacità sia nient'altro che una proprietà complessa del cervello, che si attua grazie ad una struttura neurale evoluta e organizzata, allora, sarebbe anche verosimile la costruzione di un sistema cosciente³³, al quale attribuire una soggettività piena. Ma, se così fosse, come sarebbe la coscienza di un sistema intelligente? Sarebbe sufficiente la presenza di una rete neurale artificiale, capace di “produrre” una simile coscienza, o anche in questo caso sarebbe necessario un lungo processo di sviluppo e perfezionamento dell'esperienza, come è accaduto per gli esseri umani?

Di fronte ad una ricostruzione così individualista e antropocentrica della soggettività, che abbisognerebbe di rivisitazioni in un'ottica rinnovata, forse il quesito più rilevante al proposito è il seguente: davvero si identificano nell'autocoscienza, nei sentimenti o nelle emozioni gli elementi caratterizzanti la soggettività umana, e allo stesso tempo gli elementi discriminanti utili per l'attribuzione ad un sistema di intelligenza artificiale di una forma di soggettività e poi anche di personalità?

Si tratta evidentemente di disquisizioni che sono alla base del dibattito etico e giuridico, infatti, se si riconoscesse una simile eventualità allora si dovrebbe ripensare la collocazione non solo sociale, ma anche, e soprattutto, giuridica di tali sistemi, individuando per essi una autonomia morale, una personalità, un complesso di diritti e doveri ed anche un regime di responsabilità e di libertà *ad hoc*.

4. Natura e limiti della personalità elettronica

Gli orientamenti prevalenti³⁴ sono quelli che propongono una apertura verso lo studio di uno *status* giuridico specifico per i *robot*: in Europa ne è testimonianza la Risoluzione del Parlamento europeo del 2017 la quale al paragrafo 31 lettera *f* propone espressamente

³³ Tuttavia, non tutti sono di questo avviso, ad esempio si veda la posizione contrapposta espressa da Faggin, 2015, 1-13.

³⁴ Zimmerman, 2015.

l'istituzione di uno status giuridico specifico per i robot nel lungo termine, di modo che almeno i robot autonomi più sofisticati possano essere considerati come persone elettroniche responsabili di risarcire qualsiasi danno da loro causato, nonché eventualmente il riconoscimento della personalità elettronica dei robot che prendono decisioni autonome o che interagiscono in modo indipendente con terzi³⁵.

Ciò conduce alla conseguenza che se i *robot* venissero riconosciuti come dotati di personalità giuridica dovranno anche essere considerati alla stregua persone fisiche, aspetto pacifico nel dibattito scientifico³⁶, ove ai sistemi di intelligenza artificiale si applicano le categorie di tutela tipiche dei soggetti umani, come i diritti soggettivi³⁷. Tuttavia, l'avanzamento tecnologico è tale da ritenere che «sia insufficiente applicare ai robot i regimi e le dottrine giuridiche esistenti alla luce delle tradizionali categorie giuridiche (persone fisiche, persone giuridiche, beni mobili)»³⁸. E proprio per questo motivo che, per colmare la lacuna giuridica, si è pensato di introdurre il concetto di *personalità elettronica*, attribuibile a quei sistemi intelligenti che risulteranno così dotati di personalità piena, di capacità negoziale, e di capacità di essere titolari di diritti di proprietà, come anche di conti bancari.

In un simile scenario, pertanto, ci si interroga sulla possibilità di creare una sorta di *tertium genus*, offrendo una nuova categorizzazione giuridica adattabile alle capacità di un sistema intelligente. Ma come deve essere sviluppata tale categoria? È necessario ricondurre la personalità elettronica al modello delle persone giuridiche, anche se questo implicherebbe la presenza costante di un soggetto al quale attribuirne la rappresentanza e la direzione, e soprattutto la responsabilità penale? Oppure, si potrebbe ricondurre la personalità elettronica al modello delle persone fisiche, segnando, così, il passaggio del sistema intelligente da *res* a persona?

Non si deve, tuttavia, trascurare che soprattutto in quest'ultimo caso si potrebbero provocare gravi difficoltà di accoglimento dal punto di vista giuridico ed etico: infatti, non si rischierebbe forse di umanizzare troppo un ente artificiale?

Considerare il sistema intelligente come un vero e proprio soggetto, e non solo come un oggetto, non appare in contraddittorietà con l'ordinamento giuridico italiano che attribuisce soggettività anche alle persone giuridiche come le società; pertanto, compiere la stessa operazione anche per le macchine intelligenti non dovrebbe comportare alcun tipo di problema. Anzi, procedendo per analogia ci si rende immediatamente conto che da un lato, a

³⁵ Parlamento Europeo, 2017.

³⁶ Moro, in Ruffolo (a cura di), 2021, 56.

³⁷ Taddei Elmi, in Lombardi Vallauri, 1990, 685-711.

³⁸ Comitato Nazionale per la Bioetica e Comitato Nazionale per la Biosicurezza, 2017.

differenza di una società, la macchina intelligente non sarebbe dotata di un substrato umano, ma dall'altro lato sarebbe caratterizzata (quantomeno per i sistemi dotati di corporeità sintetica) da una fisicità che la avvicinerrebbe maggiormente alla persona umana: «sarebbero persone “fisiche” entrambe, quella robotica come quella umana»³⁹.

Ecco allora che un parallelismo tra persona fisica e giuridica già lo si può vedere, ma si tratta solo di apparenza, in quanto dietro la persona giuridica si celano comunque persone fisiche autocoscienti e capaci di decisioni logiche causali. Attribuire personalità elettronica ad una macchina significherebbe riconoscere che nella stessa si conciliano sia la capacità decidente, che quella dichiarante, le quali si esplicherebbero secondo modalità “artificiali”, ossia sulla base di logiche di tipo inferenziale.

Inoltre, l'entificazione di un soggetto robotico sarebbe utile solo nel momento in cui le sue decisioni fossero soddisfacenti l'interesse del sistema intelligente stesso, e non invece l'interesse altrui, come oggi ancora accade. Come sostiene Ruffolo, delegare il processo decisionale alla macchina non conduce necessariamente all'esigenza di una sua personalizzazione e di riconoscimento di una qualche forma di capacità, dal momento che l'unica personalità e capacità essenziali sono quelle già presenti in capo al soggetto nel cui interesse viene assunta la decisione.

Se così fosse, la manifestazione di volontà del soggetto umano sarebbe riferibile *per relationem* alla comunicazione dello «schiavo digitale sapiente»⁴⁰; pertanto, a prescindere dalla predicabilità della scelta algoritmica (dettata da algoritmi più o meno sofisticati), a monte si presenta la volontà del soggetto umano di «farsi imputare una dichiarazione formata sulla base di meccanismi di scelta digitale non programmabili né prevedibili, affidandosi alla saggezza ed “esperienza” del proprio schiavo elettronico sapiente»⁴¹. Così dicendo, diviene però privo di significato interrogarsi continuamente sulla necessità di parlare di rappresentanza, personalità giuridica, capacità giuridica parziale, essendo sufficiente fare riferimento alla figura del soggetto umano.

Rispetto a questa “equiparazione” tra persona fisica e robotica, per di più, si pongono forti attriti dovuti principalmente al fatto che nella maggior parte dei casi, al fine del riconoscimento della personalità a soggetti di diritto, si ritiene necessaria almeno una forma di autocoscienza, invocando il principio cartesiano del *cogito ergo sum*, o quantomeno una componente o una manifestazione emotiva e sentimentale. Tuttavia, ridurre queste argomentazioni alla sfera profondamente antropocentrica e antropomorfa che caratterizza lo

³⁹ Ruffolo, 2021, 117.

⁴⁰ Ruffolo, 2021, 123.

⁴¹ Ruffolo, 2020, 227.

studio del diritto, non conduce ad altro che ai modelli giuridici attuali ed ai pregiudizi che la fondano. Per di più, si consideri, come sottolineato anche da Ruffolo, che le forme di tutela degli esseri umani sono costruite a prescindere dal grado della loro intelligenza nonché della loro coscienza, dal momento che si riconoscono anche al neonato o all’embrione; da questo punto di vista «reputare che la macchina idonea ad assurgere a persona sia solo quella pensante e cosciente secondo il modello umano è un inconsapevole regresso verso l’antica tesi che la intelligenza venga dall’anima e che, dunque, solo gli esseri umani possano esserne dotati»⁴².

È bene tenere a mente, però, che si parla di “Intelligenza Artificiale” proprio perché si pone come somigliante a quella umana, anzi costruita in via analogica studiando il sistema di ragionamento umano; ma non si pone come corrispondente o equivalente all’intelligenza degli uomini. Si è, pertanto, di fronte a due forme di intelligenza entrambe meritevoli di riconoscimento e tutela, ma per accettare la presenza innegabile di intelligenze di altro tipo rispetto a quella umana è necessario abbandonare definitivamente i pregiudizi di matrice antropocentrica che contraddistinguono gli studi etici, giuridici, sociali e filosofici. Le riflessioni odierne, influenzate dall’evoluzione tecnologica in corso, esortano l’accoglimento della *de-antropomorfizzazione* della definizione di “soggetto agente autonomo”, ricollegandola non tanto all’autocoscienza e agli stati psicologici interiori, ma piuttosto alla capacità decisionale e comunicativa dell’entità intelligente, che le farebbe acquisire «un’identità sul piano etico e sociale»⁴³.

Se da un lato, infatti, allo stato dell’arte si possono avere macchine intelligenti in grado di calcolare e simulare i sentimenti e le emozioni, ma non in grado di provarle o di prevederle; dall’altro lato, è pur vero che l’autonomia autocosciente degli algoritmi più sofisticati non costituisce l’unico elemento che possa essere richiamato per fondare la personalità elettronica. Da questo punto di vista, estremamente convincente è la tesi di Teubner, così come dettagliatamente illustrata nei paragrafi successivi, che prescinde dalle logiche di funzionamento interno del soggetto “attante”, per concentrarsi sulla autonomia comunicativa e porre l’accento sulle conseguenze giuridiche scaturite dalla comunicazione stessa.

A supporto di queste argomentazioni si pone la recente e sempre più diffusa spersonalizzazione dell’elemento soggettivo delle fonti dell’obbligazione: da un lato, infatti, per ciò che concerne la manifestazione di volontà, «il baricentro migra verso il tenore oggettivo della dichiarazione, anche a sca-

⁴² Ruffolo, 2020, 220-221.

⁴³ Ruffolo, 2020, 117.

pito, ove divergente, della volontà soggettiva del dichiarante»⁴⁴; e dall'altro lato, per ciò che riguarda la responsabilità aquiliana, nell'individuazione dei soggetti responsabili, frequenti sono le ipotesi di riconduzione al modello della responsabilità oggettiva, con una conseguente oggettivizzazione della colpa⁴⁵. A tutto ciò si aggiunga che categorie riferibili agli elementi soggettivi, come appunto la manifestazione di volontà o la colpa, assumono rilievo anche se accostati ad associazioni o società, quali enti non umani. Si evidenzia chiaramente questa oggettivizzazione dell'elemento soggettivo che fa percepire il soggetto umano come sempre più distante rispetto alla nascita delle obbligazioni nei rapporti giuridici diretti; ove, appunto «la personalità elettronica sembrerebbe, anzi, l'approdo ultimo di quel processo di parallelo oggettivarsi delle due grandi fonti della obbligazione, volontà nel contratto e colpa nell'illecito»⁴⁶. Questo processo accorcia il divario tra uomini e enti non umani, e di conseguenza renderebbe ancor più accreditabile l'introduzione della categoria della personalità elettronica.

Pertanto, se da un lato risulta certo che la soluzione della introduzione di una personalità elettronica, dal punto di vista pratico e teorico, potrebbe essere attuabile; dall'altro lato, ci si potrebbe chiedere se questa alternativa possa essere coerente con il sistema sociale e giuridico nel quale si inserisce. L'aspetto che più rallenta l'accettazione dell'introduzione di una personalità elettronica è il rischio che dalla personificazione del sistema intelligente derivi una forma di responsabilità diretta, potenzialmente meno soddisfattiva, dal punto di vista patrimoniale, rispetto ad una responsabilità vicaria, essendo comunque limitata all'ammontare del patrimonio della macchina.

Dal punto di vista meramente pratico, una forma di responsabilità diretta e con patrimonio proprio delle macchine programmate secondo l'intelligenza artificiale non si pone come indispensabile ai fini della tutela dei soggetti, dal momento che non solo non amplia il ventaglio di strumenti messi a disposizione per far valere le proprie pretese a livello risarcitorio/riparatorio, ma addirittura potrebbe ridurre la tutela garantita a queste categorie di soggetti, i quali potrebbero essere meglio soddisfatti e garantiti da una forma di responsabilità delle persone fisiche chiamate a rispondere per la macchina intelligente, come il programmatore, il produttore o il proprietario. Infatti, da questa prospettiva, il *robot* non potrà più essere considerato un ausiliario dell'essere umano, ma dovrà rispondere a titolo "personale"; ciò condurrebbe all'esigenza di dotare questi sistemi intelligenti di diritto di proprietà di un

⁴⁴ Ruffolo, 2021, 120.

⁴⁵ Solo per inciso si noti, inoltre, che le categorie relative agli elementi soggettivi come appunto la colpa o la manifestazione di volontà, sono rilevanti anche per le società o le associazioni, quindi gli enti non umani.

⁴⁶ Ruffolo, 2021, 121.

capitale utile per ristorare gli eventuali danni cagionati, o quanto meno di una forma di copertura assicurativa.

5. Tutela riconoscibile ai sistemi intelligenti

«I predatori pericolosi che penetrano negli ambiti del diritto civile sono gli incontrollabili algoritmi del mondo digitale, i *robot*: per dirla in modo più ampio, agenti-*software* ad elevate prestazioni, provvisti di elevata intelligenza e capacità di apprendimento, i quali costituiscono pericoli nuovi e impensati persino per gli esseri umani»⁴⁷.

Le parole «pericoli nuovi e impensati» qui usate da Teubner non alludono a qualcosa di lontano e futuribile, bensì ad una pura attualità. Appare urgente quindi un intervento capace di orientare il diritto, il quale si trova ora dinnanzi ad un bivio, e dovrà scegliere quale strada percorrere: riconoscere ai sistemi autonomi programmati secondo l'intelligenza artificiale uno *status* giuridico indipendente, e di conseguenza trattarli come responsabili diretti; oppure adeguarsi al fatto che si proporranno sempre più incidenti per i quali non si riuscirà ad individuare un responsabile.

Il dilemma che sta assillando generazioni è se le macchine intelligenti possano essere, in qualche modo, assimilate agli uomini, se possano essere considerate dotate di un'anima e «se la loro intelligenza [possa] essere definita cosciente, seppur artificiale»⁴⁸.

