RELAZIONE DI MARIA BEATRICE MAGRO

Biorobotica, robotica e diritto penale

1. Biorobotica, interfacce cervello-macchina e potenziamento umano: filosofia precauzionale e euristica di avversione al rischio.

Neuroscienze, genetica comportamentale e diritto penale. Nell'ambito delle neuroscienze, molto interesse suscitano recenti studi di biologia molecolare e di genetica comportamentale che, partendo dal presupposto che il processo volitivo ha una base biologica, sono volti ad individuare rispettivamente il genoma umano e l'influenza del patrimonio genetico sul comportamento e sulla personalità dell'uomo. In particolare si ritiene che un'influenza sul comportamento criminale potrebbe essere esercitato da un tipo di geni, c.d. di suscettibilità, come il MAOA, nel senso che se pure non in termini assoluti, i soggetti che li possiedono, specie se sottoposti ad esperienze stressanti, hanno una probabilità maggiore di svilupparlo.

Queste scoperte scientifiche dissolvono un pregiudizio culturale assai radicato: la separazione tra corpo e anima, tra mente e cervello, tra *res extensa* e *res cogitans*. Il dualismo cartesiano prende il posto ad un dualismo "di proprietà" alla Karl Popper, secondo cui corpo e mente hanno la stessa materia, ma diverse proprietà. I rischi di questa concezione riduzionista biologica-materialista sono un rigoroso ed infallibile determinismo materialista. D'altro lato vengono rifiutati approcci ed interpretazioni rigorosamente biodeterministe anche dai più accesi sostenitori, sulla base dell'argomento secondo cui non esistono cervelli identici e quindi individui identici che agiscono in modo identico, posto che sia le neuroscienze che la genetica comportamentale non conoscono modelli di sperimentazione scientifica sul comportamento umano. Ciò dovrebbe mettere al riparo da ogni riduzionismo e da una concezione troppo meccanica della nostra vita.

In ogni caso, la scoperta di una base biochimica alla base del processo decisionale, insieme a studi di filosofia della mente, mettono in discussione importanti caposaldi della cultura occidentale ed incidono su importanti categorie del diritto e del processo penale. Viene in rilievo la categoria della imputabilità e, a monte , il concetto di libero arbitrio[1]: si propone un concetto di libero arbitrio come "potere di veto", come capacità di inibire il comportamento impulsivo, quindi un concetto "sociale" e "adattivo di libero arbitrio, ovvero frutto e risultato di un processo di condizionamento culturale, razionale, una interazione tra componenti genetiche ed ambientali che operano a diverso livello (per contro, vi sono coloro che sostengono che la volontà cosciente sia un'illusione). Ne deriva una definizione della capacità di volere come capacità di controllare l'impulso motorio e della capacità di intendere come comprensiva di empatia, pensiero morale e ragionamento controfattuale; ma tali risultanze possono influire sulla capacità a stare in giudizio, sulla valutazione della prova dichiarativa, nell'ambito dei reati in materia di abuso di sostanze stupefacenti, con finalità di prevenzione del reato, nella fase

[1] Occorre richiamare il celeberrimo esperimento di Libet il quale dimostra che il soggetto agente è consapevole della decisione di muovere un arto solo dopo che il cervello si è attivato per avviare il movimento; la consapevolezza dell'azione avviene circa mezzo secondo dopo l'instaurarsi del "*readiness potential*" (potenziale di prontezza); il processo volitivo si avvia inconsciamente e l'azione inizia prima che l'individuo ne acquisti consapevolezza. secondo Libet il libero arbitrio non descrive un processo decisionale volontario, ma si esprime nel controllarne il risultato; quindi le "decisioni coscienti e volontarie" non sono determinazioni all'agire, ma funzioni coscienti di veto rispetto ad impulsi di natura istintiva e biologica.

esecutiva della pena, laddove la concezione preventiva della pena esprime quella funzione inibitoria, di orientamento culturale e di veto che agisce a livello neurologico sulle funzioni di controllo, avvalorando quella concezione della colpevolezza in senso normativo, oltre quella della colpevolezza in senso psicologico, che sembra più radicata su un sostrato biologico.

La robotica e la biorobotica e le c.d. interfacce uomo-macchina. Il campo della robotica e della biorobotica è per eccellenza interdisciplinare, in quanto in esso convergono ricerche d'informatica, di ingegneria, di matematica elettronica, neuroscienze, biologia. La robotica comprende anche lo studio delle Intelligenze Artificiali, ossia la costruzione di macchine capaci di sentire, adattarsi all'ambiente, di imparare, dotate di capacità evolutiva e che persino sembrano "capaci di empatia". Queste macchine presentano le seguenti caratteristiche: sono interattive, reattive all'ambiente, agiscono in modo autonomo e imprevedibile e non determinato, flessibile e influenzabile; sono quindi dotate di autonomia, interattività, adattabilità, in quanto sono in grado di migliorare le loro *performance*. L'aspetto più attuale e con maggiori risvolti pratici della robotica riguarda la biorobotica, ovvero la combinazione di parti robotiche nel corpo umano (computer indossabili, cellulari telefonici, impianti di computer nel corpo) degli arti bionici artificiali, della interfacce uomo-macchina e nel potenziamento delle facoltà attentive e di memoria dell'uomo mediante impianti cerebrali su soggetti sani a scopi terapeutici o migliorativi.

Infine un aspetto nuovo è quello della biologia sintetica, ovvero dell'uso di nuove tecnologie in cui componenti molecolari naturali o sintetiche sono combinate e riorganizzate in modo da creare circuiti genetici nuovi e quindi nuovi organismi (le c.d. reti neurali artificiali).

La biorobotica consiste nel fenomeno della ibridazione uomo-macchina ovvero dell' innesto nel corpo umano supporti informatici, *hardware*, *software* e robot con finalità terapeutiche o di potenziamento fisico. Il dato sorprendente, che distingue l'ibridazione uomo-macchina dalle altre protesi non cibernetiche risulta proprio nell'interazione tra sistema nervoso, impulsi cerebrali e animazione della protesi, con possibilità di innescare flussi non solo in uscita (cervello-terminazioni nervose-chip-braccio robotico), al fine di comandare il movimento della protesi attraverso gli impulsi cerebrali, ma anche in entrata (braccio robotico, chip, terminazioni nervose, cervello), al fine di restituire al soggetto la percezione del movimento. La biorobotica si avvale infatti delle c.d. interfacce cervello-macchina: canali che offrono la possibilità di influenzare gli stati mentali/di coscienza di una persona e che permettono di trasmettere direttamente al cervello dei segnali elettrici esterni. L'arto bionico è in grado di riconoscere la volontà del soggetto ed eseguire gli ordini motori del cervello in tempo reale. Le interfacce tra il cervello umano e una macchina consentono di leggere e utilizzare i segnali neurali associati all'attività cognitiva per controllare un arto artificiale o la traiettoria di una piattaforma robotica mobile. Esistono anche interfacce cervello-macchina che, convogliando segnali verso il sistema nervoso centrale o periferico di un essere umano, ne modificano significativamente l'attività, come avviene nel caso delle interfacce usate per il controllo del tremore in soggetti affetti dal morbo di Parkinson. Queste ricerche bioniche si propongono soprattutto di ripristinare o di vicariare funzioni senso- motorie perdute, ma aprono la strada al potenziamento di apparati senso-motori e cognitivi che funzionano regolarmente. Sul piano del potenziamento fisico e cognitivo, si annoverano l'applicazione di *hardware* interni, *brain* computer interni, che creano legami più intimi tra sistemi di *software* e individui: si tratta di computer indossabili, a contatto con il corpo, incorporati nell'uomo, controllati direttamente attraverso l'attività cerebrale, attraverso elettrodi impiantati stabilmente nel cervello.

