ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

DIRITTO E NUOVE TECNOLOGIE

CICLO XXVIII

Settore concorsuale di afferenza: 12/H3
Settore scientifico disciplinare: IUS/20

Responsabilita Penale e Automazione nell'E-Health

Presentata da:
Francesca LAGIOIA

Coordinatore Dottorato:
Prof. Giovanni SARTOR

Relatore:
Prof. Giovanni SARTOR

Esame finale anno 2016

«***HAL**: I enjoy working with people. I have a stimulating relationship with Dr. Poole and Dr. Bowman. My mission responsibilities range over the entire operation of the ship, so I am constantly occupied. I am putting myself to the fullest possible use, which is all I think that any conscious entity can ever hope to do.»*

2001: A Space Odyssey.
Stanley Kubrick, 1968.

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

Sommario

CIRSFID - Centro Interdipartimentale di Ricerca in Storia del Diritto, Filosofia e Sociologia del Diritto e Informatica Giuridica dell'Università di Bologna

Dottorato di Ricerca in Diritto e Nuove Tecnologie

Responsabilita Penale e Automazione nell'E-Health

Francesca LAGIOIA

Questo lavoro di ricerca indaga i problemi relativi alla responsabilità penale legata all'uso di sistemi di automazione e d'intelligenza artificiale nel settore dell'e-health. Tale indagine è stata svolta inquadrando il sistema sanitario all'interno di una visione socio-tecnica, con particolare attenzione all'interazione tra uomo e macchina, al livello di automazione dei sistemi e al concetto di errore e gestione del rischio. Sono state approfondite alcune specifiche aree di interesse quali: la responsabilità penale per danno da dispositivi medici difettosi; la responsabilità medica, connessa all'uso di sistemi a elevata automazione e legata a difetti del sistema; e, in particolare, la responsabilità penale legata all'uso di sistemi d'intelligenza artificiale e i modelli elaborati dalla dottrina per regolare tale fenomeno. Sono stati esaminati: il modello zoologico, il modello dell'agente mediato, il modello della conseguenza naturale e probabile e il modello della responsabilità diretta. È stata esaminata la possibilità che un agente autonomo intelligente sia in grado di soddisfare i requisiti dell'*actus reus* e della *mens rea*, quali condizioni necessarie all'attribuzione di responsabilità penale, qualora un AI ponga in essere una condotta astrattamente riconducibile a una fattispecie criminosa. I profili di responsabilità sono analizzati sulla base di casi e scenari. Infine sono state evidenziate possibili soluzioni e rimedi, anche alla luce della teoria degli agenti normativi. ...

Ringraziamenti

Desidero ringraziare tutti coloro che, in modi diversi, hanno contributo al mio lavoro in questi anni, con suggerimenti, osservazioni, critiche e pazienza: a loro va la mia gratitudine, anche se a me spetta la responsabilità per qualsiasi errore sia presente in questa tesi.

Ringrazio prima di ogni altro il Professor Sartor, per la sua grande disponibilità, per le conversazioni di ispirazione scientifica e umana e perché per me è stato ed è un maestro prezioso.

Un ringraziamento particolare va al dott. Contissa, a cui devo molto di ciò che ho imparato in questi anni. Le nostre conversazioni sono state importanti spunti di riflessione, che mi hanno permesso di guardare le cose da punti di vista a me prima poco conosciuti.

Ringrazio i colleghi, gli amici e tutti coloro che mi hanno incoraggiato o speso parte del proprio tempo ad ascoltare e discutere le mie idee.

Vorrei infine ringraziare le persone a me più care, i miei amici, la mia famiglia, e in modo particolare Ernesto. La scelta di occuparmi di temi legati ai sistemi d'intelligenza artificiale nasce da una conversazione avuta con lui alcuni anni fa. Ringrazio lui anche per il prezioso aiuto nella mia personale battaglia con LaTeX. ...

Indice

Elenco delle figure

Elenco delle tabelle

Introduzione

1.1 Introduzione

Negli ultimi anni abbiamo assistito alla diffusione di robot e sistemi d'intelligenza artificiale nei settori più disparati, come la difesa militare, il settore aerospaziale, i trasporti aerei e terrestri, la sanità, i sistemi bancari e ancora i processi industriali e i mercati finanziari.