Se si ritiene valida la considerazione alla quale è giunto Turing, secondo cui i calcolatori sono macchine capaci, al pari degli esseri umani, di esprimere un pensiero definibile autentico⁴⁹, allora si accoglie la posizione dei sostenitori dell'esistenza dell'Intelligenza Artificiale Forte per la quale un sistema intelligente compie gli stessi processi intellettuali umani. Secondo queste valutazioni, quindi, non solo si presenterebbe una «identità ontologico-qualitativa»⁵⁰ tra il cervello umano e quello meccanico/elettronico, *rectius* tra intelligenza umana e intelligenza artificiale; ma l'unica differenza tra queste due forme di intelligenza risiederebbe sul supporto fisico, quindi, la testa fatta di elementi biologici ed un *computer* fatto di supporti metallici ed elettrici⁵¹.

⁴⁷ Teubner, 2019, 19.

⁴⁸ Taddei Elmi, Romano, 2010, 146. Cfr. anche con Taddei Elmi, 2016, 115-137.

⁴⁹ Turing, 1950, 433-460. In questo articolo l'autore spiega che se un intervistatore non riesce ad operare una distinzione tra le risposte date da un computer e quelle date da un essere umano, allora se ne deduce che il computer si comporta come un uomo, ed allora potrà dirsi che «è» intelligente tanto quanto un essere umano.

⁵⁰ Contaldo, Campara, in Taddei Elmi, Contaldo (a cura di), 2020, 77.

⁵¹ Questa è la posizione dei sostenitori della tesi «funzionalista», per la quale vi sono delle forme di intelligenza che prescindono totalmente dal supporto fisico dal quale sono espresse.

A questa corrente dell'intelligenza artificiale forte, si contrappone quella dell'Intelligenza Artificiale Debole, secondo la quale le macchine possono solo imitare, simulare il comportamento intelligente umano⁵². In questo caso, ancorché i comportamenti dei sistemi intelligenti siano indistinguibili da quelli umani, sarebbe necessario verificare che la macchina abbia effettivamente voluto ciò che ha compiuto⁵³.

Da queste due diverse prese di posizione derivano due opposte concezioni dello *status* giuridico delle macchine intelligenti: da un lato i sostenitori dell'intelligenza artificiale forte sono fermamente convinti che le macchine programmate secondo l'intelligenza artificiale possano considerarsi soggetti "pensanti", e per tale motivo possano essere ritenute soggetti a tutti gli effetti. In quanto tali, risulterebbero meritevoli di essere riconosciute come titolari di una soggettività giuridica⁵⁴, che le renderebbe destinatarie di una vasta gamma di tutele, diritti e doveri. Dall'altro lato, i sostenitori dell'intelligenza artificiale debole sono portati a considerare le macchine come *res*, e quindi meritevoli di tutela solo per il valore che esse esprimono.

Il problema qui evidenziato non è solo di matrice filosofica, ossia di attribuire uno *status* ontologico alle macchine intelligenti, ma si riverbera profondamente anche sul piano giuridico e pratico-funzionale, dovendo capire se risulti utile conferire a queste entità una tutela per soggettività.

L'ordinamento giuridico italiano si conforma a due paradigmi che, se presenti, creano l'aspettativa di una tutela: la soggettività intesa come intenzionalità, autocoscienza, capacità di autodeterminazione e sensibilità; ed il valore inteso come complessità di funzioni, utilità pratica ed originalità. Date queste prime considerazioni, è evidente come la prima forma di tutela, quella per soggettività, sarà maggiormente adattabile all'essere umano; mentre, la tutela per valore potrà essere utilizzata per permettere agli enti di essere riconosciuti nel mondo del diritto.

A questo punto lo sforzo che sono chiamati a compiere i giuristi è quello di comprendere come possano essere classificate tali macchine intelligenti, e quale forma di tutela possa essere loro riconosciuta. Si tratta di operare una scelta di enorme portata,

tra la possibilità di considerare un robot come strumento tecnologico avanzato, ma pur sempre un mezzo per lo svolgimento dell'attività

In tal senso appunto, rileva la funzione piuttosto che la struttura.

⁵² Searle, 1980, 417-457. Le argomentazioni a sostegno di questa tesi sono principalmente due: non può dirsi che le macchine abbiano coscienza, la quale è propria solo degli esseri umani; le macchine possono solo calcolare e non comprendere.

⁵³ Taddei Elmi, in Lombardi Vallauri, 1990, 695. È chiaro che in questa prospettiva di grande valore è l'elemento strutturale, piuttosto che quello funzionale.

⁵⁴ Taddei Elmi, in Lombardi Vallauri, 1990.

che deve in ogni caso essere ricondotta all'uomo che se ne avvantaggia e ne risponde e l'eventualità fantascientifica – ma non più di tanto – di valorizzare la capacità dei robot di operare delle scelte e quindi di essere imputabili dell'attività compiuta⁵⁵.

Le posizioni che potrebbero essere assunte oscillano tra l'equiparare, sul piano puramente giuridico le macchine agli esseri umani, attribuendo loro diritti e doveri; e tra il continuare a considerare i sistemi intelligenti quali *res*. Femia sintetizza in maniera formidabile il problema che i giuristi, i filosofi e i sociologi si trovano ad affrontare: «L'intelligenza artificiale, le macchine capaci di memoria e apprendimento, gli agenti software in grado di prendere decisioni non calcolabili *ex ante* dai programmatori sono un rivoluzionario banco di prova per la funzione del diritto civile. Sono soggetti o no?»⁵⁶. Le domande da porsi sono quindi le seguenti: le macchine intelligenti possono essere considerate solo oggetti, o anche soggetti; ed in questa prospettiva possono godere oltre che di una tutela per valore, anche di una tutela per soggettività?

Per dare una risposta soddisfacente le strade che possono essere percorse sono due: la prima alternativa pone al centro della riflessione lo *status* ontologico dei sistemi intelligenti, e richiede un tempestivo intervento della filosofia nell'individuazione degli elementi sufficienti, ma allo stesso tempo necessari, per riconoscere un certo livello di soggettività a questi sistemi. La seconda alternativa invece si propone di accertare l'utilità dell'attribuzione di un simile valore giuridico ad entità intelligenti, quindi, si preoccupa di effettuare una valutazione di tipo pratico e funzionale.

Per quanto concerne la prima delle due vie percorribili appena enunciate, nel pensiero giuridico contemporaneo, si ritiene che si possano distinguere due diverse forme di soggettività⁵⁷: una soggettività piena, propria degli esseri umani «senzienti, immaginanti, raziocinanti-giudicanti, coscienti e autodeterminanti»; e una soggettività ridotta, riconoscibile negli embrioni, nelle persone diversamente abili o anche negli animali, essendo tutte «entità portatrici di interessi “sentiti”, capaci cioè di soffrire, provare piacere, avere sensibilità»⁵⁸. Ricondurre i sistemi intelligenti in una di queste due soggettività, significherebbe accogliere una delle due tesi prima esposte, e

⁵⁵ Di Sabato, 2017, 389. Al proposito si veda anche Teubner, 2019, 22, il quale sostiene: «Il diritto civile è posto in tal modo di fronte ad una alternativa: o riconoscere agli agenti software autonomi uno status giuridico indipendente, e trattarli in toto come soggetti responsabili, oppure rassegnarsi a registrare un numero sempre crescente di “incidenti” privi di un responsabile».

⁵⁶ Femia, 2019, 9.

⁵⁷ Lombardi Vallauri, 1990.

⁵⁸ Contaldo, Campara, *ult. op. cit.*, 80. Cfr., Taddei Elmi, 2004, 3.

schierarsi o dalla parte della intelligenza artificiale forte oppure da quella debole⁵⁹. Attualmente si riscontra un maggior accoglimento della tesi della intelligenza artificiale debole, potendosi rilevare, ad avviso di Taddei Elmi, ben quattro diversi «livelli»⁶⁰ che testimoniano la eterogeneità tra cervello e mente.

Il primo livello è quello della sensazione e della percezione, le quali possono essere vissute in modo cosciente dall'uomo ed incosciente dall'animale; invece, un sistema informatico intelligente può «registrare» tutto ciò ed eventualmente riproporlo in modo apatico⁶¹. Il secondo livello è rappresentato dall'immaginazione e dalla fantasia, grazie alle quali la mente umana riesce ad evocare elementi, oggetti, entità anche in loro mancanza; mentre, si ritiene che l'intelligenza artificiale non lo possa fare. Il terzo livello è costituito dall'intellezione: la macchina è istruita solo per riconoscere gli enunciati ed effettuare delle inferenze tra gli stessi; dal canto suo, invece, la mente sa cogliere congiuntamente tutti gli enunciati e formulare una rielaborazione. Il quarto livello invece è dato dalla coscienza, o meglio dall'autocoscienza: solo gli esseri umani sono consapevoli e coscienti del loro essere e delle loro azioni, mentre un sistema, seppur intelligente, «trovandosi in uno stato di "sonnambulismo" e di incoscienza, non si rende conto di sé stess[o] ed è solo in grado di comprendere i significati e non anche i significanti»⁶².

Affinché un sistema intelligente possa essere riconosciuto come titolare di una soggettività pura, naturale, dovrebbe essere un soggetto in senso ontologico, e si è tali solo se si dispone della cosiddetta «mente cognitiva»⁶³, ossia delle sensazioni, della consapevolezza di sé, della coscienza. Pertanto, secondo queste argomentazioni una macchina non può essere annoverata tra i soggetti ontologici, e quindi risulterà essere destinataria solo di una tutela per valore, in quanto *res*.

Difatti, approfondendo il secondo dei due paradigmi di tutela al quale si conforma l'ordinamento italiano, quale ulteriore alternativa vagliabile,

⁵⁹ Contaldo, Campara, *ult. op. cit.*, 81: «La determinazione dei requisiti minimi [per attribuire una soggettività naturale] significa, chiaramente, adottare una delle due prospettive teoriche della SAI o della WAI, omogeneità tra mente e cervello e tra stati fisici e stati mentali, oppure quello dualista dell'eterogeneità tra mente e cervello e tra stati mentali e fisici.» Si precisa che l'acronimo SAI significa: *Strong Artificial Intelligence*; mentre l'acronimo WAI significa: *Weak Artificial Intelligence*.

⁶⁰ Taddei Elmi, 2004, 3-4. Si veda anche Contaldo, Campara, *ult. op. cit.*, 81.

⁶¹ Taddei Elmi, 2004, 3-4: «Il calcolatore non ha sensazioni autentiche, guarda, ascolta, percepisce odori ma non sente, non ode, non odora; il calcolatore effettua solo registrazioni non ha sensazioni come uomini (sensazioni coscienti) e animali (sensazioni incoscienti).»

⁶² Contaldo, Campara, *ult. op. cit.*, 75. Taddei Elmi, 2016, 116.

⁶³ Contaldo, Campara, *ult. op. cit.*, 82.

si può riscontrare che la tutela per valore viene attribuita sulla base di tre criteri: complessità di funzioni, utilità pratica e originalità.

Questi possono presentarsi combinati tra loro, oppure individualmente, ma in ogni caso saranno fonte di una specifica tutela giuridica. Gli strumenti appositi sono quelli del diritto d'autore, dei brevetti e del segreto industriale, anche se attualmente si sta optando per l'introduzione di apposite norme.

Un programma, un *software*, un sistema intelligente verranno così riconosciuti come destinatari di una specifica tutela, a prescindere dal riconoscimento di un valore ontologico, ma solo grazie alla presenza di una delle tre caratteristiche (criteri di attribuzione) enunciate prima⁶⁴.

Più precisamente, il *software*, quale «espressione di un insieme organizzato e strutturato di istruzioni in qualsiasi forma o su qualunque supporto capace, direttamente o indirettamente, di far eseguire o far ottenere una funzione o un compito o far ottenere un risultato particolare per mezzo di un sistema di elaborazione elettronica dell'informazione»⁶⁵, è considerato dall'ordinamento italiano come un bene immateriale, ed in particolare come una creazione intellettuale. In quanto tale è meritevole di due forme di tutela: attraverso il brevetto si protegge l'invenzione industriale⁶⁶, e grazie al diritto d'autore si tutela il *software* in quanto opera dell'ingegno⁶⁷.

⁶⁴ Taddei Elmi, 2004.

⁶⁵ Il modello normativo sulla protezione dei Computer Software definiva i programmi come «*a set of instructions capable, when incorporated in a machine-readable medium, of causing a machine having information-processing capabilities to indicate, perform or achieve a particular function, task or result*». Organizzazione Mondiale della Proprietà Intellettuale, 1978,12.

⁶⁶ Al proposito si rimanda all'articolo 2585 del Codice Civile il quale sancisce: «Possono costituire oggetto di brevetto le nuove invenzioni atte ad avere un'applicazione industriale, quali un metodo o un processo di lavorazione industriale, una macchina, uno strumento, un utensile o un dispositivo meccanico, un prodotto o un risultato industriale e l'applicazione tecnica di un principio scientifico, purché essa dia immediati risultati industriali». Il software è tutelabile mediante il brevetto solo nell'ipotesi in cui non costituisca esso stesso l'oggetto dell'invenzione, ma si ponga come strumentale al raggiungimento dell'effetto inventivo. I diritti sull'invenzione industriale nascono proprio dal conseguimento del brevetto, grazie al quale viene garantito lo sfruttamento dell'invenzione con riferimento specifico al suo contenuto. Cfr. Piva, D'Agostini, 2003, 68.

⁶⁷ Ai sensi dell'art. 2577 del Codice Civile: «Formano oggetto del diritto di autore le opere dell'ingegno di carattere creativo che appartengono alle scienze, alla letteratura, alla musica, alle arti figurative, all'architettura, al teatro e alla cinematografia, qualunque ne sia il modo o la forma di espressione». Nel nostro ordinamento i programmi per elaboratore sono disciplinati dalla Legge d'Autore (Legge 633/1941), essendo considerate, ai sensi dell'articolo 2.8 della suddetta legge, opere letterarie «in qualsiasi forma espressi, purché originali quale risultato della creazione intellettuale dell'autore». Il diritto d'autore sorge contestualmente all'opera ed è funzionale a proteggere direttamente l'espressione creativa intrinseca all'opera, a prescindere dal suo contenuto. Cfr. Piva, D'Agostini, 2003, 69.

6. Soggettività intenzionale: stati mentali ed intelligenza artificiale

Inoltrandosi nella discussione relativa a quale possa essere lo *status* giuridico ascrivibile ai *robot*, è necessario tenere a mente che riconoscere una soggettività al sistema intelligente significa che lo stesso può essere considerato titolare di situazioni giuridiche soggettive, attive o passive che siano; pertanto, che la macchina dovrebbe divenire titolare di diritti e doveri. Se, tuttavia, ciò può avvenire, come detto prima, solo in presenza di una *mente cognitiva* e di stati mentali, che renderebbero il sistema cosciente e consapevole, emergono forti attriti rispetto alla consolidata posizione secondo la quale solo gli esseri umani ne sarebbero provvisti.