Potenziamento umano, fisico e cognitivo: l'uomo oltre le macchine. Il tema dell' *anhancement* umano concerne le tecniche di potenziamento sul piano cognitivo- intellettivo, e sul piano fisico (del miglioramento estetico o nelle competizioni sportive) o della estensione delle aspettative di vita. L' *anhancement* cognitivo si definisce come una amplificazione, un'estensione, o un potenziamento, interno o esterno, dei sistemi e processi mentali e neurologici di immagazzinamento, organizzazione ed elaborazione delle informazioni. Incide quindi su percezione, attenzione, comprensione e sulla memoria. Con l'avanzare delle neuroscienze cognitive, le possibilità di miglioramento delle capacità cognitive è in costante espansione. Eppure fino ad oggi, è stato il progresso della tecnologia di *computing* e delle tecnologie informatiche a produrre ottime *performance* sulla capacità di elaborare informazioni. *Hardware* e *software* sono in grado di dare agli esseri umani abilità cognitive che per molti aspetti superano di gran lunga quelli del cervello biologico. Siamo abituati a ricorrere a *hardware* esterni per potenziare le nostre capacità cognitive, facendo uso di calcolatrici, personal computer, che eseguono per noi compiti di *routine* per rendere afferrabili una quantità di dati che i sistemi percettivi umani non potrebbero gestire.

La nuova sfida è oggi quella di migliorare le capacità cognitive umane e di metterle al passo con quelle informatiche delle macchine, in modo da superare i limiti della mente umana e del corpo umano. I supporti informatici forniscono una sorta di estensione esterna della nostra mente, consentendoci di conservare tutto ciò che non potremmo conservare, e di dirigere i nostri sforzi in compiti più complicati rispetto all'immagazzinamento delle informazioni.

Questioni giuridiche: il caso del cattivo funzionamento dell'arto bionico. L'avvento della cibernetica e l'ibridazione uomo-macchina solleva problemi giuridici di non poco conto. Tra questi v'è anche il profilo delle responsabilità verso terzi (civili e penali) per i disfunzionamenti dei risultati dell'ibridazione. La reale presenza dei fenomeni ibridativi uomo-macchina (cyborg) deve farci riflettere anche sui problemi giuridici connessi al funzionamento o al disfunzionamento di queste nuove tecnologie ed al regime giuridico di responsabilità per danni arrecati ai soggetti ibridati, ai prossimi congiunti ed a soggetti terzi da parte di un soggetto che ha parti bioniche controllate a livello neurologico. Il profilo di responsabilità penale, qualora siano stati realizzati reati, solleva i seguenti interrogativi: è possibile affermare se l'incidente sia stato causato da un problema di controllo degli arti artificiali sulla base di una ricostruzione scientifica delle modalità di funzionamento degli arti bionici? A chi attribuire la responsabilità qualora venga cagionata la morte di un terzo, se non è possibile accertare un vizio di costruzione o funzionamento dell'arto biologico? L' "azione" compiuta con un arto artificiale può essere considerata "cosciente e volontaria"? Più a monte, si collocano i soliti profili relativi ai limiti alla disponibilità del corpo e della integrità fisica, ove il tema si intreccia con la distinzione tra intervento terapeutico (e relativo statuto epistemologico) e intervento migliorativo a scopo non terapeutico.

Le tecnologie di potenziamento umano e la distinzione dal concetto di terapia. Le c.d. *enhancement technologies*, ovvero le tecnologie biomediche, sono volte non già a curare una malattia ma a potenziare le normali funzioni fisiologiche dell'essere umano. Si pone quindi, non tanto sotto il profilo pratico, ma quello bioetico e giuridico, il problema di distinguere tra *enhancement* medica, cioè a scopi terapeutici e potenziamento puro. Un intervento che mira a correggere una specifica patologia, una disfunzione, o un difetto di un sottosistema cognitivo può essere definito come terapeutico. Un miglioramento invece non è un intervento che serve a porre rimedio ad una disfunzione specifica. Una persona cognitivamente potenziata, è qualcuno che ha beneficiato di un intervento che migliora le prestazioni di un sottosistema senza

correggere alcune specificità identificabili come patologia o come disfunzione del sottosistema. Il potenziamento, per definizione, va quindi oltre gli scopi della medicina e il suo statuto epistemologico (consenso, informazione, disponibilità dello stato di salute). Il che suscita serie perplessità in ordine al suo fondamento giustificativo: il modello terapeutico gode di una giustificazione radicata nella società, anche qualora si tratti di terapie sperimentali, in ragione dell'assenza di valide alternative terapeutiche, della presenza di un consenso informato, e proporzionalmente alle probabilità di successo della terapia, diversamente dalle tecnologie di potenziamento che parrebbero sfornite di tale modello giustificativo fondato sull'art. 32 Cost..

Il concetto di salute in senso soggettivo. Il punto è che è assai difficile concettualmente tracciare una distinzione tra terapia e potenziamento, e si potrebbe perfino sostenere che essa manca di significato pratico. Inoltre, ogni terapia comporta un miglioramento, anzi si può affermare che obiettivo della scienza medica è migliorare la condizione umana. Anche l'accezione del concetto di salute in senso soggettivo, come sinonimo di completo (ed utopico) stato di benessere, comprensivo non solo di una dimensione fisica ma anche di quella psichica che va al di là di quanto necessario per ristabilire uno stato di alterazione, sembrerebbe annullare la differenziazione concettuale tra terapia e potenziamento puro non terapeutico: l'obiettivo del raggiungimento di uno stato di benessere mentale, tale da migliorare le relazioni umane, le proprie prestazioni lavorative, in generale i propri rapporti personali, è un obiettivo della nostra Costituzione e dell'evoluzione scientifica stessa, quale strumento di valorizzazione della personalità umana. Vi sono quindi grandi difficoltà sul piano pratico a differenziare tra recupero capacità perse e ampliamento di capacità. Possiamo notare che la terapia di miglioramento riecheggia la dicotomia corrispondente tra standard attuale di medicina contemporanea e la *medicina -come - potrebbe -essere- praticata nel futur*o. Lo standard della medicina contemporanea include molte pratiche che non mirano a curare le malattie o lesioni: per esempio, la medicina preventiva, le cure palliative, la medicina dello sport, la chirurgia plastica ed estetica, dispositivi contraccettivi e i trattamenti di fertilità. In secondo luogo, non è chiaro come classificare gli interventi che riducono la probabilità di malattia e di morte: la vaccinazione può essere vista come un potenziamento del sistema immunitario o, in alternativa, come intervento terapeutico preventivo. Analogamente, un intervento che rallenta il processo di invecchiamento può essere considerato sia come un miglioramento o come intervento terapeutico preventiva che riduce il rischio di malattia e invalidità? Si propone il criterio del "collegamento o vincolo di interiorità" per differenziare tecnologie a fini di potenziamento e a scopi terapeutici: ma come qualificare la chirurgia laser che risolve problemi della vista o l'uso di lenti a contatto o di occhiali? Anche questi miglioramenti non interni ma esterni dovrebbero essere qualificati come potenziamento?