Il problema del gap tra il diritto e lo sviluppo tecnologico è da tempo all'attenzione di numerosi giuristi[1]. Uno dei temi di maggiore attualità per gli ordinamenti giuridici contemporanei riguarda, infatti, la tutela degli interessi fondamentali della vita e della salute contro le occasioni di rischio da cosiddetto ignoto tecnologico. Tale espressione identifica un contesto di incertezza scientifica, in cui "le modalità del calcolo del rischio, come sono state sinora definite dalla scienza e dalle istituzioni legali, collassano."[2]. L'ampia diffusione di sistemi a elevata automazione e d'intelligenza artificiale avrà presto un notevole impatto sul diritto tradizionale e, in particolare, sul diritto penale, con il rischio che le categorie classiche, quali causalità, condotta, offesa, autore e colpevolezza, subiscano gravi torsioni, una volta entrate in contatto con tale fenomeno caratterizzato dal connotato del rischio all'interno di un contesto di incertezza epistemologicaPerrow, C. *Normal Accidents: Living with High Risk Technologies (Updated)*. Princeton University Press, 1999. Verso la diffusione di una cultura della prevenzione si è orientato l'ordinamento comunitario che, proprio in relazione ai rischi cd. da ignoto tecnologico, sottolinea l'importanza dell'applicazione del "principio della precauzione e dell'azione preventiva" previsto dal

[1] Stella, F. «Giustizia e modernità». In: *La protezione dell'innocente e la tutela delle vittime* (2003) 292 e ss., il quale parla di "shock da modernità". Centonze, F. *La normalità dei disastri tecnologici: il problema del congedo dal diritto penale*. Giuffrè, 20043 e passim; Paliero, C. «L'autunno del patriarca. Rinnovamento o trasmutazione del diritto penale dei codici». In: *Riv. it. dir. proc. pen.* Vol. 4. 1994, p. 1220. Con particolare riferimento all'uso di sistemi d'intelligenza artificiale si vedano anche Borruso, R. *Computer e diritto*. Computer e diritto v. 1. A. Giuffrè, 1988; Casa, F. «Dalle scienze cognitive alle applicazioni giuridiche dell'intelligenza artificiale / Federico Casa.» In: (), 67-131; Sartor, G. *Intelligenza artificiale e diritto: un'introduzione*. A. Giuffrè, 1996, passim.

[2] Beck, U. «La società del rischio, trad. it». In: *Roma, Carocci* (2000), 29.

secondo comma dell'articolo 174 TUE, oltre che dall'articolo III-223 del progetto di Trattato istitutivo della Costituzione europea. Tale principio di derivazione comunitaria assume a fondamento l'incertezza scientifica e la mancanza di dati tecnico-empirici, quindi l'impossibilità di fornire rassicurazioni sugli effetti per la salute e l'incolumità pubblica derivanti dall'applicazione di una determinata tecnologia o da un'attività economico-produttiva che utilizzi fattori tecnologici avanzati.

Per ragioni di opportunità e per interesse di chi scrive, si è scelto di indagare i problemi relativi alla responsabilità penale legata all'uso di sistemi di automazione e d'intelligenza artificiale nel settore dell'e-health. La sanità, infatti, costituisce una delle aree principali di sviluppo di robot e sistemi d'intelligenza artificiale, così come evidenziato dalla Research Agenda for Robotics in Europe 2014-2020. Basti pensare che mentre solo ieri, secondo Hans Jonas, l'arte medica tradizionalmente intesa doveva assecondare e non già forzare la natura, nella medicina moderna si assiste oggi alla nascita e allo sviluppo di ambizioni talvolta definite al limite del bizzarro, di illusioni di onnipotenza, e di sogni di immortalità. Tali caratteristiche, accanto all'emergere di problematiche per nulla risolte, contribuiscono ad attrarre l'attenzione sul tema oggetto del presente lavoro.

Prima di affrontare il tema della responsabilità penale, è sembrato opportuno prendere le mosse dall'analisi dei sistemi in uso o di prossima introduzione nel settore sanitario e le aree coinvolte, così da delineare i confini della ricerca.

Si procederà poi a inquadrare ed esaminare il sistema sanitario all'interno della teoria dei sistemi socio-tecnici . Esse hanno affrontato i problemi emergenti dall'interazione tra l'uomo e le tecnologie, a livello fisico, cognitivo e organizzativo, considerando due macro componenti all'interno di un sistema organizzativo complesso, quali il sottosistema tecnico e il sottosistema sociale. Tale approccio permette di far emergere le difficoltà che nascono dall'interazione tra essere umano e tecnologie. In tal modo, sarà più facile comprendere in che modo le capacità cognitive di tali sistemi possano coadiuvare o sostituire l'attività umana, quali rischi ne derivino e che tipo di implicazioni comportino.