Difatti, la difficoltà di attribuire stati mentali a sistemi artificiali, e ad applicare le relative qualificazioni giuridiche, può essere facilmente superata considerando che tutte le categorie mentalistiche possano essere applicate solo agli esseri umani. Questa inclinazione, però, porterebbe al necessario riesame degli istituti giuridici già esistenti, in modo da renderli applicabili anche ai rapporti con, e tra, le entità artificiali. Ciò si potrebbe raggiungere solo tramite una oggettivizzazione di questi istituti, ottenuta grazie alla eliminazione di qualunque riferimento a fatti psicologici dal diritto. Tuttavia, obbligare il legislatore e il giurista a rimuovere i collegamenti con le attitudini mentali, oppure a prevedere delle fattispecie puramente comportamentistiche per le quali non ha rilievo l'elemento dell'intenzione, condurrebbe ad una disumanizzazione della concezione classica del diritto ed anche dei fatti sociali.

Pertanto, si evidenzia come le difficoltà maggiori, che conducono in errore gli operatori del diritto, siano proprio: da un lato, la propensione verso un'applicazione ai sistemi programmati secondo l'intelligenza artificiale di modelli interpretativi che sono stati costruiti *ad hoc* per gli esseri umani; e dall'altro lato, quella di voler spiegare il comportamento delle entità intelligenti attribuendo loro gli stati mentali tipici dei meccanismi umani.

Queste aspirazioni aggravano la posizione dei giuristi, ai quali non rimane altro che eliminare completamente la componente psicologica, o duplicare la caratterizzazione dei fatti rilevanti, adottando una logica puramente comportamentalista. Secondo Sartor invece sarebbe possibile dare una lettura dei concetti mentali «in modo flessibile e neutrale. Cosicché essi divengano applicabili ad alcuni tipi di enti artificiali. Ciò consentirebbe di conservare tanto la spiritualità quanto l'unitarietà del diritto, pur in una società sempre più caratterizzata dall'elaborazione automatica dell'informazione»⁶⁸.

⁶⁸ Sartor, 2003, 24.

I tentativi di superare questa forte aderenza all'elemento psicologico e mentale sono stati numerosi⁶⁹; in particolare, Sartor approva la posizione di Dennett il quale ritiene che il modo degli esseri umani di interpretare e prevedere il comportamento degli enti con i quali si interagisce, non dipende tanto dalla natura degli enti stessi, ma dalla prospettiva con la quale ci si appropria ad essi: la prospettiva fisica (*physical stance*), la prospettiva progettuale (*design stance*) e la prospettiva intenzionale (*intentional stance*)⁷⁰.

6.1. Analisi delle tre prospettive

In ordine alla prima delle tre prospettive individuate a conclusione del paragrafo precedente, Daniel Dennett ritiene che:

[I]f you want to predict the behavior of a system, determine its physical constitution [...] and the physical nature of the impingements upon it, and use your knowledge of the laws of physics to predict the outcome of any input. This is the grand and impractical strategy of Laplace for predicting the entire future of everything in the universe; but it has more modest, local, actually, usable versions. The chemist or physicist in the laboratory can use this strategy to predict the behavior of exotic materials, but equally the cook in the kitchen can predict the effect of leaving the pot on the burner too long⁷¹.

In altre parole, quando si adotta la prospettiva fisica, non si fa altro che spiegare il comportamento di un qualsiasi soggetto (animali, persone, entità artificiali) in base al suo stato fisico e alle leggi di natura ad esso applicabili. Ad esempio, si può spiegare il motivo per il quale il lanciatore del martello in atletica non perde il controllo sullo strumento durante la fase iniziale di rotazione, antecedente al lancio. Il filo che collega la sfera metallica alla impugnatura, mentre la sfera gira su sé stessa, esercita su di essa una forza diretta in ogni istante verso il centro, ossia la forza centripeta che è funzionale a mantenere un oggetto in movimento circolare uniforme. La sfera tende per

⁶⁹ Numerose sono le teorie che si susseguono da decenni accendendo il dibattito in questione. Solo per fare qualche riferimento, per un eventuale approfondimento, si pensi a coloro che ritengono che il cervello sia una "macchina semantica", ossia uno strumento di calcolo che possa essere accostato ad un calcolatore elettronico: Haugeland, 1981, 34-50. Altri optano per il cd. «materialismo eliminativo», il quale prevede che i concetti di psicologia comuni (come quelli di volontà, credenza, coscienza), quali rappresentazione concreta di una realtà ideale, siano ingannevoli e siano destinati ad essere sostituiti da concetti elaborati dalla neuroscienza pura: Churchland, 1981, 67-90. Vi sono anche coloro che riconoscono l'intenzionalità, intesa come la capacità di produrre un significato, anche a sistemi biologici o meccanici: Dennett, 1988, 180-202; Dennett, in Beckermann, Mclaughlin, Walter, 2009, 1-22.

⁷⁰ Sartor, *ult. op. cit.*, 25. Dennett, 1996, 28 e succ.

⁷¹ Dennett, in Haugeland, 1997, 60.

inerzia a muoversi in linea retta lungo una tangente immaginaria, ma l'atleta facendola rotare la tira verso di sé obbligandola a modificare la sua direzione; la forza centripeta serve per mantenere la sfera su un'orbita centrale, impedendo che sfugga verso la tangente, cosa che accade non appena l'atleta lascia andare lo strumento.

Questa prospettiva può essere efficacemente utilizzata solo da coloro che conoscono precisamente le leggi della fisica, altrimenti non si sarà in grado di comprendere il movimento e le dinamiche degli oggetti osservati.

Guardando un oggetto attraverso la seconda prospettiva, quella progettuale, lo si considera come funzionale al raggiungimento dello scopo per il quale è stato creato:

*Where one ignores the actual [...] details of the physical constitution of an object, and, on the assumption that it has a certain design, predicts that it will behave as it is designed to behave under various circumstances. For instance, most users of computers have not the foggiest idea what physical principles are responsible for the computer's highly reliable, and hence predictable, behavior. But if they have a good idea of what the computer is designed to do [...], they can predict its behavior with great accuracy and reliability, [...]*⁷².

Dal momento che gli esseri umani sono consapevoli che premendo l'interruttore della luce questa si accende illuminando la stanza, compiendo questa azione è normale aspettarsi un simile risultato, a prescindere dalla conoscenza del meccanismo fisico/meccanico che causa questa determinata conseguenza. In altre parole, conoscendo il fine per il quale un oggetto è stato progettato ci si aspetta che «il suo funzionamento corrisponda al progetto che [si assume] essere in esso incorporato»⁷³.

È anche vero però che questa corrispondenza tra il progetto e la funzione attesa potrebbe essere messa in crisi da un errore del programmatore, o dell'esecutore del progetto, tale per cui il comportamento dell'oggetto sia diverso da quello che si voleva ottenere⁷⁴. Se ne conclude, quindi, che la prospettiva progettuale non sia infallibile, anzi a volte per approfondire la fonte della diversità di atteggiamento, è necessario procedere per livelli integrativi: si scende più in profondità rispetto al comportamento inatteso rilevato per spiegare l'eccezione verificatasi. Nello specifico, le strade che possono

⁷² Dennett, *ult. op. cit.*, 60-61.

⁷³ Sartor, *ult. op. cit.*, 26.

⁷⁴ Potrebbe trattarsi di un errore originario, l'algoritmo potrebbe essere stato erroneamente programmato tanto da dare un risultato errato di fronte a specifiche combinazioni; oppure potrebbe trattarsi di una «degenerazione successiva» come, ad esempio, un intenso flusso di energia elettrica, che conduce ad una variazione del comportamento inizialmente atteso. Sartor, *ult. op. cit.*, 27.

essere intraprese sono due: da un lato si può continuare ad applicare la prospettiva progettuale alle singole componenti dell'oggetto preso in considerazione, e quindi a fronte di un errore di programmazione dell'algoritmo operativo si andranno a verificare le istruzioni impresse nel sistema, i dati, e così via; d'altro lato, invece, si potrebbe provare a spiegare il diverso comportamento assunto sfruttando la prospettiva fisica, e magari adducendo che a causa dell'aumento incontrollato della tensione elettrica si è verificato uno scompenso al circuito elettronico del sistema, provocandone un malfunzionamento.

Qualche volta potrebbe capitare che la prospettiva di progettazione non sia accessibile, in questo caso Dennett offre l'alternativa della terza prospettiva, che viene così spiegata da Sartor:

Quando guardiamo un ente dalla prospettiva intenzionale, spieghiamo il comportamento di quell'ente ipotizzando che l'ente possieda determinati stati mentali. Tipicamente si tratterà dell'ipotesi che l'ente in questione stia cercando di raggiungere determinati scopi (abbia desideri o, se si preferisce, "stati conativi"), abbia determinate credenze (informazioni o, se si preferisce, stati epistemici), determini, sulla base di quelle credenze, i comportamenti più idonei a raggiungere i propri scopi (razionalità dei mezzi), e proceda quindi ad adottare quei comportamenti⁷⁵.

La prospettiva intenzionale è quella che viene normalmente usata per spiegare le azioni umane, in particolare per comprendere e prevedere i comportamenti altrui: si attaccherà il caricabatterie dello *smartphone* alla presa della corrente per evitare che si spenga se è scarico; l'idraulico chiuderà l'acqua prima di lavorare sul sistema idraulico della casa, per evitare perdite eccessive in corso d'opera; al gioco degli scacchi si attenderà la mossa dell'avversario per valutare quale sia la risposta più adeguata, nel tentativo di arrivare a fare scacco matto; e così via. Ma se l'avversario al gioco degli scacchi non fosse un essere umano, ma un *software*, si adotterà la stessa strategia interpretativa predittiva, utilizzando la prospettiva intenzionale? Una situazione di questo tipo si è concretamente proposta, quando nel 1997 la IBM ha creato

⁷⁵ Sartor, *ult. op. cit.*, 30; Dennett, in Haugeland, 1997, 61, a questo proposito le parole del padre fondatore dell'*Intentional stance* sono: «(...) first you decide to treat the object whose behavior is to be predicted as a rational agent; then you figure out what beliefs that agent ought to have, given its place in the world and its purpose. Then you figure out what desires it ought to have, on the same considerations, and finally you predict that this rational agent will act to further its goals in the light of its beliefs. A little practical reasoning from the chosen set of beliefs and desires will in many instances yield a decision about what the agent ought to do; that is what you predict the agent will do.»

il sistema intelligente denominato *Deep Blue*, con l'obiettivo di sfidare il campione mondiale di scacchi del momento.

Passando in rassegna le possibili prospettive che possono essere adottate nel rapportarsi con un simile sistema intelligente, è evidente come si debba immediatamente escludere la prospettiva fisica, in quanto si dovrebbe essere dotati di un bagaglio di conoscenze tecnico-scientifiche molto elevato; si tenga altrettanto presente che questa prospettiva non è utilizzabile neppure rapportandosi con un essere umano, dal momento che sarebbe come pretendere di prevedere il comportamento dell'avversario analizzando i processi fisici, chimici, biologici delle cellule del cervello. Nemmeno la prospettiva progettuale è adatta al caso concreto, infatti, sarebbe difficile riuscire a discostarsi dall'ipotesi che il progettista del software *Deep Blue* abbia voluto realizzare un sistema capace di giocare a scacchi; anche qui si evidenzia l'impossibilità di utilizzare questa prospettiva con un avversario umano, perché sarebbe come cercare di interpretare il suo comportamento sulla base delle funzioni che le componenti del cervello umano sono deputate a svolgere. Si giunge così a prendere in considerazione la prospettiva intenzionale, la quale permette di interpretare e di prevedere il comportamento sia di *Deep Blue* che di un essere umano: attribuire degli stati intenzionali all'avversario, in termini di credenze, intenzioni o conoscenze, consente di adottare una strategia di gioco calibrata sulla base degli stati che si crede che il rivale stia perseguendo.

Più precisamente, per spiegare cosa si intenda per “stati intenzionali”, partendo dal presupposto che si devono identificare in senso oggettivo, si può fare riferimento a due proposte: la prima, quella “comportamentista”, ritiene che gli stati esisterebbero solo nello sguardo di chi osserva; quindi, affermare che un ente ha stati intenzionali, vuol dire ritenere che il suo comportamento sia interpretabile e prevedibile tramite la prospettiva intenzionale. Dal punto di vista pratico ciò significa che «affinché risulti vero che il mio programma scacchistico “vuole” mangiarmi la torre, è sufficiente che l'iscrizione di questa volontà al mio programma mi consenta di prevederne con successo il comportamento (di anticipare i successivi tentativi di insidiare la mia torre)»⁷⁶. La seconda proposta, invece, che può essere definita «realistica», si fonda sulla considerazione che gli stati intenzionali facciano riferimento a specifiche caratteristiche dell'ente al quale vengono attribuiti. Quindi per appurare che un ente possieda questi stati ci si dovrebbe interrogare «sulla sua struttura interna, o meglio, sul suo modo di funzionare: ci dobbiamo chiedere se ci sono stati interni dell'ente che rappresentano informazioni relative all'ambiente (credenze), stati interni che rappresentano obiettivi che

⁷⁶ Sartor, *ult. op. cit.*, 34.

l'ente tende a raggiungere (scopi), e modi nei quali l'ente adegua il proprio comportamento alle informazioni e agli obiettivi in esso rappresentati (scelte razionali)»⁷⁷.

Uno stato interno dell'ente rappresenta un obiettivo, uno scopo, e nel momento in cui l'ente possiede questo stato si attiva per raggiungerlo; analogamente, ciò condurrà un essere umano (anche dotato di un bagaglio di conoscenze basilari) a ricercare del cibo, nel momento in cui in lui si attivano le reazioni fisico-chimiche della fame, tali da spingerlo al raggiungimento dell'obiettivo ossia alla soddisfazione di questa sua esigenza. Lo stesso può dirsi con riferimento ad un sistema informatico, il quale sia stato programmato per compiere una determinata azione: fintanto che l'azione non sia compiuta il sistema continuerà ad attivarsi. A questo punto, Sartor ritiene che «si potrà affermare che il sistema “ha voluto” adottare certi comportamenti, se il sistema ha tenuto quei comportamenti in seguito ad un procedimento interno al sistema stesso, procedimento inteso a fare in modo che il sistema realizzasse i propri obiettivi, sulla base delle proprie credenze»⁷⁸.