Invero, le soluzioni etiche o razionali devono essere differenziate e non una sola monolitica posizione: ogni tipologia di intervento migliorativo merita una presa di posizione a sé.

Gli effetti delle tecnologie di potenziamento): effetti isolati all'individuo e effetti estesi all'ambiente e alle generazioni future. Vi sono tipologie di potenziamento, quali'impianto di cellule neuronali, la biorobotica, il miglioramento genetico, che hanno effetto solo sul soggetto direttamente coinvolto, anche se in modo permanente (ad esempio gli impianti neuronali, ovvero tutti quegli interventi medici o biologici dei fenotipi che non si trasmettono agli eredi; la stimolazione magnetica transcranica, che agisce stimolando elettricamente la corteccia cerebrale, così migliorando le prestazioni motorie, di coordinamento, della memoria; la biorobotica). Vi sono inoltre tecnologie di *enhancers* con effetti che si estendono sull'ambiente e che modificano l'assetto genetico di individui futuri in modo permanente, in quanto incidono

sul patrimonio genetico dell'individuo (ad esempio i miglioramenti genetici). Rispetto questa tipologia si pongono in modo più stringente problemi di *governance* e di regolamentazione di situazioni caratterizzate da profonda incertezza sulle conseguenze dell'agire umano sotto il profilo delle loro interazioni rispetto gli altri individui e rispetto l'ambiente.

Gli effetti delle tecnologie di potenziamento (sull'individuo, sull'ambiente e sulle generazioni future) e la condizione di "doppia incertezza epistemica". Il tema del potenziamento apre una nuova frontiera del dibattito bioetico, destando una serie di preoccupazioni etiche dovute alla situazione di doppia incertezza epistemica che lo connota. L'effetto della stimolazione neurologica sul controllo della propria attività cerebrale da parte del soggetto è ancora oggetto di studio e sperimentazione e non vi sono linee guida o modelli scientifici sperimentati e corroborati ad un grado di accettazione e di alto riconoscimento nell'ambito della comunità scientifica. Dunque, il primo livello di incertezza epistemica è di tipo scientifico. Infatti, da un lato esse costituiscono delle tecniche relativamente nuove, alcune futuribili, altre allo stadio di prima sperimentazione, pertanto non vi è certezza o meglio, non vi è alcuna prova statistico scientifica sui loro possibili utilizzi in termini di efficacia, sicurezza, sia nel breve che nel lungo periodo. Anzi, deve dirsi che, al momento, il dibattito si fonda più sulle promesse dell'ingegneria genetica che sulle reali possibilità di intervento, per limiti di natura tecnica, in quanto non siamo ancora in grado di manipolare tratti influenzati da più geni. Anche gli studi scientifici sull'uso di terapie farmacologiche non sono corroborati da un numero consistente di casi clinici e quindi sono privi di attendibilità e ciò rende la letteratura scientifica sul tema una guida assai labile. Vi è quindi una condizione di incertezza della scienza circa i possibili effetti benefici (sia nel lungo che nel breve termine) sul singolo individuo e una totale incapacità di previsione di quelli dannosi per la salute umana in genere nel lungo termine. La sfida tecnologica non fornisce un sapere scientifico certo e incontrovertibile: vi sono riscontri scientifici di minoranza, i riscontri scientifici circa la beneficialità/dannosità non sono sufficientemente fondati o ancora *in itinere*. I dubbi aumentano quanto più tali tecnologie sono destinate ad incidere direttamente sulle facoltà cerebrali e fisiche dell'individuo, a volte con effetti permanenti, a volte con effetti limitati nel tempo. La scienza perde il suo ruolo di sapere informatore della politica e del diritto e al contrario, le pratiche scientifiche e tecnologiche sono tra i principali responsabili dell'incertezza del mondo, in quanto produttrici di innovazioni, i cui effetti collaterali possono essere negativi ed inaspettati, ma anche benefici.

Questioni etiche: il dibattito etico tra sostenitori e detrattori. Il secondo livello di incertezza concerne la valutazione sul piano morale ed etico degli effetti sull'individuo, sulle generazioni future e sull'ambiente. Si pone il problema del come valutare, sotto il profilo etico, l'affinarsi di tecniche di potenziamento. La scelta di alterare il proprio stato mentale o il proprio corpo nella ricerca di un sé più appagante, di migliorare se stessi, le proprie prestazioni, è moralmente e giuridicamente accettabile e rientra nella disponibilità dell'individuo? L'autenticità (è più autentica la vita di un individuo che ha migliorato le proprie potenzialità?), prima ancora della dignità costituisce un valore messo a repentaglio da queste tecnologie? Sul piano bioetico e filosofico l'evoluzione tecnologica mette in discussione il concetto di "umano" e di "umanità". Si schierano, contrapponendosi, da un lato i tecnodeterministi e pessimisti, avversi ad un mondo dominato dalla tecnologia, dall'altro i tecnocratici che viceversa professano un illimitato dominio della scienza. In ogni caso, a prescindere da queste contrapposizioni, occorre prendere atto che la tecnologia provoca condizionamenti culturali, che risolve problemi ma ne crea di nuovi, che incide sulla visione umana del mondo e che interferisce nei sistemi di interazione interpersonale. La tecnologia non è neutrale: perciò occorre porre una questione di formazione, informazione e responsabilità nell'uso delle tecnologie. La ricostruzione dei parametri di

definizione del concetto di umano e le sue possibili trascendenze è iniziata come processo qualche anno fa nell'ambito del doping sportivo e attualmente nella ricerca e combinazione di parti umane e parti macchine e quindi nella creazione di cybor o ibridi. È opportuno chiedersi se sia nella nostra disponibilità modificare la nostra dotazione "naturale" di capacità senso-motorie e cognitive attraverso interventi bionici. Una risposta positiva a tale quesito suscita a sua volta domande sulla persistenza dell'identità personale, prima e dopo l'intervento bionico. Più specificamente: una modifica delle funzioni mentali, sensoriali o motorie resa possibile dai sistemi bionici può indurre una modifica dell'identità personale? Quali alterazioni della continuità degli stati cerebrali/mentali di un soggetto possono essere accettabili da un punto di vista della tutela della identità personale? Il dibattito filosofico contemporaneo sulle condizioni psicologiche o fisiche di persistenza dell'identità personale fornisce strumenti concettualmente rilevanti per affrontare questi problemi ontologici. Il concetto di esistenza umana e di identità umana viene interpretato in termini evoluzionistici. L'evoluzione- miglioramento diviene la metafora dell'esistenza umana. All'interno del modello evolutivo, quello che noi consideriamo umano è un mero artefatto di un dato momento storico.