Saranno prese in considerazione tassonomie e sistemi di classificazione dei livelli di automazione, poiché essi costituiscono validi strumenti per l'analisi giuridica della responsabilità. Come postulato da Sheridan e Verplanck, l'automazione non è "tutto o niente", non si tratta di automatizzare del tutto o solo parzialmente una certa attività o un dato compito, ma di stabilirne il livello di automazione. La scelta di avvalersi di una tassonomia dei livelli di automazione come strumento per l'analisi giuridica della responsabilità, risiede nella possibilità di identificare il livello di automazione di una tecnologia, con riferimento a

specifiche funzioni cognitive; determinare l'esatta divisione dei compiti tra uomo e macchina; e, infine, investigare il tema della responsabilità associandola al livello di automazione e alla divisione dei compiti.

All'interno di una visione socio-tecnica, saranno esaminati anche i modelli di analisi e gestione del rischio clinico, quali strumenti per migliorare la qualità e la sicurezza delle prestazioni sanitarie e prevenire errori ed eventuali danni alla salute dei pazienti.

Le caratteristiche del sistema sanità possono avere un notevole impatto sulla ripartizione delle responsabilità, è molto importante avere presente che i sistemi ad elevata automazione, e in particolare i sistemi d'intelligenza artificiale, possono essere considerati come strumenti nelle mani degli operatori umani e, al tempo stesso, come agenti autonomi, che operano all'interno del sistema, contribuendo attivamente allo svolgimento di compiti e talvolta sostituendosi, in tutto o in parte, all'operatore umano.

Considerata la molteplicità dei possibili soggetti coinvolti, abbiamo scelto di indagare e approfondire alcune specifiche aree e in particolare (a) la responsabilità penale da prodotto difettoso, con riferimento alla responsabilità per danno da dispositivi medici difettosi; (b) la responsabilità medica, connessa all'uso di tecnologie ad alto rischio, nelle ipotesi di danni derivanti dall'uso di dispositivi difettosi; e in particolare (c) la responsabilità penale legata all'uso di sistemi d'intelligenza artificiale, caratterizzati da un elevato grado di autonomia e i modelli elaborati dalla dottrina per regolare tale fenomeno.

La letteratura ha spesso cercato di inquadrare il rapporto tra agenti autonomi intelligenti, sviluppatori e utenti, specialmente in ambito civilistico. Sul fronte del diritto penale, la responsabilità dei sistemi d'intelligenza artificiale è un territorio ancora molto inesplorato. Esistono tuttavia alcune eccezioni. Un primo modello elaborato dalla dottrina propone un'analogia tra agenti intelligenti (AI), dotati di capacità cognitive e stati mentali, e animali. Più di recente Gabriel Hallevy ha proposto tre modelli di responsabilità penale degli AI: (1) Perpetration-via-Another, (2) Natural-Probable- Consequence, e (3) Direct Liability.

Si tenterà di verificare se un agente autonomo intelligente sia concretamente in grado di soddisfare i requisiti dell'actus reus e della mens rea, quali condizioni necessarie all'attribuzione di responsabilità. Qualora un AI sia capace di sviluppare forme di intenzione e volontà, secondo le caratteristiche prescritte dal diritto penale, si procederà nella ricerca di quali soluzioni siano disponibili per ridurre le ipotesi che un agente autonomo intelligente ponga in essere condotte criminose.

Sulla base dei modelli elaborati dalla dottrina, sarà esaminato un caso concreto che, seppur non attinente alla sfera dell'e-health, appare a chi scrive esemplificativo del fenomeno. Non sempre, infatti, è possibile attribuire la responsabilità per il reato compiuto da un AI a programmatori o utilizzatori, con il rischio che si verifichi un vuoto di responsabilità davanti a tali fenomeni e che questi diventino incontrollabili. A tal fine saranno presi in considerazione alcuni dei modelli informatico- giuridici, utili a regolare i casi in cui un AI violi una norma, ponendo in essere una condotta criminosa che, se compiuta da un essere umano, sarebbe punibile secondo le leggi del diritto penale.

Infine, torneremo all'area dell'e-health, provando a costruire uno scenario che prenda in considerazione un sistema d'intelligenza artificiale utilizzato come supporto alla diagnostica, e ad indagare i rapporti tra uomo e macchina, eventuali aspetti problematici e a delineare infine profili di responsabilità, nel caso in cui, da una diagnosi errata del sistema derivino danni al paziente.