6.2. Considerazioni conclusive in merito alla soggettività intenzionale

Riconoscere l'intenzionalità ai sistemi informatici comporterebbe due conseguenze: *in primis*, interagendo con un sistema intelligente si potrebbe, in virtù del principio della tutela dell'affidamento, «attribuire al sistema stesso gli stati psicologici che esso appaia possedere, e quindi in particolare gli stati che esso dichiara di possedere, o che siano presupposti dagli atti linguistici da esso compiuti»⁷⁹. Pertanto, il titolare del sistema intelligente poi non potrà sottrarsi, nei confronti della controparte, alle conseguenze prodotte dalla attribuzione di stati intenzionali al sistema: da questo punto di vista, se un sistema informatico compie un atto che possa essere configurato come una dichiarazione di volontà, ad esempio concludendo un negozio giuridico, si potrà ritenere che creda in quanto abbia dichiarato, e il titolare del sistema non può negare gli effetti derivati dal negozio adducendo una mancanza della sua volontà.

In secundis, quale seconda conseguenza, chiunque si interfacci con un sistema informatico deve accettare le interpretazioni del comportamento del sistema che «corrispondano al contenuto degli stati intenzionali effettivamente posseduti dal sistema, e inoltre siano ascrivibili ad esso in base alle convenzioni applicabili all'interazione di cui si tratta»⁸⁰. In altre parole, si-

⁷⁷ Sartor, *ult. op. cit.*, 34.

⁷⁸ Sartor, *ult. op. cit.*, 36.

⁷⁹ Sartor, *ult. op. cit.*, 40.

⁸⁰ Sartor, *ult. op. cit.*, 41.

gnifica che l'interlocutore non potrà sottrarsi agli effetti del negozio giuridico che è stato concluso dal sistema, in relazione agli stati intenzionali che sono annoverati in quel comportamento e al contenuto che al negozio stesso viene riconosciuto convenzionalmente.

Concentrandosi, a questo punto, su una discussione puramente giusfilosofica, l'intenzionalità che si vorrebbe ascrivere alla macchina per riconoscerle una forma di soggettività è sicuramente una intenzionalità derivata⁸¹; in questo caso, in particolare, il sistema programmato secondo l'intelligenza artificiale *significa* qualcosa solo perché sono gli esseri umani a dirgli cosa deve significare⁸². Se così fosse allora si potrebbe dire che i fenomeni mentali, *rectius* gli stati mentali, sono dotati di intenzionalità, ma solo agli occhi degli osservatori che sono proiettati verso l'attribuzione di una interpretazione mentalistica al fenomeno. Dennett, al proposito, parla di intenzionalità "come se", in forza della quale un ente sarebbe dotato di intenzionalità solo perché coloro che ne entrano in contatto lo trattano come se la possedesse. La svolta di quella che può essere chiamata *Teoria dell'intenzionalità* di Dennett, tuttavia, risiede nel fatto che, a differenza dei funzionalisti (tra i quali Searle), egli sostiene che si giunga a tale conclusione applicando «le stesse morali, le stesse regole pratiche di interpretazione al caso dell'uomo»⁸³. Difatti, tanto quanto quella dei sistemi programmati secondo l'intelligenza artificiale, anche l'intenzionalità degli esseri umani è derivata, in quanto sarebbe la risultante dell'applicazione della teoria dell'evoluzione della specie di Darwin. Ciò significa che non vi è Madre Natura a possedere una intenzionalità originaria o intrinseca, che poi viene attribuita agli esseri umani, ma ciò da cui deriva l'intenzionalità è proprio il millenario processo di sviluppo e selezione naturale; pertanto, per usare le parole di Dennett: «la nostra intenzionalità è derivata dall'intenzionalità dei nostri geni "egoisti"! »⁸⁴.

Alla luce di questa lettura la coscienza, ed in particolar modo l'intenzionalità, non potrebbero essere considerate come qualità interiori ed inaccessibili se non soggettivamente, come era proprio del dualismo cartesiano⁸⁵, ma piuttosto come qualità indagabili e sperimentabili oggettivamente. Quest'ul-

⁸¹ Solo per completezza si precisi sin da ora che la intenzionalità derivata si contrapporrebbe a quella intrinseca. Si tratta di una bipartizione argomentata da Searle, e sostenuta dall'esperimento della stanza cinese, la quale viene utilizzata per argomentare che non si potrà mai avere una IA forte, in quanto la sua intenzionalità sarà sempre frutto di derivazione di progettazione o realizzazione da parte di un agente con intenzionalità intrinseca, come appunto l'essere umano.

⁸² Fini, Milano, 2002, 102.

⁸³ Dennett, 1992, 394.

⁸⁴ Dennett, *ult. op. cit.*, 399-400.

⁸⁵ La criticità inerente al dualismo tra mente e corpo è risalente agli studi di Cartesio, il quale separava la sostanza pensante, *res cogitans*, dalla sostanza estesa, *res extensa*.

timo passaggio segna il punto cruciale per poter quindi avviare un'indagine più approfondita sulla possibilità di strutturare una soggettività intenzionale ai sistemi intelligenti.

In conclusione, ci si chiede: risulta più adeguato adottare un sistema in grado di oggettivizzare le relazioni uomo-macchina, per meglio permettere una convivenza tra i due soggetti; o, la strada migliore è quella di «spiritualizzare i rapporti con e tra le macchine, imponendo anche ad essi un'interpretazione intenzionale»⁸⁶?

6.3. L'intenzionalità tecnologica nella prospettiva di Benanti

Uno dei padri della filosofia della scienza e della tecnologia, Don Ihde, già nel 1990 parlava di «intenzionalità tecnologica»⁸⁷, per rilevare il fatto che un artefatto tecnologico sia in qualche misura in grado di imprimere una direzione o uno scopo a chi agisce mediante esso⁸⁸: «*[P]henomenologically speaking, it would be more adequate to localize “technological intentionality” not in technologies themselves but in the relation, technologies co-shape “intentionality”; they facilitate specific relations between humans and world, and in so doing play a role in the constitution of both*»⁸⁹.

Ihde ritiene che ogni utilizzatore di un artefatto sia assoggettato ad una serie di suggestioni e di condizionamenti culturali che influenzano grandemente la sua responsabilità. A sua volta, Paolo Benanti osserva che questo orientamento ci obbliga a considerare non solo il rapporto stretto tra l'agente e lo strumento, ma anche e soprattutto un'analisi approfondita sul contesto dell'azione⁹⁰.

Per meglio comprendere questa presa di posizione l'esempio seguente sarà sicuramente chiarificatore: si pensi ad un'arma, come potrebbe essere un fucile. Questo oggetto può essere considerato allo stesso tempo come uno strumento di difesa e di offesa, e ciò dipende dall'esercizio di una libera e consapevole scelta dell'utilizzatore, in altre parole si potrebbe dire che il fucile è a disposizione di chi lo impugna. Tuttavia, il fucile non è solo espressione dello sviluppo tecnico scientifico nell'ambito bellico e della scelta morale dell'utilizzatore, ma ad esso si potrebbe guardare anche come ad uno strumento perfettamente in grado di forgiare la consapevolezza e la responsabilità di colui che lo utilizza. «Chi imbraccia un fucile – precisa Ihde – guarda il mondo attraverso il suo mirino e percepisce la realtà nella prospettiva che questa

⁸⁶ Sartor, *ult. op. cit.*, 42.

⁸⁷ Ihde, 1990.

⁸⁸ Benanti, 2016.

⁸⁹ Verbeek, 2001, 145.

⁹⁰ Benanti, *ult. op. cit.*

fornisce»⁹¹, questa affermazione potrebbe suggerire quindi che il fucile, per il sol fatto di essere stato progettato e creato come uno strumento di offesa, racchiude in sé le aspettative dell'utilizzatore che è portato però a vedere gli altri come nemici dai quali difendersi. Il fucile ha intrinseca in sé la natura di "arma di offesa" che lo qualifica e che innegabilmente ne suggerisce le modalità di utilizzo.

Analogamente si può dire lo stesso prendendo in considerazione un'automobile: l'autista utilizza il mezzo di trasporto per recarsi da un punto A ad un punto B, imprimendo su di esso una volontà che si potrebbe definire pura. Ma proprio per il fatto che si tratta di un'automobile, l'autista è costretto ad assumere specifici comportamenti, che uniti all'ambiente contemporaneo contribuiscono a determinare la consapevolezza e la responsabilità dell'autista.

In altre parole, l'intenzionalità tecnologica consiste nel fatto che le modalità di utilizzo di un artefatto dipendano non solo dalla intenzionalità umana ma anche dal contesto storico e culturale in cui si è inseriti; nell'esempio di prima, un'automobile non è un semplice strumento per raggiungere una destinazione: in quanto artefatto tecnologico richiede una stretta relazione con l'utilizzatore e anche con il mondo circostante, determinandone il modo di essere.

Le considerazioni di Ihde e di Benanti qui enunciate non si discostano troppo dalle conclusioni alle quali è giunto Sartor, infatti, a ben vedere in entrambi i casi l'attribuzione di una intenzionalità al sistema intelligente dipende anche dall'accettazione del comportamento e della funzione che convenzionalmente connatura il sistema stesso.

Evitare di attribuire innanzitutto una soggettività, e poi anche una responsabilità, alle macchine intelligenti, solo ed esclusivamente perché ritenute prive di intenzionalità, non regge il paragone con l'intenzionalità umana, laddove quest'ultima risulti essere in ogni caso influenzata dal contesto storico e culturale. Infatti, in nessuno dei due casi si potrà parlare di intenzionalità pura, quindi, perché ricercarla nelle macchine? Pretendere di continuare a procedere per analogia non fa altro che allontanare dall'obiettivo, perché non permette di avere un'apertura mentale che da un foglio bianco conduca alla scrittura di un nuovo modo di pensare.

7. La soggettività funzionale e la teoria di Teubner

Alla luce della trattazione effettuata in precedenza si evince che un *robot* non possa essere considerato un soggetto in senso ontologico: pertanto, ci

⁹¹ Ihde, *ult. op. cit.*

si chiede se possa risultare utile che il diritto sfrutti una delle sue facoltà e attribuisca a questi enti non umani una soggettività, indipendentemente dalla presenza dei caratteri ontologici prima descritti e solo in virtù di un più ampio discorso di opportunità.

A tal proposito, Gustav Teubner definisce tale soggettività “ascritticia”, intendendo una *species* di soggettività la quale «viene conferita attraverso atti formali da una autorità secondo criteri pratico- funzionali»⁹². Il diritto, infatti, è libero di attribuire una soggettività, anche solo motivato dall’utilità o dall’opportunità, a prescindere dalla presenza dei caratteri propri della soggettività ontologica; questo è quello che è accaduto, storicamente, per le associazioni non riconosciute o per le società, le quali di certo non possono essere trattate alla stregua di soggetti ontologici⁹³. Si tratta, pertanto, di una vera e propria creazione dell’ordinamento utile per garantire la regolamentazione del regime di responsabilità, in termini prettamente risarcitori (quale prima grande spinta che ha mosso i giuristi a ricercare una disciplina della soggettività), ma allo stesso tempo permette di mantenere saldo il rifiuto del riconoscimento di una soggettività naturale ai *robot*, i quali non saranno mai considerati alla stregua di persone in senso ontologico.

Questa teoria si basa sul postulato che solo gli esseri umani possono essere considerati persone, e si concede l’attribuzione, grazie all’espedito della *finzione*, di una soggettività giuridica a sistemi intelligenti solo ed esclusivamente al fine di soddisfare esigenze pratiche, funzionali e soprattutto giuridiche⁹⁴: si permette che il diritto «formalmente attribui[sca], in termini di *fictio iuris*, lo *status* di soggetto a una entità non umana in base a esigenze o criteri per lo più funzionali»⁹⁵.

Si giunge, quindi, alla conclusione che allo stato attuale non si può riconoscere una piena soggettività giuridica agli agenti robotici, i quali possono solo essere considerati «schiavi digitali»: nel loro agire autonomo non può certo ravvedersi il perseguimento di un interesse proprio, ma solo ed esclusivamente di interessi di soggetti terzi, persone fisiche o giuridiche che siano⁹⁶. Si osserva infatti come, per ora, si possa solo approfondire la stretta relazione tra uomo e macchina che si sta sempre più intensificando: la tendenza non dovrebbe essere quella di dar luogo a macchine intelligenti considerate nella

⁹² Taddei Elmi, Romano, 2010, 146, riferimento alla nota numero 6.

⁹³ Ma ancor prima si pensi al *servus* romano, ossia la persona ridotta in stato di schiavitù, che era considerato una *res*, ma se liberato tornava soggetto.

⁹⁴ Berti Suman, in Alpa (a cura di), 2020, 265.

⁹⁵ Troncarelli, 2018, 109.

⁹⁶ Teubner, 2019, 31; sul punto si veda anche il contributo di Schirmer, 2016, 665.

loro singolarità, ma piuttosto quella di intervenire sull'«agire congiunto di un ibrido uomo-macchina o di una complessiva interazione tra reti»⁹⁷.

Di fronte ad una simile prospettiva non appare, appunto, per nulla appropriata l'attribuzione di una soggettività piena ai sistemi animati da intelligenza artificiale, la quale non farebbe altro che «incrementare la tendenza a svilire il contributo dell'azione dell'uomo entro la complessa compagine di interazioni uomo-macchina e offr[ire] una rappresentazione che trascura il momento della padronanza dell'uomo sul suo agire, quale potenziale di responsabilità»⁹⁸. La costruzione giuridica alla quale puntare, per garantire una tutela a tutto tondo, sarebbe quella di uno «*status* giuridico (...) determinato funzionalmente, muovendo dal ruolo servente assunto dagli agenti *software* nelle interazioni uomo-macchina di assistenza digitale, ruolo per il quale, una volta che sia così definito, non è necessaria una capacità giuridica generale»⁹⁹.

In questo contesto l'esempio della soggettività ascrittiva, quale forma di soggettività parziale¹⁰⁰ costruita *ad hoc* per tutelare la società giuridica dal rischio di autonomia, di associazione e da interconnessione¹⁰¹, introduce una suggestione di grande valore nella trattazione qui presa in considerazione, dal momento che segna allo stesso tempo una battuta di arresto rispetto all'affannosa ricerca di ricondurre alle categorie esistenti la disciplina del fenomeno dell'intelligenza artificiale, e un vero e proprio cambiamento di metodo per porre rimedio al problema della personalità degli agenti *software*. Il sistema giuridico italiano spesso costringe ad una rigida categorizzazione, la quale però conduce ad una irrinunciabile precisione e specificazione, che non sono sempre raggiungibili, come nel caso in esame. Staccarsi dalla inflessibilità di questi schemi concettuali permette di indagare più a fondo l'utilità degli istituti giuridici ed il loro concreto modo di operare.

Doveroso è tener presente che l'ordinamento italiano, sebbene sia in costante divenire, abbia posto l'individuo al centro della trattazione giuridica. Per oltrepassare una simile concezione antropocentrica, che appare debitrice della cultura razionalistica dell'epoca moderna e che sembra inadeguata a disciplinare complesse relazioni della società tecnologica e postindustriale,

⁹⁷ Teubner, *ult. op. cit.*, 32.

⁹⁸ Teubner, *ult. op. cit.*, 32-33.