Il dibattito attuale vede fronteggiare sostenitori che articolano teorie e argomentazioni a difesa del potenziamento (teoria libertaria, teoria utilitarista, dei c.d. "tecnofili" Nicholas Agar, Allen Buchanan, Nick Bostrom, John Harris, Julian Savulescu) e d'altrolato detrattori o c.d. "bioconservatori" (Francis Fukuyama, Jurgen Habermas, Leon Kass, Michael Sandel) che analizzano le possibili minacce aperte all'uomo e alle generazioni future. In questa prospettiva il problema del potenziamento viene affrontato nell'ambito delle teorie della giustizia, con specifico riferimento al problema della disuguaglianza (potenziati/non potenziati), delle possibili ripercussioni dell'accesso al potenziamento sulla non accettazione della disabilità (dato il divario sempre crescente tra dis-abili, abili, super-abili o potenziati). Ci si chiede, quali le conseguenze sociali, in termini di disuguaglianza, di accentramento di potere, di potere tecnocratico, nel lungo periodo? Queste tecnologie accentuano o creano diseguaglianze sociali o realizzano un obiettivo di miglioramento morale e sociale? I disabili, i meno attrezzati cognitivamente verrebbero ancor più discriminati rispetto ad individui potenziati? E se tali tecnologie fossero accessibili a tutti, non si correrebbe il rischio opposto di un livellamento delle condizioni?

<u>L'incertezza scientifica e il modello di regolazione: il fondamento del divieto ultraprudenziale: il divieto di cmportamenti i cui effetti sono benefici</u>. Se gli effetti ipotizzabili delle tecnologie di potenziamento bionico e biorobotico sono dannosi (per l'individuo, per le generazioni future o per l'ambiente) o in parte benefici (per l'individuo) ma dannosi non le generazioni future e per l'ambiente, l'eventuale penalizzazione delle tecnologie di potenziamento umano di biorobotica trova fondamento nel paradigma classico di legittimazione del divieto di cagionare danno a terzi o a se stessi. Tuttavia, se come si diceva, vige una totale incertezza scientifica sulla dannosità di tali interventi, e di essi si accerta solo la beneficialità sotto il profilo fisico-psichico, sia pure sul solo singolo individuo su cui vengono praticate, si pone il problema sotto il profilo etico se il biopotenziamento non debba costituire un dovere morale e sociale (se poi gli effetti benefici per tutti, sia a livello personale che generale). La ipotetica e futuribile penalizzazione di queste tecnologie in questo caso rifletterebbe una inversione del paradigma paternalistico forte e debole: si vieta ciò che è bene per l'uomo sul presupposto che il potenziamento è espressione di motivi futili. Si va oltre il divieto di cagionare danno a se stessi o ad altri: si intravede una norma di divieto ultraprudenziale che va oltre al modello di giustificazione del paternalismo forte e che recepisce l'euristica di precauzione forte.

Il principio di precauzione e l'euristica della avversione al rischio. In questo contesto culturale compare sulla scena il principio di precauzione. Il principio di precauzione postula una strategia di gestione di analisi del rischio nelle situazioni di incertezza, di complessità della scelta e di limiti della conoscenza umana sui possibili effetti della tecnologia. Intesa in senso forte, la filosofia precauzionale impone la regola della totale astensione in presenza di qualunque fattore di rischio potenziale, riguardo al quale non si abbia certezza delle conseguenze, né della loro portata, paralizzando qualunque decisione che non si risolva in *nihil agere*, finché non venga addotta la prova della totale innocuità dell'attività intrapresa. In senso meno forte, il principio di precauzione impone di adottare tutte le misure necessarie per azzerare o contenere i rischi connessi allo svolgimento di un'attività, soprattutto se relativi a beni di rilevanza primaria, quali la salute umana o l'ambiente. L' intensa produzione normativa europea degli ultimi anni appare ispirata, in diversi contesti di "rischio" alla realizzazione di un "elevato livello di tutela". Il miglior strumento per la realizzazione di questo obiettivo è costituito dalla progressiva estensione dell'applicazione del principio di precauzione ad ambiti di tutela sempre più ampi e ulteriori rispetto al settore originario della tutela ambientale, trovando applicazione nelle materie incidenti sulla tutela della salute umana. La norma precauzionale, recependo una euristica di *loss adversion*, si dirige in direzione del divieto o della limitazione della attività, sollecitando la decisione che minimizza il rischio di danno, piuttosto quella che valorizza il rischio di benefici, a prescindere da un giudizio scientifico circa la probabilità che la misura riduca o escluda la verificazione del danno. In tal modo il principio di precauzione vieta cautelativamente ciò che, forse e successivamente, si rivelerà del tutto innocuo (se non addirittura utile e benefico). La filosofia di pensiero e di regolazione pubblica ispirata al principio di precauzione concerne situazioni in cui è impossibile stabilire alcuna probabilità di verificazione di un determinato danno, incuneandosi nel momento della valutazione dei rischi, ovvero nelle politiche di gestione e di percezione del rischio e nelle richieste di sicurezza sociale, e così sganciando l'intervento prudenziale da un sostrato di verità- necessità ed orientandolo verso la scelta di cautele adeguate di contenimento di un rischio (che è solo presuntivamente supposto) ad un livello ritenuto giuridicamente accettabile. L'applicazione del metodo precauzionale svolge il ruolo di criterio guida delle decisioni assunte in condizioni di incertezza, in direzione della prevenzione delle conseguenze peggiori tra tutte le opzioni disponibili e del contenimento del rischio, quando le conoscenze scientifiche non consentano di escludere, ma nemmeno provano, il carattere dannoso dell'attività svolta.

3. La robotica, i droni e le Intelligenze Artificiali.

La robotica e l' intelligenza artificiale (forte o debole): nessuna differenza tra robot e mente umana? La costruzione di macchine dalle prestazioni *human like* si configura come uno dei nodi più difficili della scienza. In questa prospettiva di ricerca si colloca la disciplina delle Intelligenza Artificiale. L'espressione "Intelligenza Artificiale" è stata coniata nel 1956 da Mc Charty durante un seminario interdisciplinare. Essa indica sia un termine che una disciplina che parte dal presupposto secondo cui si possa emulare ogni aspetto dell'intelligenza umana. Sebbene sia state proposte parecchie, talora anche contraddittorie, definizioni di tale disciplina, ciò che le accomuna e connota lo spirito della IA è quello di imitare, riprodurre per mezzo di macchine elettroniche l'attività mentale umana, ovvero quella che costituisce la facoltà più essenziale dell'uomo.

Invero l'idea che le macchine possano compiere operazioni proprie dell'intelligenza umana, o almeno alcune funzioni di essa, risale al seicento ed oggi si presenta nella forma del computer, quale incarnazione di "macchine pensanti".

Robot intelligenti e dotati di autonomia decisionale. L'utilizzo di agenti artificiali complessi e organizzati come *decision making*. Nelle organizzazioni complesse esistono sistemi informatici e tecnologici intelligenti in grado di prendere delle decisioni autonome (decisioni che non prevedono alcuna esplicita autorizzazione da parte di un essere umano) e che operano come supporto a manager soprattutto nella gestione di infrastrutture tecnologiche ad alto rischio. Le organizzazioni artificiali sono strutture mutuate da quelle umane, all'interno delle quali ogni agente intelligente, denominato *personoide* occupa un ruolo ben preciso (che comprende accesso a informazioni, doveri responsabilità) e producono diversi processi decisionali. La *decision making* è un'attività di ragionamento che implica la necessità di compiere una scelta, provocata da una informazione ricevuta; esso inizia quando sono sconosciuti i criteri di scelta, o quando non si conoscono le alternative, e termina quando si compie una scelta*;* è un processo organizzativo guidato da i*nformazioni*, *conoscenze* e *preferenze* per il raggiungimento di un Goal predefinito.