Questo lavoro si concentra sul rapporto tra responsabilità penale e sistemi di automazione e d'intelligenza artificiale senza avventurarsi nelle aree che riguardano l'etica, compresa la roboetica, e la moralità. L'obiettivo principale è sviluppare un quadro adeguato per la discussione concreta e funzionale dei problemi che derivano dall'introduzione di sistemi di automazione e d'intelligenza artificiale all'interno di sistemi complessi e in particolare nel sistema sanità.

In particolare, il sistema oggetto dello scenario è Watson, un'applicazione avanzata di elaborazione del linguaggio naturale, information retrieval, rappresentazione della conoscenza, ragionamento automatico e tecnologie di apprendimento automatico nel campo del cd. open domain question answering, lanciata nel 2011 dall'IBM all'interno del programma televisivo americano Jeopardy. Watson è stato costruito sulla base di DeepQA, una tecnologia IBM per la formulazione di ipotesi, raccolta massiva di controprove, analisi e scoring. Oggi è utilizzato, in via sperimentale, come supporto alla diagnostica in campo medico. Nel settore sanitario, infatti, il linguaggio naturale di Watson, la generazione di ipotesi, e le capacità di apprendimento basate sull'evidenza, ne consentono l'utilizzo come sistema di supporto per le decisioni cliniche, da parte del personale medico. Il sistema consente di porre una domanda specificando i sintomi riportati dal paziente e gli altri fattori correlati ed è in grado di elaborare le informazioni per identificarne le parti più importanti, elaborare i dati per trovare fatti rilevanti nella storia clinica ed ereditaria del paziente, esaminare le informazioni disponibili all'interno della letteratura scientifica per formulare e testare ipotesi, e fornire una lista di raccomandazioni individualizzate e classificate per livello di evidenza. Le sorgenti di dati che Watson utilizza per le analisi includono linee guida di trattamenti sanitari, registri medici elettronici, annotazioni del personale medico e sanitario, materiali di ricerca, studi clinici, articoli di riviste

e informazioni sul paziente. Dal 2011 l'IBM collabora con diverse cliniche americane per l'utilizzo di Watson come assistente decisionale del personale medico. Fino a che punto un medico può legittimamente affidarsi a tali tipi di tecnologie? A chi dovrà essere attribuita la responsabilità qualora si verifichi una situazione irregolare, pericolosa e dannosa per il paziente? A chi sarà attribuita la responsabilità penale in caso di morte o lesioni qualora dipendano, ad esempio, da una diagnosi errata o da un malfunzionamento del sistema? Ai produttori? Al personale medico e sanitario? Ed entro quali limiti?

Chi scrive, spera in questo modo di contribuire all'avanzamento dello stato dell'arte in materia di responsabilità penale e uso di sistemi d'intelligenza artificiale.

Chi scrive, spera in questo modo di contribuire all'avanzamento dello stato dell'arte in materia di responsabilità penale e uso di sistemi d'intelligenza artificiale.

1.2 Il carattere interdisciplinare della ricerca e la metodologia

La natura di questo lavoro ha carattere interdisciplinare. Data la varietà delle domande poste e degli oggetti di indagine – tecniche d'intelligenza artificiale e sistemi di automazione e analisi giuridica dei profili di responsabilità – saranno coinvolte discipline e metodologie appartenenti a domini differenti:

1. Teorie dei sistemi socio-tecnici.

   Esse hanno affrontato i problemi sociali emergenti dall'interazione tra l'uomo e le tecnologie, a livello fisico, cognitivo ed organizzativo. Tale approccio ci invita a considerare due macro componenti all'interno di un sistema organizzativo complesso:

   - il sottosistema tecnico, costituito non solo dalla tecnologia in senso stretto, ma anche dalle altre risorse materiali e da tutte le prescrizioni, esplicite o tacite, connesse al funzionamento del sistema stesso, come ad esempio procedure, regole e sanzioni e
   - il sottosistema sociale che riguarda le modalità d'interazioni tra le persone, gli schemi cognitivi utilizzati per dar senso e affrontare i problemi emergenti, gli habit, ovvero i comportamenti consuetudinari e i valori di riferimento adottati: in sintesi, l'insieme delle condizioni sociali di funzionamento del sistema

Uno dei capisaldi dell'approccio socio-tecnico - rifiutando l'idea del "determinismo tecnologico" - è che se si interviene solo su uno dei due sottosistemi, e tipicamente quello tecnico, si corre il rischio di ottenere risultati deludenti. È facile incontrare sistemi tecnici perfettamente disegnati che falliscono sul piano del funzionamento reale della struttura organizzativa che li ospita. Di qui l'esigenza della così detta "ottimizzazione congiunta", che lancia un ponte nella progettazione organizzativa tra discipline tecniche (come ad esempio industrial engeneering e informatica) e discipline umanistiche, per integrare tali tecnologie in uno spazio fisico e virtuale condiviso. In tal modo sarà più facile comprendere in che modo le capacità cognitive di tali sistemi possano coadiuvare o sostituire l'attività umana, quali rischi ne derivino e che tipo di implicazioni comportino.