⁹⁹ Teubner, *ult. op. cit.*, 32. In particolare si veda la nota numero 36, la quale recupera il contributo di Gruber, 2015, 198.

¹⁰⁰ Teubner, *ult. op. cit.*, 33.

¹⁰¹ Al proposito doveroso è il rimando alla Proposta di Regolamento del Parlamento Europeo e del Consiglio, *Che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) e modifica alcuni atti legislativi dell'unione*, la quale si concentra molto sulla disciplina dei sistemi di intelligenza artificiale "ad alto rischio", in particolare si veda il Titolo III e successivi della Proposta.

introducendo quale ulteriore perno di riferimento un sistema intelligente, è necessario un dialogo contemporaneo che incida anche sulla metodologia di trattazione. In questa prospettiva, non può certo dubitarsi che il riconoscimento di una soggettività e di una personalità giuridica a sistemi intelligenti non siano finalizzati a tutelare la macchina, ma piuttosto l'essere umano rispetto ad essa¹⁰².

Non a caso Teubner ritiene che alla soggettività si dovrebbe guardare come ad uno strumento utile a perseguire specifici obiettivi, tra i quali la tutela dei cd. rischi da responsabilità:

1. Il rischio di autonomia, che discende dalle “decisioni” indipendenti che l'agente software assume;
2. Il rischio di associazione, che dipende dalla stretta cooperazione tra uomo e agente software;
3. Il rischio di interconnessione, che consiste nell'agire non isolato, ma in stretta interdipendenza di una pluralità di computer in rete¹⁰³.

7.1. Rischio di autonomia e soggettività giuridica limitata: introduzione alla teoria attore-rete

Il rischio di autonomia appare molto penetrante e trae la sua origine dalla condotta, tendenzialmente non prevedibile, degli enti programmati secondo l'intelligenza artificiale, in particolare governati da algoritmi con autoapprendimento. Diversi sono gli autori¹⁰⁴ che sostengono che un simile rischio si concretizzi nella teoria attore-rete, o *actor-network theory*.

La Teoria Attore Rete (ANT) è un modello teorico che può essere collocato all'interno delle categorie di studio della scienza, della sociologia, della tecnologia, e si presenta in numerose aree applicative. I suoi ideatori, i sociologi francesi Bruno Latour e Michel Callon¹⁰⁵, collaborarono con l'obiettivo non solo di ridefinire i rapporti tra scienza, tecnologia e società, ma soprattutto con il fine di dare nuova vita al concetto di società.

¹⁰² Teubner, 2006, 516. «*The result ist that law is opening itself for the entry of new juridical actors – animals and electronic agents. The differences in the outcomes, however, are striking. (...) Animal rights and similar constructs create basically defensive institutions. Paradoxically they incorporate animals in human society in order to create defences against destructive tendencies of human society against animals. (...) For electronic agents, the exact opposite is true. Their legal personification, especially in economic and technological context, creates aggressive new action centers as basic productive institutions. Here, their inclusion into society does not protect the new actors, just the oppite, it is society that needs to defend itself against the new actor*».

¹⁰³ Teubner, 2019, 37.

¹⁰⁴ Law, 1992.

¹⁰⁵ I due autori furono i primi a parlare della teoria attore-rete nel seguente contributo: Callon, Latour, in Knorr-Cetina, Cicourel, 1981, 277-303. Per approcciarsi all'argomento è utile anche il contributo: Zampieri, 2015, 61-88.

Uno dei pilastri portanti di questa teoria è la considerazione della conoscenza, non tanto come frutto di una metodologia scientifica, ma piuttosto come un prodotto sociale. Nello specifico, la conoscenza viene studiata come un insieme di singole componenti, che di fatto godono di una vita autonoma, ma che se inserite in una rete di elementi, sono in grado di comunicare superando la loro individualità¹⁰⁶. Al proposito è interessante la spiegazione offerta dall'antropologo britannico Law, il quale precisa:

“Knowledge”, then, is embodied in a variety of material forms. But where does it come from? The actor-network answer is that it is the end product of a lot of hard work in which heterogeneous bits and pieces--test tubes, reagents, organisms, skilled hands, scanning electron microscopes, radiation monitors, other scientists, articles, computer terminals, and all the rest--that would like to make off on their own are juxtaposed into a patterned network which overcomes their resistance¹⁰⁷.

La proposta di Law sarebbe poi quella di adattare tali considerazioni, valide per la scienza, anche alla famiglia, alle organizzazioni, ai sistemi computerizzati, all'economia e alla tecnologia, dal momento che in ognuno di questi casi si può intravedere una rete ordinata di materiali eterogenei, la cui resistenza (intesa nel senso di autonomia ed individualità del materiale) è stata superata. Ecco allora che gli autori della ANT non fanno altro che avanzare la suggestione secondo la quale «*the social is nothing other than patterned networks of heterogeneous materials*»¹⁰⁸.

Si tratta chiaramente di una proposta rivoluzionaria e radicale, la quale sottende che queste reti sono composte, non solo da persone, ma anche da animali, macchine, ed ogni altra entità artificiale; la conclusione alla quale si giunge è che gli esseri umani creano reti sociali, non solo perché si relazionano reciprocamente, ma anche perché creano collegamenti e legami con le altre entità materiali.

Una simile presa di posizione getta le basi per un mutamento del concetto di *società*, intesa come «insieme di uomini organizzato sulla base di un sistema più o meno strutturato di rapporti naturali, economici, culturali, politici; in tale accezione il termine è per lo più accompagnato da attributi che ne

¹⁰⁶ Nello specifico, Law, *ult. op. cit.*, 381, afferma: «*With others in the sociology of science, they argued that knowledge is a social product rather than something generated through the operation of a privileged scientific method. And, in particular, they argued that “knowledge” (but they generalize from knowledge to agents, social institutions, machines, and organizations) may be seen as a product or an effect of a network of heterogeneous materials.*»

¹⁰⁷ Law, *ult. op. cit.*, 381.

¹⁰⁸ Law, *ult. op. cit.*, 381.

specificano meglio il significato sulla base di criteri diversi»¹⁰⁹, ed anzi, proprio Latour propone di non parlare più *society*, ma di *collective*:

*Unlike society, which is an artifact imposed by the modernist settlement, [the concept of collectives] refers to associations of humans and nonhumans associations of humans and no-humans. While a division between nature and society renders invisible the political process by which the cosmos is collected in one livable whole, the word “collective” makes this process central*¹¹⁰.

Si presenta, in queste poche parole, da un lato una forte esigenza di distacco dalla concezione antropocentrica del mondo, delle scienze, della tecnologia e anche del diritto; e dall'altro lato un importante tentativo di allontanarsi dalle convenzioni sociali e civili proprie della nostra epoca. Se queste sono le premesse, appare chiaro il motivo per il quale questa teoria, nonostante l'impegno costante di sociologi, filosofi e antropologi, sia ancora circoscritta da una cortina di incertezza rispetto al suo accoglimento. Ciononostante, numerosi sono stati gli autori che si sono serviti della ANT per le loro analisi argomentative.

7.1.1. La teoria attore-rete nella soggettività degli agenti *software*

Date queste prime considerazioni, la logica teorica e metodologica della teoria attore-rete¹¹¹ sfalda la distinzione tra soggetto ed oggetto: l'azione è il risultato dell'interazione tra soggetti umani e oggetti in movimento connessi in rete. In particolare, all'interno di questa rete, i protagonisti umani e non, vengono definiti “attanti”, proprio per sottolineare il loro non essere solo ed esclusivamente “attori” legati alla dimensione umana. Dal loro relazionarsi reciprocamente attraverso le connessioni di rete nascono degli *ibridi*: un essere umano alla guida di un'automobile non deve essere semplicemente considerato come un soggetto, ma va considerato come un tutt'uno con la macchina e la strada percorsa, al quale imputare il comportamento in continua evoluzione.

L'applicazione della teoria attore-rete all'argomento qui trattato relativo alla soggettività degli agenti intelligenti, porta a considerare i soggetti non umani, che operano nel mondo del diritto, come «attori collettivi»¹¹², quindi

¹⁰⁹ Definizione tratta dal Vocabolario Treccani Online, consultabile al link: <https://www.treccani.it/vocabolario/societa/>.

¹¹⁰ Latour, 1999, 304, così come citato in Bryant, 2011.

¹¹¹ Questa teoria ha iniziato ad essere sviluppata da Callon, Latour, in Knorr-Cetina, Cicourel, 1981, 277-303; ed è poi stata compiutamente sviluppata in Latour, 2005.

¹¹² Per il concetto di «attore collettivo» in campo sociologico cfr. Luhmann, 1987, in particolare, i capitoli 3 e 4. In merito a questo tema si veda anche: Teubner, 1987, 61-85; Sprenger, 2017, 108-122.

non come un insieme di singoli individui, ma piuttosto come una catena di comunicazioni; in questa accezione, quindi, un ente o un'organizzazione, ai quali normalmente l'ordinamento giuridico italiano riconosce personalità giuridica, non sono considerati meri insiemi di strutture, risorse ed esseri umani, ma vere e proprie catene di decisioni. Ciò, a patto che la catena di comunicazioni sia autodescrittiva, in quanto capace di comunicare su sé stessa, e che a questa autodescrizione siano attribuiti dei significati che possano essere compresi dagli agenti.

Lo stesso discorso può essere fatto per gli agenti *software*: nel momento in cui nella relazione comunicativa tra uomo e *computer*, si individuano degli eventi che possono essere «compresi» come «comunicazioni» dell'algoritmo (che per definizione è proprio un flusso di informazioni) che possano trasmettere delle informazioni, allora tra uomo e *computer* si sarà instaurato un vero e proprio sistema sociale. Infatti, le risposte che giungono agli esseri umani dagli agenti *software* soddisfano comunque i tre passaggi essenziali della comunicazione: emissione, informazione e comprensione¹¹³; non a caso, Esposito sostiene che:

Algorithms are social agents. Their presence and role are now central and indispensable in many sectors, both as tools to do things (as machines) and as communicative partners. Algorithms are involved in communication not only on the web where the active role of bots is now taken for granted, but also (explicitly or not) in more traditional forms, a sprint communication and even voice communication¹¹⁴.

Attualmente si sta iniziando a guardare alle macchine intelligenti, non solo come ad efficienti algoritmi che imparano ad operare come la mente umana, ma anche e soprattutto quali forme di comunicazione. In questi termini si evidenzia però, come inevitabilmente la comunicazione tra uomini e algoritmi non possa presentarsi come simmetrica, per tre motivi: agli algoritmi non si possono riconoscere le stesse capacità psichiche ed intellettive, proprie degli esseri umani; nel legame uomo macchina la “doppia contingenza” è unidirezionale; non si può parlare di reciproca capacità di comprendersi.

Per ciò che concerne la prima distinzione, vale la pena solo sottolineare che si può sostenere che si tratti di una differenza sostanziale: da un lato le macchine si scambiano informazioni tramite operazioni matematiche ed impulsi elettrici; dall'altro lato invece gli esseri umani comunicano coscientemente solo mediante una interiorizzazione dell'informazione ricevuta¹¹⁵. Questo elemento non turba la ricostruzione del rapporto uomo-macchina

¹¹³ Luhmann, 1987. Esposito, 2017, 257.

¹¹⁴ Esposito, 2017, 249.

¹¹⁵ Esposito, 2017, 253.

che si sta delineando ai fini della soggettività parziale. Infatti, ciò che rileva al proposito è quella che Teubner definisce «irritazione comunicativa»¹¹⁶, ossia la modalità di reazione di un sistema allo stimolo esterno. Gli uomini e gli algoritmi sono due sistemi autonomi e differenti: quando interagiscono, ciascuno riceve degli stimoli dall'altro, e si attiva di conseguenza dando luogo all'evento comunicativo.

Il secondo punto di asimmetria recupera il concetto di «doppia contingenza»¹¹⁷ che caratterizza la Teoria dei Sistemi Sociali di Luhmann. Si tratta di una situazione in cui uno dei partecipanti alla relazione presa in considerazione riconosce nell'altro un soggetto che dipende dalle sue azioni comunicative: per gli esseri umani questa doppia contingenza è bidirezionale, mentre per i sistemi intelligenti è soltanto unidirezionale, dal momento che all'attuale stato dell'arte non si può ancora sapere se l'algoritmo sia in grado di comprendere la sua dipendenza dalla azione comunicativa umana¹¹⁸. Fortunatamente questa forma di doppia contingenza solo unilaterale non intralaccia una vera e propria comunicazione. Si consideri, ad esempio, il paragone con Dio offerto dall'antropologo tedesco Sprenger: durante le pratiche della preghiera o di altri rituali, infatti, si realizza comunque l'evento comunicativo nei suoi tre momenti dell'emissione, della informazione e della comprensione¹¹⁹, a condizione che si proceda alla personificazione della parte non umana, in modo da renderle imputabile l'azione.

Per finire, il terzo elemento di disarmonia riguarda la comprensione di sé, ossia se i protagonisti della relazione siano capaci di riconoscersi come entità separate ed autonome, e allo stesso tempo di comprendere l'altruità. È chiaro che l'uomo possiede questa caratteristica dell'autoreferenzialità, ma la questione per le macchine è tutt'ora aperta. Ciononostante, l'evento comunicativo si realizza pienamente e si caratterizza per il suo essere superficiale, prescindendo infatti dalla reale e completa comprensione dell'inte-

¹¹⁶ Teubner, 2019, 43.

¹¹⁷ Luhmann, *ult. op. cit.*, 148.

¹¹⁸ Esposito, *ult. op. cit.*, 261: «Algorithms are not human and do not be human. But the discriminating factor, as we saw, is not whether the utterer is a person but whether there is double contingency, which so far normally required the participation of two people». In particolare, l'autrice avanza la proposta di una «contingenza virtuale» che spiega che nel momento in cui un soggetto interagisce con una macchina intelligente, ciò che accade è identico a quello spiegato di seguito: «(...) in the interaction with robotics toys. Therefore we can speak of virtualization of contingency. The interaction is meaningful because it produces information that did not exist previously, neither for the user nor for the machine. But this contingency is the result of the duplication of the perspective of the user, who observes his own contingency from a different perspective. The observers do not duplicate, what duplicates is the perspective of the same observer. No authentic double contingency is produced as reflected (and unpredictable) contingency between two communicating parties.», 263.

¹¹⁹ Sprenger, 2017, 119.

riorità dell'uomo e della macchina. Ad un simile processo importa solo che reciprocamente uomo e macchina reagiscano con una differente emissione e informazione.