Attualmente le IA sono applicate alla robotica cioè nella costruzione di macchine intelligenti capaci di interagire fisicamente con il mondo esterno e di manipolare oggetti, spesso dotate di un sistema esperto cioè di sistemi computazionali dotati di conoscenze specialistiche. Ma anche nelle applicazioni più sofisticate ed evolute, le IA hanno mostrato alcuni limiti di fondo che esamineremo oltre.

> Gli ambiti di applicazione. Gli ambiti di applicazione delle Intelligenze Artificiali sono i più disparati: si va dagli agenti artificiali e reti neurali artificiali in ambito economico (i robot agiscono come agenti economici intelligenti privi di emozioni, e quindi realizzano le condizioni di equilibrio competitivo nella domanda e nell'offerta, ovvero il modello di efficienza distributiva perfetta), creazione di robot di servizio e assistenziali (badanti robot). In medicina, con l'avvento della chirurgia mini-invasiva, i robot sono utilizzati nella c.d. telechirurgia. Applicazioni assai vaste sono in ambito militare, con l'uso di droni come armi e di robot soldati. In proposito si richiama il monito di molti scienziati a vietare o a usare cautele nell'uso di robot autonomi letali, come sistemi intelligenti in ambito militare in cui la decisione è autonoma, le cui conseguenze possono essere assai devastanti. Questi robot armi possono innescare armi letali senza un intervento umano nel processo decisionale. La Corte penale internazionale ha dichiarato l'illegittimità dell'uso di armi robot e droni.
>
> Applicazioni nel lavoro, ma anche in casa, come robot di servizio o ludici (un esempio di questa proiezione-attribuzione affettiva è offerto dal robot cane *Aibo*, di cui la Sony ha interrotto la produzione dopo averne costruito, dal 1999 al 2006, oltre 150.000 esemplari), o come supporto tecnologico nelle grandi organizzazioni complesse finalizzati a esprimere decisioni. In ambito economico si obietta che non è l'abilità dell'agente artificiale a realizzare l'equilibrio del mercato, ma l'intelligenza implicita del meccanismo e delle regole di mercato che realizzano l'equilibrio, se non sono ostacolate da fattori esogeni (Sundel 2004).

Il dibattito sulle IA risponde alle domande: "le macchine possono pensare?", "è possibile riprodurre l'intelligenza umana?" "il pensiero può avere sede fuori dal cranio umano?". Ed ancora: le Intelligenze artificiali, in quanto dotate di autonomia, sono soggetti realmente capaci di manifestare una "propria" intenzionalità e quindi un proprio agire? Sono soggetti capaci di esprimere decisioni? Sono capaci di provare emozioni e di manifestare intenzionalità? Il paradigma teorico di base è quello dell'analogia mente-computer e dell'isomorfismo mente computer, da non intendersi però in senso ontologico-strutturale, ma in modo analogico e

funzionale. Su questi temi si sono confrontati filosofi della mente di questo ultimo mezzo secolo, contrapponendo un approccio funzionalista che valorizza le funzioni svolte dalla mente umana e dalla macchine, in quanto in grado di svolgere funzioni simbolico computazionali da un approccio più critico e strutturalista.

La nozione di Intelligenza Artificiale; la IA forte e il "gioco dell'imitazione" di Turing. a) intelligenza come reattività comportamentale. Quali sono le caratteristiche che connotano la soggettività (anche non umana)? Per definire se una macchina è intelligente si ricorre al test di Turing detto anche "gioco dell'imitazione". Il test di Turing dimostra che il computer è un sistema computazionale (programma *software* più eventuale supporto *hardware*) che interagisce con l'ambiente circostante, che reagisce agli stimoli di tale ambiente (reattivo); che è in grado di prendere decisioni, e di conseguenza di agire, in modo autonomo, per raggiungere un obiettivo predefinito o negoziato (proattivo); è in grado di comunicare (coordinarsi, cooperare, negoziare) con altri agenti e/o con esseri umani (capacità di interazione sociale).

Se si assume un concetto di intelligenza di tipo comportamentale e interattivo e se per psicologia e intendiamo una scienza che descrive il comportamento di qualsiasi specie di sistemi il cui comportamento sia suscettibile di analisi e di interpretazione in termini di costrutti comportamentali di base quali stimolo, risposta, impulso, saturazione, allora potremmo dire che i calcolatori possiedono una psicologia perché obbediscono a leggi psicologiche e sono agenti intelligenti. In altre parole, l'intelligenza artificiale è una macchina intelligente che manipola e usa simboli in modo formale e astratto.

Alan Turing è il fondatore della corrente di pensiero detta intelligenza artificiale forte che ritiene i calcolatori macchine capaci di esprimere autentico pensiero e in grado di produrre processi intellettuali identici a quelli umani. Non vi sarebbe alcuna differenza ontologico-qualitativa tra cervello umano e cervello elettronico e tra intelligenza umana e intelligenza artificiale, poiché la sola differenza risiederebbe nella sede o nel supporto fisico: la testa umana fatta di carne, ossa e altri materiali biologici e la struttura di un calcolatore fatto di metallo e energia. Parte integrante di quest'ottica funzionalista è l'assimilazione del cervello allo *hardware* (materiali duri: le parti meccaniche) e della mente al *software* (materiali morbidi: i programmi e istruzioni). Non conta la struttura ma la funzione ed esistono intelligenze in sé indipendentemente dalla sede fisica (il cervello- la macchina) in cui risiedono. Perciò un computer è paragonabile ad un essere umano quanto ad intelligenza, se le prestazioni svolte dal computer non possono essere distinte da quelle svolte dagli esseri umani (tesi funzionalista).

Detto altrimenti: lo stesso *software* può essere realizzato da tipi differenti di *hardware* (Putnam).

Tuttavia si osserva dai fautori della IA debole che tutto ciò ovviamente non comporta che i robot siano anche coscienti e cioè che abbiano una coscienza. Le macchine di Turing sono psicologicamente isomorfe all'uomo ossia hanno una identità di organizzazione funzionale ma da ciò non si può ritenere che abbiano una coscienza o una consapevolezza del proprio agire. Avere stati psicologici rilevabili *ab extra* non significa possedere autentica consapevolezza dei comportamenti. Anche l' uso di vocabolario antropomorfo nel descrivere le caratteristiche, le proprietà e il funzionamento delle Intelligenze Artificiali e dei sistemi informatizzati non riguarda tanto la macchina (il calcolatore) in sé, ma il modo in cui noi vediamo la macchina, e indirettamente noi stessi (Minsky rivendica la legittimità/utilità dell'uso di termini antropomorfici).

Seguendo questo ordine di idee si è perfino perseguito lo scopo conoscitivo di studiare, attraverso la costruzione di macchine intelligenti macchine, la mente umana e le sue modalità di funzionamento. La mente umana opera come entità formale ed astratta e gli stati mentali sono qualificati dalla loro funzione svolta e non dalla specificità materiale. Quindi è possibile supporre che gli stati mentali siano slegati dal sistema neurale organico e che sia riproducibile lo stesso sistema neurale. Si sviluppa il paradigma connessionista delle "reti neurali" e la tentativo pioneristico della cibernetica di riprodurre artificialmente l'architettura neuronale del sistema nervoso umano: se l'intelligenza è espressione dell'attività neuronale, la riproduzione artificiale di tale attività avrebbe per ciò stesso riprodotto l'intelligenza.

La confutazione del test di Turing: La IA debole e l'Intelligenza come intenzionalità e come "comprensione" disignificati. L'altra tesi, quella della intelligenza artificiale debole, ossia della non autenticità del pensiero meccanico e della diversità ontologica tra intelligenza artificiale e intelligenza naturale è stata sostenuta da autorevoli studiosi. Secondo la IA debole le macchine simulano e riproducono soltanto i processi intellettuali umani di quali ne rappresentano delle copie. Weizenbuam nel libro "*Computer Power and Human Reason: From Judgment to Calculation*", discute i limiti dei calcolatori affermando che le visioni antropomorfiche dei computer sono delle ingiustificate riduzioni degli esseri umani. Afferma quindi una nozione di intelligenza come intenzionalità (dimensione essenziale della coscienza): caratteristica che contraddistingue certi stati mentali, quali le credenze, i desideri e le intenzioni, diretti verso oggetti e situazioni del mondo e che differenzia l'uomo dalle macchine. L'assunto fondamentale è l'impossibilità per una macchina computazionale di manifestare l'intenzionalità che caratterizza gli esseri umani e, sia pure in forme diverse, gli animali.

Il test della stanza cinese di Searle. Per contro, secondo Searle, l'intenzionalità è un dato di fatto empirico circa le effettive relazioni causali tra mente e cervello, che consente unicamente di affermare che sussistono certi processi cerebrali; ma l'esecuzione di un programma su un dato input non è di per sé una condizione sufficiente per l'intenzionalità. Il test della stanza cinese dimostra che il computer può anche essere un ottimo manipolatore formale di simboli, ma questa attività non comprende quella della "comprensione" dei significati dei tali simboli e di conseguenza non coincide con l'intelligenza umana, la quale è sempre intenzionale e cosciente. Il sistema simbolico-computazionale si comporta come "se" capisse il cinese, ma non riproduce l'attività umana, la simula perché manca la coscienza dei significati.

Searle prende in esame i lavori sulla simulazione della capacità umana di comprendere narrazioni, che richiede l'abilità di rispondere a domande che coinvolgono informazioni non fornite in modo esplicito dalla narrazione, ma desumibili da essa sfruttando conoscenze di natura generale. Risultato: la capacità (di un uomo/una macchina) di manipolare le informazioni ricevute secondo regole formali ben definite non è sufficiente a spiegare il processo di comprensione. Si afferma il carattere non intenzionale, e, quindi, semanticamente vuoto, dei simboli elaborati da un sistema artificiale e, in conclusione, che i processi mentali non possano essere ridotti a processi di natura computazionale che operano su elementi formalmente definiti. I robot intelligenti non hanno capacità di connessione e di critica, ovvero non hanno categorie concettuali.

Il senso comune. In più si fa notare che ai robot dotati di intelligenza artificiale, dotati di conoscenze altamente specialistiche, manca, al disotto di queste conoscenze, il livello di conoscenze comuni, il senso comune, ciò che tutti gli umani posseggono senza aver fatto studi particolari. Il senso comune è quello che consente di collegare conoscenze specialistiche di campi diversi e di affrontare i problemi e di risolverli senza la rigidità tipica dell'approccio

simbolico dell'intelligenza. Spesso una reazione intelligente ad una certa situazione è quella che sì, tiene in considerazione il contesto, ma che non è in grado di selezionare quale aspetto del contesto sia rilevante. A questo tipo di conoscenza e di approccio corrisponde il concetto di intelligenza emotiva quella che orienta verso il "senso comune": da qui il paradosso che le macchine riescono a riprodurre il più altro rigore logico ma non riescono a essere programmate su un livello minimo di intelligenza comune (non dimentichiamo gli studi sul ruolo delle emozioni nel processo decisionale e gli studi sull'intelligenza emotiva di Golemann).

Conclusione: contro i funzionalisti, occorre affermare lo strutturalismo: ciò che conta è la struttura e non la funzione ossia la sede dove si svolge l'attività non come si svolge. L'intenzionalità trova sede nel cervello, la mente è un prodotto causale del cervello. Ne segue quindi che qualunque meccanismo in grado di causare la mente e di produrre intenzionalità dovrebbe necessariamente possedere poteri causali simili a quelli del cervello, cosa che non è possibile.

La tesi secondo cui i programmi non sono menti implica una presa di posizione contro il paradigma funzionalista e la sua maniera di impostare il *mind-body problem*. Infatti l'ipotesi funzionalista secondo cui la mente sarebbe concettualmente e empiricamente separabile dal cervello è pericolosamente dualistica: una forma di dualismo che non ricalca quello tradizionale cartesiano che dichiara che si sono due generi di sostanze, ma cartesiana nel senso che afferma che ciò che è specificatamente mentale non ha alcuna connessione intrinseca con le proprietà del cervello. Sebbene la letteratura della IA contenga molte denunce contro il dualismo classico, anche essa propone un neo dualismo che pretende di sganciare i processi cognitivi dalla biochimica concreta delle loro origini materiali.

L'intelligenza emotiva, i robot e le emozioni. Infine, ricordiamo che vi sono ricerche finalizzate a dotare i robot di emozioni artificiali e di coscienza, sempre artificiale, rendendoli così sempre più simili a noi. Se oggi si progettano agenti capaci di *manifestare* emozioni (con l'espressione, con l'atteggiamento e così via), un domani si vorrebbero costruire agenti capaci addirittura di *provare* (oltre che manifestare) emozioni. La domanda è: si può dire che un agente artificiale manifesta emozioni quando si comporta in modi che, negli umani, presuppongono emozioni? Le emozioni sono per gli umani un tratto costitutivo fondamentale, inseparabile dalle altre nostre caratteristiche: sono strettamente intrecciate alla razionalità computante, ma anche alle funzioni fisiologiche, alla memoria, all'esperienza, sono profondamente innestate nel corpo, inteso sia come insieme di organi sia come depositario della nostra identità, dei nostri ricordi e della nostra storia. Le emozioni sono tanto pervasive che ogni nostro atto si colora di esse e ogni nostra relazione con noi stessi e con l'"altro" ne è condizionata. Ma nelle macchine non si si tratta di "vere" emozioni "artificiali", cioè qualcosa che vada oltre la *simulazione* dei sentimenti e sostenuta dalla nostra proiezione.