Da questa breve discussione se ne evince che, nonostante questa asimmetria intrinseca alla natura della relazione dei protagonisti, si realizza comunque un sistema comunicativo autentico. In particolare, così costruiti, i sistemi intelligenti appaiono come meri flussi di informazioni, che possono essere elevati a “persone” nel momento in cui grazie al processo comunicativo viene riconosciuta loro una identità sociale, con conseguente attribuzione di una vera e propria capacità di agire. Tale conquista, in combinato disposto con le opportune e necessarie disposizioni organizzative come quelle relative alla disciplina sulla rappresentanza, renderebbe le macchine programmate secondo l'intelligenza artificiale nuovi possibili protagonisti, e forse anche nuovi *soggetti*, della realtà giuridica e sociale.

7.1.2. Soggettività giuridica e teoria della comunicazione

La dottrina propone spesso, quale criterio determinante per l'attribuzione di una soggettività giuridica agli algoritmi, l'intelligenza o la capacità di pensare; tuttavia, in questo caso si incorre in un errore che conduce ad un vicolo chiuso: nel momento in cui si analizza la capacità giuridica non si devono indagare le qualità delle quali un agente *software* deve essere dotato per poter essere considerato attore, come l'intelligenza o la capacità riflessiva; infatti, i processi psicologici interiori non sono determinanti al fine di attribuire questo tipo di soggettività alle persone giuridiche.

Interessante al proposito è quanto affermato da Esposito: «*After all, what is interesting in the interaction with algorithms is not what happens in the machine's artificial brain, but what the machine tells its users and the consequences of this*»¹²⁰, dalle cui parole si può dedurre con Teubner che «non le qualità interiori dell'agente, ma le interazioni sociali, cui prendono parte le operazioni correnti dell'algoritmo – e cioè una sequenza chiusa e autonoma di comunicazione ricorsiva – costituiscono l'algoritmo come persona, come artefatto semantico, attore dotato di capacità comunicativa»¹²¹. Quanto precede è perfettamente coerente con l'esposizione della teoria della comunicazione effettuata nel paragrafo precedente: gli algoritmi, essendo flussi continui di dati, esistono quali attori del mondo giuridico non di per sé, ma solo se a monte vi è un sistema sociale che li costruisca come persone e li accolga al suo interno. Per di più si evidenzia nuovamente il fatto che in una “comunicazione” «non importa la capacità interiore di pensare dell'algoritmo, né la

¹²⁰ Esposito, *ult. op. cit.*, 250.

¹²¹ Teubner, 2019, 51. Riferimento specifico anche a: Luhmann, 1987; Sprenger, 2017, 108-122.

“effettiva” intelligenza artificiale (qualunque cosa significhi); importa invece la loro specifica partecipazione alla comunicazione sociale»¹²². Pertanto, è da ritenersi che risulti maggiormente decisivo l’approccio critico alla cosiddetta *artificial communication*, piuttosto che alla *artificial intelligence*¹²³, dal momento che la soggettività non dipende dalla coscienza o dal pensiero, ma proprio dalla modalità di comunicazione, ponendosi il «principio del dialogo, che giustifica la coesistenza reciproca degli esseri umani»¹²⁴ la nuova frontiera etica e giuridica dello studio dell’accoglimento di una forma di soggettività in capo ai sistemi programmati secondo l’intelligenza artificiale.

7.2. Rischio di associazione

Nel paragrafo 7 si accennava alla soggettività parziale, quale *species* di soggettività funzionale, introdotta con l’obiettivo di tutelare la società da tre rischi di responsabilità emersi con l’intervento dei sistemi intelligenti. Avendo già analizzato dettagliatamente il primo nei paragrafi precedenti, si passi ora all’osservazione del secondo rischio, ossia quello di associazione.

Considerando l’associazione uomo-macchina, si pone un duplice problema di fronte al riconoscimento di una soggettività parziale: da un lato riconoscere diritti e doveri solo in capo all’uomo, quale componente dell’associazione esaminata, significa sottovalutare il rischio digitale sotteso, giacché si finirebbe con il trascurare gravemente l’autonomia dei sistemi di intelligenza artificiale. Dall’altro lato, attribuire a questi sistemi le loro azioni, condurrebbe al recupero, per i soggetti umani, delle categorie della procura e della rappresentanza, ma ciò mal si concilia con la figura dell’associazione.

La soluzione è offerta proprio dalla teoria attore-rete di Latour: assumendo che uomini e macchine non agiscono nella società in qualità di attanti individuali, ma quali ibridi, ossia in quanto associazioni, si conclude che da questa unione di risorse risulti un unico centro di imputazione. Così, questa relazione stretta tra entità umane e non, che si crea negli ibridi, ed in virtù della quale ciascun partecipante non agisce per sé stesso ma per l’ibrido, consente la partecipazione dei sistemi intelligenti alle relazioni giuridiche. È evidente, in questo caso, la possibilità di effettuare un paragone con un rappresentante di una impresa: questi agisce, non per sé stesso, ma per l’impresa intesa come «sistema sociale»¹²⁵.

Questa lettura della teoria di Latour è coerente con la trattazione qui in esame: infatti, permette sia di attribuire una regolamentazione giuridica dei

¹²² Teubner, *ult. op. cit.*, 52.

¹²³ Esposito, *ult. op. cit.*, 249.

¹²⁴ Moro, in Provolo, Riondato, Yenisey, 2014, 533-544.

¹²⁵ Teubner, 1987, 61.

rapporti interni all'associazione uomo-macchina, in termini di rappresentanza; sia di segnare la prevalenza dell'associazione nelle relazioni. In questo modo le singole condotte degli enti umani e non confluiscono direttamente in capo alla associazione uomo-macchina: «è l'unione uomo-algoritmo, come sistema di azione autonomo, ad entrare in relazione comunicativa con i terzi, non gli esseri umani coinvolti o gli algoritmi»¹²⁶, rendendo così l'associazione imputabile di una forma di responsabilità. Una simile ricostruzione, che si avvale del diritto collettivo applicato alla associazione uomo-macchina, risponde perfettamente alla realtà di questa relazione, rispetto all'utilizzo delle categorie del diritto individuale.

Attualmente però sono preferite le soluzioni individualiste, come accade ad esempio per lo “scopo contrattuale”, quale comune alle parti in rapporti di natura sinallagmatica. Lo scopo contrattuale, infatti, non viene certo considerato come un tutt'uno con i contraenti, ma contribuisce a ridurre ad unità la relazione tra gli stessi, senza però sfociare nella configurazione di una forma associativa. Applicando queste considerazioni in via analogica alla relazione uomo-*computer* si potrebbe addirittura arrivare alla introduzione di uno «“scopo associativo” [ponendo] così rimedio ai difetti di una costruzione puramente individualistica e l'unità dell'associazione uomo-macchina riceverebbe valore giuridico»¹²⁷. Entrambe le soluzioni prospettate, sia di diritto collettivo che di diritto individuale, partono dal presupposto del riconoscimento di una limitata capacità giuridica ad entità non umane e sono appropriate per due diverse situazioni: la opzione associativa è più appropriata in presenza di una stretta relazione uomo-macchina; la opzione individualista, invece, appare adeguata alle ipotesi in cui si evidenzia un essere “attore” del sistema intelligente nella compagine sociale di riferimento.

7.3. Rischio di interconnessione

Nel momento in cui gli algoritmi sono inseriti in un sistema multi-agente la personificazione di ibridi e attanti riscontra un grave limite, rappresentato dal cosiddetto rischio di interconnessione. Questa prospettiva annienta la individualità, comportando importanti ricadute in tema di attribuzione delle azioni e della responsabilità: è il singolo attore ad aver compiuto l'azione, oppure essa è la conseguenza dell'insieme delle singole condotte? Per rispondere ad un simile quesito Spiecker¹²⁸ propone una responsabilità ripartita tra tutti coloro i quali partecipano alla “interconnessione”, che si troveranno così a rispondere per le azioni degli algoritmi interconnessi. Ma è chiaro che

¹²⁶ Teubner, 2019, 106.

¹²⁷ Teubner, *ult. op. cit.*, 113.

¹²⁸ Spiecker, 2016, 703.

questa soluzione porti a nient'altro che ad una strada senza uscita: lo spazio digitale e la rete sono delle aree nelle quali si creano infinite interconnessioni di algoritmi, e pertanto sarà impossibile determinare i soggetti responsabili.

L'ulteriore proposta risolutiva promossa da Teubner sarebbe quella di «attribuire la responsabilità non più a “persone”, identificate quali portatrici di azioni, ma direttamente e soltanto alle azioni stesse, identificate in sé, cioè ad una *matrice anonima*¹²⁹ data da processi sociali e digitali: una matrice di azioni, quindi, non una struttura soggettivata»¹³⁰. Da ciò si evince che la responsabilità andrebbe imputata alla decisione autonoma, e non a chi ha assunto quella decisione, o alla rete, o alle interconnessioni, o agli agenti *software*. A questo punto, però, si dovrebbe mutare l'intera prospettiva del “soggetto agente” radicatasi nel mondo del diritto, perché si prescinderebbe dallo stesso, sia esso individuale o collettivo, focalizzandosi solo ed esclusivamente sulle decisioni in sé¹³¹, e sul loro essere rischiose. Lo sforzo dei giuristi sarebbe, di fatto, concentrato solo sulla individuazione di categorie di rischio, che verrà poi trattato in base alle tecniche di amministrazione dei rischi che l'ordinamento mette già a disposizione. Per soddisfare questo obiettivo il diritto ha bisogno di costruire una “collettività” che domini questi rischi, li prevenga ed eventualmente risarcisca i danni, e questo è possibile solo attraverso una attenta programmazione degli algoritmi e delle condizioni del loro utilizzo.

7.4. Uno *status* giuridico variabile

L'emergere dei tre rischi appena analizzati porta il giurista ad individuare uno *status* giuridico variabile o flessibile per gli agenti informatici programmati secondo l'intelligenza artificiale. Per soddisfare le esigenze di tutela

¹²⁹Teubner, 2006. Nello specifico Teubner sostiene in un passaggio, 177: «*Sie beschränken sie aber auf die Gefährdungen, die von den eher anschaulichen „Kollektiv-akteuren“ (Staaten, politischen Parteien, Wirtschaftsunternehmen, Konzernen, Verbänden) ausgehen und verfehlen die Gefährdung durch die anonyme „Matrix“, durch verselbständigte kommunikative Prozesse (Institutionen, Funktionssysteme, Netzwerke), die nicht als Kollektive personifiziert sind. Schon die politischen Menschenrechte dürfen nicht als Beziehungen zwischen politischen Akteuren (Staat vs. Bürger), also als Ausdruck von Person-Person-Beziehungen gesehen werden. Vielmehr geht es um Beziehungen zwischen anonymen Machtprozessen auf der einen Seite und den geschundenen Körpern und den verletzten Seelen auf der anderen. Das wird in der Kommunikation nur sehr unvollkommen, um nicht zu sagen irreführend, als Verhältnis der „Person“ Staat zu den „Personen“ Individuen ausgedrückt.*

¹³⁰Teubner, *ult. op. cit.*, 118.

¹³¹Hildebrandt, 2015. La quale ritiene di doversi concentrare sull'azione, 26: «*Because di agents may be distributed on and possibly mobile between different hardware applications and because as a multiagent system it is capable of changing shape (polymorphous), it is not always easy to identify where the emerging agent is located and what is and is not a part of fit at any point time. However, in so far ad the emergent behaviours oft he system allow is identification as a unity of action, it can be qualified as an agent, whatever the underlying embodiment*», citato alla nota numero 215 di pagina 119 in Teubner, 2019.

poste da questi rischi, si dovrebbe in ciascuna occasione riconoscere uno *status* specifico in ragione del ruolo assunto di volta in volta dal sistema informatico.

Di fronte al rischio di autonomia una valida soluzione sarebbe quella, proposta da Teubner, di attribuire agli agenti *software* lo *status* di «attori parzialmente provvisti di capacità giuridica»¹³². Alle loro autonome decisioni è riconosciuta una valenza giuridica e, in caso di illiceità, è previsto un regime di responsabilità *ad hoc*. Inoltre, in questo contesto in cui gli algoritmi sono dotati di soggettività giuridica parziale, gli stessi diventano capaci di concludere contratti in quanto rappresentanti. Allo stesso modo, poi, sono riconosciuti anche come «ausiliari dotati di capacità giuridica», in modo tale che qualsiasi comportamento “difettoso” della macchina possa essere imputato all’impresa produttrice in quanto tale.

Una reazione al rischio di associazione potrebbe invece essere quella di riconoscere agli agenti *software* uno *status* giuridico in qualità di partecipanti ad una associazione uomo-macchina. La soluzione prospettabile concretamente è quella di introdurre il nuovo concetto di «scopo associativo», in virtù del quale sarebbe possibile interpretare «le dichiarazioni negoziali degli agenti software»¹³³, per poi ricostruire i loro diritti e doveri. Tuttavia, la soluzione in assoluto migliore sarebbe quella di considerare l’associazione stessa quale centro di imputazione dei diritti, dei doveri e delle azioni.

Per quanto concerne l’ultimo rischio, quello di interconnessione, lo *status* dovrebbe essere attribuito ai “poli di rischio”: in questo caso è lo *status* che forgia la posizione giuridica degli algoritmi nell’ambito del flusso di informazioni che caratterizza un agente *software*, e la responsabilità del polo emerge solo in relazione ad una condotta illecita.

Presupposto per il riconoscimento di questi tre diversi *status* è poter considerare gli agenti *software* come dotati di una soggettività giuridica parziale. Questi risultano approfondimenti concretizzabili nel mondo del diritto, che permettono di dettare delle soluzioni sostenibili a livello giuridico, senza ricercare il supporto di finzioni. Ma ancor più rilevante, come sottolineato in chiusura da Teubner, è che una simile triplice lettura concede il privilegio di «soddisfare l’imperativo che anche nello spazio digitale l’eguale sia trattato in modo eguale e il diseguale in modo diseguale»¹³⁴, quale obiettivo tutt’altro che semplice da raggiungere. Infatti, è bene rimarcare il fatto che utilizzare delle categorie giuridiche e semantiche, che si sono radicate nel corso dei secoli e sono state costruite *ad hoc* per gli esseri umani, appare un

¹³² Teubner, 2019, 125.

¹³³ Teubner, *ult. op. cit.*, 126.

¹³⁴ Teubner, *ult. op. cit.*, 127.

tentativo forzato di creare una similitudine tra macchine ed umani, la quale non può essere prospettata.

8. Considerazioni conclusive. Sono pensabili diritti e doveri per le macchine intelligenti?

L'accoglimento di un'altra forma di intelligenza, sia essa non *human-like* in termini di emotività e coscienza, impone doveri di matrice etica e giuridica volti al riconoscimento di specifiche forme di tutela e garanzia per gli esseri umani, *in primis*, ma anche per gli stessi sistemi intelligenti. Tenendo presente, però, che l'obiettivo primario è la protezione del soggetto leso da una condotta di una macchina intelligente, specificatamente in termini risarcitori e restitutori, le soluzioni declinabili dipendono dall'esito delle riflessioni sulle teorie della tutela riconoscibile ai sistemi intelligenti.