La computazione emotiva e nuovi problemi di tutela della privacy: le macchine intelligenti che "leggono" le emozioni umane. In questi casi si pongono principalmente problemi di tutela della privacy. La computazione emotiva è un sistema che ha accesso a vari indizi dello stato emotivo degli utenti, elaborati a partire sia da dati biometrici correlabili a risposte emotive (sudorazione, dilatazione delle pupille e così via), sia dall'esame di gesti, postura e altri comportamenti manifesti dei soggetti. Il sistema valuta questo complesso di informazioni per scoprire se un soggetto è interessato all'argomento, se si sta distraendo, se è scoraggiato. In base a tali supposizioni, il sistema sceglierà una forma di interazione con il proprio utente che sia contestualmente più appropriata a facilitare il processo di apprendimento. Non si può escludere che un tale sistema, una volta sviluppato, sia dolosamente utilizzato all'insaputa degli utenti, con la finalità di raccoglierne un

profilo emotivo e di sfruttare a fini commerciali le risposte emotive indotte in base alle informazioni raccolte. È dunque opportuno chiedersi in quali circostanze la raccolta e l'elaborazione di dati personali da parte di tali robot sia legittima ovvero si configuri come una minaccia per la dignità degli utenti, portando a una violazione del loro diritto alla sicurezza, alla riservatezza (privacy).

Lo statuto ontologico delle macchine. La responsabilità giuridica per l'agire o dell'agire del robot intelligenti. I robot come agenti non-umani o come strumenti? Dalle diverse visioni e concezioni sulla Intelligenza artificiale (strutturalista e funzionalista) derivano due opposte concezioni a proposito dello statuto ontologico delle macchine. Secondo la tesi strutturalista gli automi, anche molto intelligenti, tali che i loro comportamenti siano indistinguibili allo osservatore da quelli umani, sono comunque degli oggetti strumentali e dunque, sul piano morale e giuridico, meritevoli di tutela solo per il valore che esprimono; secondo la tesi funzionalista i robots intelligentissimi potrebbero essere considerati come dei soggetti e dunque meritevoli, sul piano filosofico, morale e giuridico, di una eventuale tutela per soggettività.

Si profilano due ordini di problemi. Il primo teorico filosofico, ossia la questione dello statuto ontologico di macchine particolarmente evolute: sono semplici oggetti, oppure travalicano la soglia dei requisiti minimi per il riconoscimento di un livello di soggettività? Il secondo pratico-funzionale-giuridico, ossia, la questione se sia opportuno e utile ai fini giuridici conferire a simili artefatti un livello di soggettività. Il problema quindi è stabilire se la tecnologia moderna ha creato una nuova tipologia di soggetto, il soggetto non-umano, l'agente non-umano per il quale è possibile declinare il linguaggio giuridico dei diritti e delle responsabilità (il primo a ipotizzare una loro soggettività anche giuridica è Putnam che nel 1960 già scriveva " è giusto che i robot abbiano i diritti civili ?).

Ovviamente dalla risposta a tali questioni deriva anche il tipo di 'diritto' e il tipo di 'tutela' applicabile a questi sistemi intelligenti artificiali (in tal caso possono essere oggetto di tutela penale). Le domande vanno scisse e affrontate separatamente. La prima si muove su un piano filosofico, la seconda su un piano giuridico, la prima aspetta una risposta a livello teorico, la seconda a livello pratico. Due sono i paradigmi che nella nostra cultura fondano l'aspettativa di una tutela, il valore che suggeriamo di intendere come elevatezza-complessità di funzioni, e/o utilità pratica, e/o rarità-originalità e la soggettività intesa in un crescendo di sensibilità emotiva, immaginazione e creatività, intellezione come pensiero logico, intenzionalità, coscienza, autocoscienza e autodeterminazione. Tutela per valore o per soggettività?

Torniamo alla questione di fondo: il robot agisce o è agito? È oggetto- strumento o soggetto, anche se mediato, o meglio, un "nuovo" soggetto per il diritto? Le Intelligenze artificiali e i robot intelligenti sono dotati di soggettività autonoma, come gli enti collettivi, e quindi sono soggetti responsabili?

Il problema della prevedibilità dei comportamenti dei sistemi robotici intelligenti. Nel 1950 Isaac Asimov elaboraò tre leggi fondamentali della robotica: 1) Un robot non può danneggiare con azioni o omissioni un essere umano né permettere che un essere umano riceva danno; 2) un robot deve obbedire agli ordini impartiti dagli esseri umani, a meno che tale ordini non contravvengano alla regola 1; 3) Un robot deve proteggere la propria esistenza, purché questa autodifesa non contrasti con la regola n. 2 e n.1. Queste tre leggi fondamentali sono ovviamente contradditorie: cosa accade se si ordini ad un robot di ferire un uomo per il bene di qualcun altro? Che cosa succede se il comandante ordina ad un robot poliziotto di arrestare un sospetto che fa resistenza all'arresto? E

ancora cosa accade se il robot chirurgo si ostina ad effettuare un intervento salvavita contro la volontà del paziente?

I problemi di comprensione e prevedibilità degli esiti di un intervento bionico rivelano un aspetto particolare della nostra limitata capacità di prevedere il comportamento delle macchine della robotica e dell'intelligenza artificiale. Tali limitazioni previsionali furono principalmente discusse in relazione alla formulazione generale dei teoremi di indecidibilità algoritmica; nuovi spunti per la riflessione epistemologica provengono da studi più recenti nel campo dell'apprendimento automatico. Le interfacce bioniche cervello-macchina, alle quali è stato accennato, richiedono un processo di addestramento per classificare e riconoscere i segnali neurali associati all'attività cognitiva di un essere umano. Non si può realisticamente pensare di fornire a un robot una specifica sufficientemente dettagliata di ciò che esso deve fare o aspettarsi di incontrare negli spazi non rigidamente definiti e controllati di una casa, di un ufficio o di altri ambienti nei quali si svolge la nostra vita quotidiana. Per questo motivo, un robot che sia abbastanza autonomo e adattabile per assistere un anziano nel suo appartamento, deve essere un robot capace di apprendere dall'esperienza. Ma questa esigenza deve confrontarsi con il fatto che i metodi di apprendimento automatico non sempre consentono di appurare se il robot abbia veramente imparato (o anche solo approssimato in modo soddisfacente) ciò che vogliamo insegnargli. Si pongono quindi problemi di valutazione dei risultati ottenuti nei processi di apprendimento automatico da parte di robot. La riflessione epistemologica sull'apprendimento automatico si inquadra dunque nella problematica più generale della prevedibilità dei comportamenti di sistemi robotici.

Gli studi dei metodi formali di verifica del *software* nascono dall'esigenza di verificare se ogni esecuzione di un determinato programma per un calcolatore soddisfa alcuni requisiti fondamentali. Più recentemente, queste metodologie sono state estese al problema di specificare e verificare le proprietà di sistemi, generalmente detti "ibridi", che comprendono varie tipologie di sistemi robotici. Alcuni risultati limitativi che sono stati ottenuti a proposito dei sistemi ibridi indicano che, in generale, non è possibile verificare con queste metodologie se un sistema robotico soddisfi o non soddisfi determinati vincoli spaziali o temporali nell'esecuzione di un dato compito.

Certo è che i robot agiscono in modo imprevedibile, in modo non regolato da rigorose leggi scientifiche: il comportamento dei robot può essere imprevedibile, in quanto le Intelligenze Artificiali hanno capacità di apprendimento, di adattamento all'ambiente, capacità decisionale e creatività rispetto una rosa di scelte.

Inoltre, l'interazione tra un sistema aperto di apprendimento condizionato proprio delle Intelligenze artificiali e il suo ambiente (comunicazione asincrona) è spesso soggetta a vincoli temporali che non consentono di assicurare la presenza di un essere umano nei cicli di controllo. Quindi certamente non sono completamente controllabili dall'uomo. Le Intelligenze Artificiali, i droni e robot intelligenti solleveranno pertanto insormontabili problemi di accertamento della responsabilità nell'ordinamento giuridico. Si tratta quindi di definire se tali macchine siano capaci di compiere azioni e comunque di capire a chi tali azioni, e le conseguenze di tali azioni, debbano essere imputabili. Ma, quando si parla di decisioni assunte da intelligenze artificiali, si tratta di vere decisioni, o di programmazione di una serie vincolata e limitata di scelte?

Quali sono le implicazioni pratiche normative delle riflessioni epistemologiche sulla nostra limitata capacità di prevedere il comportamento delle macchine? Il produttore o il

programmatore di un sistema intelligente che apprende dall'esperienza e che non ne previsto il disfunzionamento, anche in condizioni normali d'uso, è responsabile?

Il problema è che i robot pensanti non agiscono in modo prevedibile in quanto la capacità di apprendimento tratteggia uno spazio di imprevedibilità che si sottrae alle leggi scientifiche. Ma i robot sono "davvero" imprevedibili? Le caratteristiche di questi sistemi informatizzati esperti è che sanno reagire a situazioni previste o prevedibili, ma non a situazioni che, non essendo prevedibili da un essere umano, non hanno neppure essere previste nel programma e nelle regole che l'essere umano ha costruito e che costituiscono l'intelligenza artificiale. La loro presunta "imprevedibilità" – essendo macchine e non uomini - non è forse, a sua volta, prevedibile da parte del costruttore ? Una reazione intelligente ad una situazione non prevista e realmente nuova è quella che individua soluzioni sufficientemente accettabili.

I profili di responsabilità civile. Nell'ambito civile la questione si pone sul piano della responsabilità contrattuale per i vizi e conseguenze connesse ala progettazione, costruzione e utilizzo del robot; e della responsabilità extracontrattuale a carico del programmatore, del costruttore e dell'utilizzatore per i danni arrecati a terzi e non solo alle controparti del contratto. Gli schemi richiamati sono quelli della rappresentanza, del mandato, del contratto di assicurazione, forme di autenticazione, della responsabilità contrattuale ed extracontrattuale.

I profili di responsabilità penale: l'ipotesi fantastica del robot come agente non-umano responsabile penalmente. Sul piano penalistico si pongono due questioni: A) ipotesi dei robot intelligenti come agenti non umani, nuovi soggetti di diritto penale. Questa ipotesi solleva problemi di definizione della capacità soggettiva, della nozione di azione, della nozione di colpevolezza: i robot sono rimproverabili e motivabili? Esprimono una loro volontà, possono essere "autori" di reati e i loro agire si può definire cosciente e volontario? Il tema presenta un'analogia con quello della responsabilità penale-amministrativa degli enti collettivi e cozza con una concezione antropomorfica del diritto penale; gli enti collettivi non hanno corpo né anima, ma sono soggetti alla legge penale; i robot hanno un "corpo", una materia su cui far ricadere la sanzione penale (ad esempio, la disattivazione della macchina o la sua distruzione), ma non hanno un'anima, pur essendo dotati di autonoma capacità decisionale ed essendo in grado di interagire con l'ambiente. Si presenta lo schema della responsabilità dell'ente: l'agente umano agisce, l'ente ne risponde accanto l'individuo.

B) Ipotesi della responsabilità penale della persona fisica: se i robot sono mezzi e non agenti - non umani. Se si ritiene che i robot non siano agenti in senso stretto, si pone quindi il problema della responsabilità esclusivamente umana del programmatore o utilizzatore. A che titolo può essere imputata la responsabilità all'uomo, se si assume che il sistema informatico intelligente è un *decision maker*? Come abbiamo detto, si assume che i robot intelligenti agiscono in modo non programmato e imprevedibile e questa imprevedibilità pone problemi di attribuzione della responsabilità penale a carico dei programmatori, dei costruttori e degli utilizzatori. Se i robot e le Intelligenze artificiali sono artefatti in grado di migliorare le capacità degli esseri umani, mezzi di supporto e di comunicazione, si pone un problema inverso di imputazione penale: qui l'agente non--umano agisce, la persona fisica ne risponde.

La responsabilità penale del programmatore o utilizzatore: che modello di responsabilità penale a carico dell'agente umano? Responsabilità diretta o indiretta? A titolo di dolo, di colpa? A) a titolo di dolo, perché l'azione "volontaria" dell'agente non-umano incardina responsabilità sull'uomo e perché l'azione del robot si indentifica e rappresenta una *longa manus* della

persona umana. B) a titolo di colpa, individuata nel difetto di programmazione, costruzione, scelta, utilizzo, manutenzione funzionamento di robot intelligenti.

La colpa del programmatore: occorre elaborare un nuovo concetto normativo di colpevolezza che consente di elaborare un concetto di "colpa da programmazione", ove si può ipotizzare di prospettare una causa di esclusione della colpevolezza consistente nella predisposizione di misure di sicurezza che prevengano la realizzazione di reati da parte di robot intelligenti (in pratica che mettano in atto le tre leggi della robotica di Asinov). In definitiva, è colposo il comportamento de programmatore che non preveda l'imprevedibilità del robot intelligente! La diligenza e le regole cautelari nella programmazione, scelta, uso, controllo e manutenzione di robot. Anche la prevedibilità astratta e generica di futuri danni ancora non ben identificati incardina la responsabilità per colpa, anche quando non sono noti tutti gli anelli del processo causale: rischio di identificare la colpa con la precauzione e il prudenzialismo.

Quando l'agente artificiale è evoluto e gode di margine di creatività e decisionalità (è un *decision maker*) e non vi è neppure un visibile difetto nel controllo, uso, manutenzione dei robot, a che titolo sarà l'imputazione soggettiva del reato a carico della persona fisica? Nella letteratura anglosassone si parla di prevedibilità e di logico sviluppo prevedibile, similmente al 116 c.p. in tema di responsabilità del concorrente per l'evento diverso più grave non voluto, purchè prevedibile, e anche se non sono perfettamente noti tutti gli anelli della catena causale di produzione dell'evento. La differenza è che il programmatore o l'utilizzatore non concorrono nella realizzazione di un reato di base e quindi la regola del *versari in re illicita* non vale, ma la condotta della persona fisica crea una condizione di rischio lecito tollerato. La prevedibilità opera come componente della colpa, anche sulla base di leggi scientifiche non corroborate da studi e non consolidate, o nella totale ignoranza di leggi scientifiche di spiegazione causale.

Limiti alla responsabilità del programmatore, costruttore, utilizzatore sono naturalmente da individuare nel caso di uso improprio del robot, o nel caso fortuito o nella forza maggiore.