Focalizzandosi, pertanto, sulla questione della soggettività, in quanto preordinata alla problematica del regime di responsabilità configurabile in capo ai sistemi programmati secondo l'intelligenza artificiale, emerge con evidenza la primaria esigenza di condurre un'indagine che concili il rigore concettuale tipico degli studi giuridici attuali e un'analisi sociale, rivedendo la portata semantica di alcune categorie tipiche della tradizione giuridica, filosofica, sociologica e antropologica, quali ad esempio la agentività, la soggettività o la personalità. Dopodiché, spogliandosi da qualunque forma di radicamento e pregiudizio di ordine puramente antropocentrico, è utile soffermarsi sugli aspetti manchevoli di quelle che sono state individuate come le due teorie della soggettività, quella intenzionale e quella funzionale.

Per ciò che concerne la soggettività denominata intenzionale o psicologica, frutto dell'originale tentativo di superare le aderenze all'elemento psicologico e mentale, nonché quello di superare il naturale ed inevitabile paragone continuo con l'essere umano, proposta da Sartor, sulla scia degli studi condotti da Dennett e dai funzionalisti¹³⁵, si può osservare che la stessa esibisca la propria origine moderna, ponendo al centro l'essere umano nella sua esclusiva individualità¹³⁶. Una simile idea di soggettività unitaria, che fondamentale è il risultato dei condizionamenti culturali di matrice razionale ed individuale che connotano l'indagine storica, filosofica e giuridica attuale, non solo mal si concilia con l'obiettivo di tutela primario, che riguarda l'essere umano, ma soprattutto diviene difficilmente sostenibile a livello giuridico, richiedendo lo sforzo di rivoluzionare le categorizzazioni individualistiche.

¹³⁵ Per approfondimenti: Di Francesco, 2002; Gozzano, 1997.

¹³⁶ Cfr. Cavalla, 2011; Cotta, 1997.

Diversamente, la seducente teoria attore-rete proposta da Latour e Callon, che in combinazione con la teoria del sistema sociale di Luhmann conduce alla rottura della distinzione tra soggetto ed oggetto, configurando l'azione come il risultato dell'interazione tra soggetti umani e oggetti in movimento connessi in rete, i quali vengono definiti "attanti", fornisce ottimi elementi per l'introduzione di una nuova *soggettività relazionale*.

Come osservato nei paragrafi dedicati alla soggettività funzionale, la capacità comunicativa di uomini e macchine intelligenti, diverrebbe lo strumento per eccellenza per raggiungere il superamento della singolarità giuridica e approdare alla figura *dell'uomo in relazione*, al quale, appunto, «l'umanità (...) viene attribuita in quanto membro, o parte, della società in senso ora stretto ora lato»¹³⁷.

Questa appare una soluzione più convincente dell'attribuzione di ragione e volontà all'ente intelligente, traendo la sua ragion d'essere dal classico schema di derivazione romanistica, che prende in considerazione la correlazione di azioni di più soggetti umani, e non, nell'ambito di un medesimo rapporto eterogeneo. Il sistema potrà trarre riconoscimento sociale, e di conseguenza anche un riconoscimento e una tutela giuridica grazie alla sua capacità comunicativa, che gli permette di partecipare, in quanto *actor*, o meglio "attante", al *network*.

Inoltre, per concludere, l'*escamotage* di considerare protagonista del diritto l'ibrido nato dalla relazione tra uomo e macchina, permette di *bypassare* la scelta obbligata tra il continuare a considerare il sistema intelligente come *res* od elevarlo a persona fisica, permettendo di considerare, appunto, solo la relazione in quanto tale.

Queste riflessioni sono essenziali al fine di porre la questione della cd. "personalità elettronica", la quale si presenta come punto di partenza per affrontare consequenzialmente due altri problemi: da un lato, attribuire una personalità ad una entità robotica significherebbe riconoscerla anche titolare di una sfera di doveri e obblighi, nonché di un patrimonio con il quale la stessa potrà rispondere direttamente nei confronti dei soggetti lesi; dall'altro lato, sarebbe necessario, altresì, riconoscere un sistema intelligente come dotato di un assetto di diritti e di tutele.

Entrando più nel dettaglio delle questioni, risulta che dal punto di vista etico e tecnogiuridico, è chiaro come doveri e responsabilità possano essere attribuiti a soggetti anche non "persone fisiche", come accade con il ruolo attivo delle persone giuridiche in seguito al d.lgs. 231/2001. Tuttavia, riconoscere la macchina intelligente come meritevole di una forma di tutela incontra delle difficoltà maggiori, perché ciò comporterebbe la traslazione dell'entità intel-

¹³⁷ Montanari, 2009, 7.

ligente dall'essere considerata mera *res*, all'essere elevata a soggetto animato degno di tutela, al pari di una persona.

Pertanto, la personalità elettronica potrebbe essere sfruttata per innestare una reazione a catena, la quale, a partire dall'accettazione sociale della personificazione del sistema intelligente, giunge ad attribuire alla stessa una dignità "di persona (o di entità)", alla quale garantire tutele, diritti, libertà, ed anche una capacità patrimoniale. Quest'ultimo riconoscimento sarà funzionale ad elevare l'ente intelligente a "soggetto", al quale poi imputare una responsabilità diretta¹³⁸.

Questo passaggio da *res* a soggetto degno di tutela non è una novità: infatti, nella tradizione romanistica si è già assistito ad un progressivo riconoscimento dello schiavo come dotato di una natura umana, anche se inizialmente era considerato come una *res*¹³⁹; egli non venne subito elevato ad essere ritenuto una "persona", ma almeno un soggetto degno di tutela¹⁴⁰. Queste considerazioni non sono troppo lontane da alcune situazioni correnti; infatti, in modo equivalente, la vicenda si sta riproponendo a proposito degli animali: questi restano dei beni assoggettati ad un proprietario, ma ad essi viene accordata una tutela¹⁴¹, concedendo loro di essere portati dallo stato di mera *res* allo stato di soggetti degni di tutela, o per usare le parole di Ruffolo, a «esseri senzienti».

Ma se, in un futuro non troppo lontano, la società civile arrivasse a liberare la macchina dai rapporti esclusivamente relegati alla sfera del possesso di cui è oggi protagonista, e quindi dallo *status* di *res*, di quali diritti si potrebbe parlare? Ma soprattutto, in presenza di diritti si dovrà, conseguentemente, anche arricchire la gamma dei doveri?

Tutte le considerazioni esposte sinteticamente in questo paragrafo conclusivo portano ad un vero e proprio paradosso; infatti, la corsa alla personaliz-

¹³⁸ Ruffolo, 2021,125.

¹³⁹ In particolare, si trattava di *res mancipi*, ossia le cose di maggior pregio come appunto gli schiavi, i fondi e gli animali da tiro e da soma. Marrone, 2000, 292.

¹⁴⁰ Al proposito è sufficiente ricordare che la posizione dei giuristi romani, rispetto alla condizione giuridica degli schiavi, era complessa e addirittura contraddittoria: da un lato in quanto esseri umani erano considerati appartenenti alla categoria delle *personae*; d'altro lato, potendo essere oggetti di proprietà e di altri diritti soggettivi erano allo stesso tempo *res*, rispetto alle quali il *dominus* godeva del forte potere del *ius vitae ac necis*.

¹⁴¹ Si pensi ad esempio agli articoli 544ter (Maltrattamento di animali), 638 (Uccisione o danneggiamento di animali altrui) e 727 (Abbandono di animali) del codice penale; ma anche ad alcune sentenze della Cassazione che si esprimono in merito alla tutela degli animali: Sentenza n. 26368 del 6 luglio 2011, Terza Sezione Penale della Corte di Cassazione, che ritiene integrante il reato di maltrattamento di animali (ex articolo 544ter) legare un cane ad una catena/corda troppo corta; Sentenza n. 46291 del dicembre 2003, Terza Sezione Penale della Corte di Cassazione, la quale statuisce che prendere a calci un cane per futili motivi è reato (ex articolo 727), «in quanto anche gli animali sono esseri dotati di sensibilità e devono essere trattati con umanità».

zazione delle entità artificiali ritenute capaci di assumere decisioni autonome non ha fatto altro che allontanare dall'obiettivo originario e condurre alla entificazione della comunicazione, anziché dell'ente intelligente¹⁴². Il fraintendimento alla base di questo paradosso è dato dalla sopravvalutazione dell'esigenza di creare una personalità per le entità artificiali, al sol fine di responsabilizzarle in via diretta; tralasciando invece che la personalizzazione è funzionale anche e soprattutto allo scopo di garantire diritti e tutele alle macchine intelligenti, con le conseguenti problematiche¹⁴³.

Inoltre, al di là della persuasività delle posizioni su esposte, si tratta comunque di una questione ancora aperta, anche a causa della molteplicità di modelli di sistemi programmati secondo l'intelligenza artificiale, per i quali non è possibile giungere ad un trattamento giuridico generalizzato. Non si può prescindere da altri elementi che necessitano di essere presi in considerazione, come la natura del *robot* e l'ambiente in cui opera, nonché la qualità e quantità del controllo che l'uomo imprime su di esso¹⁴⁴.

A ciò si aggiunge la preoccupazione espressa da un gruppo di esperti di intelligenza artificiale e robotica in una lettera aperta alla Commissione Europea¹⁴⁵, la quale non è trascurabile: la posizione da loro assunta è tassativa nel rifiutare un approccio puramente giuridico e legale, che non tenga in considerazione anche la componente etica, psicologica e sociale¹⁴⁶; infatti, questo

¹⁴²Degna di attenzione è, al proposito, l'affermazione di Garrido Martin, 2022, 173. L'autore sostiene che nella nostra modernità «*no se trata por tanto de la humanización de la máquina, sino de la tecnologización de la mente, y esto a través del constante y regular contacto con el dispositivo digital, que atraviesa nuestro comportamiento diario*».

¹⁴³Ruffolo, 2021, 121-122.

¹⁴⁴Bisol, Carnevale, Lucivero, 2014, 248 ss: «[...] una comprensione profonda della tecnologia robotica, prima di tutto da un punto di vista tecnico, ma anche rispetto a una prospettiva che nel suo insieme può essere definita fenomenologica. Per un'analisi etica accurata, infatti, devono essere tenuti in considerazione aspetti quali la natura del robot (automa, sistema bionico ibrido); i compiti che è in grado di svolgere (navigazione, locomozione, manipolazione), l'ambiente in cui è in grado di operare, il tipo di controllo che l'uomo esercita sul robot durante il suo funzionamento (robot autonomo, automatico, teleoperato). Sono inoltre rilevanti criteri come l'interazione tra uomo e robot (fisica, non fisica, puntuale o continua, invasiva o non invasiva), il rapporto di utenza e infine il grado di prossimità del robot, che può operare entrando o non entrando in contatto con l'essere umano. Fondamentale è inoltre comprendere in quale contesto lavorativo o sociale e per quali fini viene impiegato il robot [...]».

¹⁴⁵Riferimento al sito: <http://www.robotics-openletter.eu/>. Nello Specifico si tratta di un gruppo di Leaders politici, ricercatori nel campo della robotica e dell'IA, esperti in psicologia e salute mentale, esperti in diritto ed etica, i quali uniscono la loro voce per sottolineare le conseguenze negative della costruzione di uno status giuridico per i robot nell'Unione Europea.

¹⁴⁶In particolare: «*a. A legal status for a robot can't derive from the Natural Person model, since the robot would then hold human rights, such as the right to dignity, the right to its integrity, the right to remuneration or the right to citizenship, thus directly confronting the Human right. This would be in contradiction with the Charter of Fundamental Rights of the European*

ristretto gruppo di esperti esclude che la personalità elettronica possa essere forgiata non solo sulla base della personalità riconosciuta alle persone fisiche ma anche su quella propria delle persone giuridiche. Nel primo caso perché questo condurrebbe all'attribuzione al *robot* di diritti ascritti tradizionalmente agli esseri umani, mentre nel secondo caso perché vi sarebbe la necessità della presenza di un soggetto umano a monte dell'entità robotica che la rappresenta e la dirige.

È chiaro, in relazione alle osservazioni appena espresse, che si sta cercando di prendere le distanze dalla ammissione del paragone uomo-macchina per il timore, tutto filosofico, di doversi trovare a rispondere criticamente alle domande che la cultura odierna spesso pone: cosa individua davvero l'uomo? E cosa lo distingue da un sistema intelligente sofisticato? Si potrebbe argomentare partendo dalle emozioni, dai sentimenti e dalla capacità di pensare, ma sarebbe sufficiente?¹⁴⁷ Queste sono solo alcune delle suggestioni che potrebbero sorgere spontanee in conseguenza delle riflessioni condotte in questa trattazione, le quali dimostrano la necessità di indagare in profondità e prima d'altro una teoria critica della soggettività.

Bibliografia

Amoroso D., Tamburrini G. 2019. *I sistemi robotici ed autonomia crescente tra etica e diritto: quale ruolo per il controllo umano?*. In *BioLaw Journal-Rivista di Biodiritto*, n. 1: 33-51.

*Union and the Convention for the Protection of Human Rights of the European Union and the Convention for the Protection of Human Rights and Fundamental Freedoms. b. The legal status for a robot can't derive from the Legal Entity model, since it implies the existence of human persons behind the legal person to represent and direct it. And this is not the case for a robot. c. The legal status for a robot can't derive from the Anglo-Saxon Trust model also called Fiducie or Treuhand in Germany. Indeed, this regime is extremely complex, requires very specialized competences and would not solve the liability issue. More importantly, it would still imply the existence of a human being as a last resort – the trustee or fiduciary – responsible for managing the robot granted with a Trust or a Fiducie». A questa conclusione è giunta anche Nathalie Nevejeans, la quale, nello Studio richiesto dal Parlamento Europeo – Commissione Affari Legali, e commissionato, supervisionato e pubblicato dal Policy Department for "Citizens' Rights and Constitutional Affairs", ritiene: «When considering civil law in robotics, we should disregard the idea of autonomous robots having a legal personality, for the idea is as unhelpful as it is inappropriate [...] From a scientific, legal and even ethical perspective, it is impossible today – and probably will remain so for a long time to come – for a robot to take part in legal life without a human being pulling its strings [...]. Doing so risks not only assigning rights and obligations to what is just a tool, but also tearing down the boundaries between man and machine, blurring the lines between the living and the inert, the human and the inhuman», *European civil law rules in robotics*, 2016.*

¹⁴⁷ Moro, in Ruffolo (a cura di), 2020, 45-61.

- Balconi M. 2014. *Consapevolezza e senso di agentività: un'integrazione possibile?*. In Il Mulino – Rivista web, n. 1. Bologna: 79-101.
- Benanti, P. 2016. *The Cyborg: corpo e corporeità nell'epoca del post-umano. Prospettive antropologiche e riflessioni etiche per un discernimento morale*. Cittadella Editrice, Assisi.
- Berti Suman, A. 2020. *Intelligenza artificiale e soggettività giuridica: quali diritti (e doveri) dei robot?*. In Alpa G. (a cura di), *Diritto e intelligenza artificiale*. Pacini Giuridica, Pisa: 251-273.
- Bryant, L. R. 2011. *The Democracy of Objects*. In Open Humanities Press, Michigan.
- Buttazzo G. 2002. *Coscienza artificiale: missione impossibile?*. In *Mondo Digitale*, n.1: 16-25.
- Callon, M., Latour, B. 1981. *Unscrewing the big Leviathan: how actors macro-structure reality and how sociologist help them to do so*, in Knorr-Cetina K., Cicourel V., *Advances in social theory and methodology. Toward an integration of micor- and macro- sociologies*. Routledge& Kegan Paul Boston, London and Henley, London-Boston: 276-303.
- Cavalla, F. 2011. *All'origine del diritto. Al tramonto della legge*. Jovene Editore.
- Comitato Nazionale per la Bioetica e Comitato Nazionale per la Biosicurezza, le Biotecnologie e le Scienze della Vita, 2017, *Sviluppi della robotica e della roboetica*.
- Chen B., Kwiatkowski R., Vondrick C., Lipson H. 2022, *Fully body visual self-modeling of robot morphologies*. In *Science Robotics*: 1-13.
- Churchland, P. M. 1981. *Eliminative Materialism and the Propositional Attitudes*. In *The Journal of Philosophy*, Vol. 78, n. 2: 67-90.
- Commissione Etica, o Ethik-Kommission (nominata dal Ministero dei Trasporti e delle Infrastrutture Digitali tedesco). 2017. *Guida automatica e connessa*.
- Commissione Europea. 4 aprile 2021. *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*. COM (2021) 206 final. Bruxelles.
- Commissione Europea. 19 febbraio 2020. *Libro Bianco sull'intelligenza artificiale - Un approccio europeo all'eccellenza e alla fiducia*. COM (2020) 65 final. Bruxelles.
- Commissione Europea. 19 febbraio 2020. *Relazione sulle implicazioni dell'intelligenza artificiale, dell'Internet delle cose e della robotica in materia di sicurezza e di responsabilità*. COM (2020) 64. Bruxelles.

- Commissione Europea. 25 aprile 2018. *L'Intelligenza Artificiale per l'Europa*. COM (2018) 237 final. Bruxelles.
- Contaldo A., Campara F. 2020. *Intelligenza artificiale e Diritto. Dai sistemi esperti "classici" ai sistemi esperti "evoluti": tecnologia e implementazione*, in Taddei Elmi G., Contaldo A. (a cura di), *Intelligenza Artificiale. Algoritmi giuridici. Ius condendum o "fantadiritto"?*. Pacini Giuridica, Pisa: pp. 1-96.
- Cotta, S. 1997. *Soggetto umano. Soggetto giuridico*. Giuffrè Editore.
- Daly A., Hagendorff T., Hui L., Mann M., Marda V., Wagner B., Wei Wang W. 2022. *AI, Governance and Ethics: Global Perspectives*. In Michlitz H. W., Pollicino O., Amnon Reichman A., Simoncini A., Sartor G., De Gregorio G., *Constitutional Challenges in the Algorithmic Society*. Cambridge University Press: 182-201.
- Dennett, D. C. 2009. *Intentional System Theory*. In Beckermann A., Mclaughlin B. P., Walter S., *The Oxford handbook of philosophy of mind*. Oxford University Press: 1-22.
- Dennett, D. C. 1997. *True Believers: The Intentional Strategy and Why It Works*. In Haugeland J., *Mind Design II*. MIT Press.
- Dennett, D. C. 1996. *Kinds of Minds: Towards an Understanding Consciousness*. New York.
- Dennett, D. C. 1992. *L'atteggiamento intenzionale*. Il Mulino, Bologna.
- Dennett, D. C. 1988. *The Intentional Stance in theory and practice*. In Machiavellian Intelligence. Oxford University Press:180-202.
- Di Francesco, M. 2002. *Introduzione alla filosofia della mente*. Carocci Editore, Roma.
- Di Sabato, D. *Gli smart contracts: robots che gestiscono il rischio contrattuale*. In Perlinger G, Fachechi A., *Ragionevolezza e proporzionalità nel diritto contemporaneo*. Edizioni Scientifiche Italiane, Napoli: 387-419.
- Esposito, E. 2017. *Artificial Communication? The Production of Contingency by Algorithms*. In *Zeitschrift für Soziologie*. Vol. 4, n. 46: 249-256.
- Faggin, F. 2015. *Sarà possibile fare un computer consapevole?*. In *Mondo Digitale*, n. 61: 1-13.
- Femia, P. 2019. *Introduzione. Soggetti responsabilità. Algoritmi e diritto*. In Teubner G., *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, (a cura di) Femia, P. Università della Campania Luigi Vanvitelli – Dipartimento di Scienze Politiche Jean Monnet. Edizioni Scientifiche Italiane, Napoli: 7-16.

- Fini M., Milani P. 2002. *Intelligenza e coscienza. L'IA tra Searle e Dennett. Sviluppo dell'Intelligenza Artificiale*. In Di Francesco M., Introduzione alla filosofia della mente. Carocci Editore, Roma: 94-118.
- Floridi L., Cowls J., Beltrametti M., Chatila R., Chazerand P., Dignum V., Luetge C., Medeline R., Pagallo U., Rossi F., Schafer B., Valcke P., Vayena E. 2018. *AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations*. In *Minds and Machines*, n. 28: 689-707.
- Garrido Martin, J. 2022. *Inteligencia (artificial) y automatismo. Anatomía de un conflicto*. In Llano Alonso F. H. (director), *Inteligencia Artificial y Filosofía del Derecho*. Ediciones Laborum, S.L.: 169-187.
- Gozzano, S. 1997. *Storia e teoria dell'intenzionalità*. Laterza, Bari.
- Gruber, M. C., 2015. *Was spricht gegen Maschinenrechten?*. In Gruber M. C., Bung J., Ziemann S., *Autonome Automaten: Künstliche Körper und artifizielle Agenten in der technisierten Gesellschaft*. Berlin.
- Haugeland, J. 1981. *Semantic Engines. An Introduction to Mind Design*. In *Mind Design*, MIT Press. Cambridge: 34-50.
- Hildebrandt, M. 2015. *Smart Technologies and the End(s) of Law*. Edward Elgar Pub.
- Hofstadter R., Dennett D. 2001. *The Mind's I. Fantasies and reflections on self and soul*. Basic Book, New York.
- Ihde, D. 1990. *Technology and the Lifeworld: from Garden to Earth*. Indiana University Press.
- Latour, B. 2005. *Reassembling the Social. An introduction to Actor-Network-Theory*. Oxford University Press, New York.
- Latour, B. 1999. *Pandora's Hope: Essays on the Reality of Science Studies*. Harvard University Press, Cambridge.
- Law, J. 1992. *Notes on the Theory of the Actor-Network: Ordering, Strategy, and Heterogeneity*. In *Systems Practice*, Vol. 5, n. 4: 379-393.
- Lombardi Vallauri, L. 1990. *Il Meritevole di Tutela*. Giuffrè, Milano.
- Luhmann, N. 1987. *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Suhrkamp- taschenbuch Wissenschaft, Frankfurt am Main.
- Manzotti R., Tagliascio V. 2002. *Si può parlare di coscienza artificiale?*. In *Sistemi Intelligenti*. Il Mulino: 89-108.
- Marrone, M. 2000. *Istituzioni di diritto romano*. Palumbo, Italia.

- Montanari, B. 2009. “Soggetto umano – Soggetto giuridico”. *Il diritto nella prospettiva ontologico-esistenziale di Sergio Cotta*. In *Teoria e Critica della Regolazione Sociale*. Centro Studi TCRS: 1-34.
- Moro, P. 2021. *Alle frontiere della soggettività: indizi di responsabilità delle macchine intelligenti*. In Ruffolo, U. (a cura di), *XXVI lezioni di diritto dell’intelligenza artificiale*. G. Giappichelli Editore, Torino: 55-65.
- Moro, P. 2020. *Macchine come noi. Natura e limiti della soggettività robotica*. In Ruffolo U. (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l’etica*. Giuffrè Francis Lefebvre, Milano: 45-61.
- Moro, P. 2014. *Biorobotica e diritti fondamentali. Problemi e limiti dell’intelligenza artificiale*. In Provolo D., Riondato S., Yenisey F., *Genetics, Robotics, Law, Punishment*. Padova University Press, Padova: 533-544.
- Organizzazione Mondiale della Proprietà Intellettuale. 1978. *Model provisions on the protection of computer software*. Ginevra.
- Parlamento Europeo. 2021. *Proposta che stabilisce regole armonizzate sull’intelligenza artificiale (legge sull’intelligenza artificiale) e modifica alcuni atti legislativi dell’unione*. {sec(2021) 167 final} - {swd(2021) 84 final} - {swd(2021) 85 final}.
- Parlamento Europeo. 16 febbraio 2017. *Norme di diritto civile sulla robotica*, Risoluzione del Parlamento europeo recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)).
- Piva A., D’agostini D. 2003. *La tutela giuridica dei programmi per elaboratore*. In *Mondo Digitale*: 66-77.
- Ruffolo, U. 2021. *La personalità elettronica tra “doveri” e “diritti” della machina*. In Ruffolo U. (a cura di), *XXVI lezioni di diritto dell’intelligenza artificiale*. G. Giappichelli Editore, Torino: 115-127.
- Ruffolo, U. 2020. “*La personalità elettronica*”. In Ruffolo, U. (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l’etica*. Giuffrè Francis Lefebvre, Milano: 213-236.
- Ruffolo, U. 2020. *La responsabilità da artificial intelligence, algoritmo e smart product: per i fondamenti di un diritto dell’intelligenza artificiale self learning*. In Ruffolo U. (a cura di), *Intelligenza artificiale. Il diritto, i diritti, l’etica*. Giuffrè Francis Lefebvre, Milano: 93-124.
- SAE International, *Taxonomy and Definitions for Terms Related to Driving Automation System for On-Road Motor Vehicle (J3016)*, giugno 2018.

- Sarra, C. 2021. *Dalla Cibernetica alla Data Ethics. Linee di sviluppo dell'etica applicata alla rivoluzione informatica*. In Moro P. (a cura di), *Etica, Diritto e Tecnologia. Percorsi dell'informatica giuridica contemporanea*. Franco Angeli, Milano: 25-44.
- Sartor, G. 2003. *L'intenzionalità dei sistemi informatici e il diritto*. In *Rivista Trimestrale di Diritto e Procedura Civile*, n.1: 23-53.
- Sartor, G. 1996. *Intelligenza artificiale e diritto. Un'introduzione*. Giuffrè Editore, Milano.
- Searle, J. R. 1980. *Minds, brains, and programs*. In *Behavioral and Brain Science*, Vol. 3, n. 3: 417-457.
- Schirmer, J. E. 2016. *Rechtsfähige Roboter*. In *Juristenzeitung*, n.13.
- Spiecker, I. 2016. *Zur Zukunft systemischer Digitalisierung – Erste Gedanken zur Haftungs- und Verantwortungszuschreibung bei informationstechnischen Systemen. Warum für die systemische Haftung ein neues Modell erforderlich ist*. In *Computer und Recht*: 698-704.
- Sprenger, G. 2017. *Communicated into being: System theory and the shifting of ontological status*. In *Anthropological Theory*, Vol. 17: 108-122.
- Taddei Elmi G., Contaldo A. 2020. *Algoritmi giuridici. Ius condendum o "fantadiritto"?*. Pacini Giuridica, Pisa.
- Taddei Elmi, G. 2016. *Il robot tra ius condendum e ius conditum*. In *Informatica e diritto*, Vol. XXV, n. 1: 115-137.
- Taddei Elmi G., Romano F. 2010. *Robotica: tra etica e diritto. Un seminario promosso dal Dipartimento Identità Culturale e dall'ITTIG del CNR*. In *Informatica e diritto*, Vol. XIX, n.1-2:146.
- Taddei Elmi, G. 2004. *Soggettività artificiali e diritto*. Relazione tenuta il 3 giugno in occasione della chiusura del Master sul Diritto delle nuove tecnologie informatiche organizzato dal CSIG sezione di Firenze.
- Taddei Elmi, G. 1990. *I diritti dell'intelligenza artificiale tra soggettività e valore: fantadiritto o ius condendum?*. In Lombardi Vallauri L., *Il Meritevole di Tutela*. Giuffrè, Milano: 685-711.
- Tamburrini, G. 2020. *Etica delle macchine. Dilemmi morali per robotica e intelligenza artificiale*. Carocci editore, Roma.
- Teubner, G. 2019. *Soggetti giuridici digitali? Sullo status privatistico degli agenti software autonomi*, (a cura di) Femia, P. Università della Campania Luigi Vanvitelli – Dipartimento di Scienze Politiche Jean Monnet. Edizioni Scientifiche Italiane, Napoli.

- Teubner, G. 2006. *Die anonyme Matrix: Zu Menschenrechtsverletzungen durch „private“ transnationale Akteur*. In *Der Staat*, Duncker & Humblot GmbH. Vol. 45, n. 2:161-187.
- Teubner, G. 2006. *Rights of Non-Human? Electornic Agents and Animals as New Actors in Politics and Law*. In *Journal of Law and Society*, Vol. 33: 497-521.
- Teubner, G. 1987. *Unternehmenkorporatismus. New Industry Policy und das „Wesen“ der Juristischen Person*. In *Kritische Vierteljahresschrift für Gesetzgebung und Rechtswissenschaft*: 61-85.
- Troncarelli, B. 2018. *Soggettività umana e diritto del lavoro nella trasformazione digitale*. In *Jus Online*, n. 2: 80-115.
- Turing, A. 1950. *Computing Machinery and Intelligence*. In *Mind*, n. 49: 433-460.
- Turing, A. M. 1936. *On Computable Numbers, with an application to the Entscheidungsproblem*. In *Proceedings of the London Mathematical Society*, Vol. 42: 230-265.
- Verbeek, P. P. 2001. *Don Ihde: The Technological Lifeworld*. In *American Philosophy of Technology: The Empirical Turn*. Indiana University Press, Bloomington (USA): 119-146.
- Veruggio, G. 2005. *The birth of roboethics*. In *IEEE International Conference on Robotics and Automation, Workshop on Robo-Ethics*. Barcelona: 1-4.
- Wallach W., Allen C. 2009. *Moral Machines. Teaching Robots Right from Wrong*. Oxford University Press.
- Zampieri, A. 2015. *Latour in azione*. In *Tecnoscienza - Italian Journal of Science and Tecnology Studies*, n. 6: 61-88.
- Zimmerman E. J., 2015. *Machine Minds: Frontiers in Legal Personhood*.