2. Teorie giuridiche e normative.

   Esse forniscono una comprensione delle strutture di norme e sistemi normativi, dei concetti di diritti e responsabilità. Tale approccio è indispensabile per valutare:

   - come un sistema normativo possa regolare l'utilizzo di sistemi dotati d'intelligenza artificiale capaci di agire autonomamente nello spazio fisico e virtuale, in riferimento ai principi generali dell'ordinamento e al diritto penale in particolare;
   - standard e norme tecniche relative alla costruzione di dispositivi medici software caratterizzati dall'impiego di tecniche d'intelligenza artificiale nelle diverse aree del settore sanitario

3. Intelligenza artificiale.

   Essa fornisce modelli computabili di conoscenza, apprendimento e ragionamento la cui comprensione e classificazione è indispensabile per la creazione di una tassonomia dei sistemi d'intelligenza artificiale e una classificazione dei livelli di automazione, nel settore disciplinare di riferimento.

Solo un approccio interdisciplinare consente di ricostruire una visione quanto più possibile completa ed esaustiva dei problemi legati all'uso di sistemi di automazione e d'intelligenza artificiale nell'e-health e delle possibili soluzioni. Tuttavia, considerando che chi scrive ha una formazione prettamente giuridica, non si ha la pretesa di applicare tali approcci in maniera sempre corretta ed esaustiva.

Healthcare, robotica e sistemi d'intelligenza artificiale

2.1 Premessa

Prima di poter indagare il tema della responsabilità penale legata all'uso di sistemi di automazione e d'intelligenza artificiale nell'e-Health, è necessario verificare quali siano i sistemi in uso o di prossima introduzione in questo settore e le aree coinvolte. Si procederà poi ad inquadrare ed esaminare il sistema sanitario all'interno della teoria dei sistemi socio-tecnici e il ruolo dell'automazione al loro interno, facendo emergere le difficoltà che nascono dall'interazione tra essere umano e tecnologie. Saranno prese in considerazione tassonomie e sistemi di classificazione dei livelli di automazione, poiché essi costituiscono validi strumenti per l'analisi giuridica della responsabilità. All'interno di una visione socio-tecnica, saranno esaminati i modelli di analisi e gestione del rischio clinico, quali strumenti per migliorare la qualità e la sicurezza delle prestazioni sanitarie e prevenire errori ed eventuali danni alla salute dei pazienti.

2.2 Aree e applicazioni della robotica nell'healthcare

A causa dei cambiamenti demografici, in molti paesi, i sistemi sanitari sono costretti a fornire assistenza a una popolazione che invecchia, e saranno sottoposti a uno sforzo sempre maggiore. Inoltre, la domanda di assistenza è in aumento e cresce parallelamente al miglioramento delle procedure e dei risultati relativi a una più ampia gamma di condizioni mediche. Assistiamo a un crescente aumento dei costi mentre la percentuale di operatori sanitari umani sembra inesorabilmente destinata a diminuire con il trascorrere del tempo. L'uso della tecnologia, e in particolare della robotica e dei sistemi d'intelligenza artificiale, è generalmente considerata come parte della soluzione.

Per poter procedere a una classificazione dei sistemi di automazione e d'intelligenza artificiale nell'healtcare, è utile prima di tutto individuare tre macroaree principali e le relative applicazioni:

- *Robotica clinica*: essa è definita come l'insieme di sistemi robotizzati e le tecniche d'intelligenza artificiale che supportano la "cura" e i "processi di cura". Riguarda principalmente la diagnosi, il trattamento, l'intervento chirurgico e farmacologico, ma anche l'assistenza sanitaria di emergenza. Questi sistemi sono generalmente gestiti da personale clinico o altro personale di assistenza qualificato.

  All'interno della robotica clinica possiamo distinguere molteplici applicazioni. Queste possono essere classificate nelle seguenti sotto-aree:

  - Sistemi che migliorano le abilità chirurgiche e l'efficacia degli interventi
  - Sistemi che consentono diagnosi e interventi a distanza
  - Sistemi di assistenza diagnostica e terapeutica
  - Sistemi di assistenza nelle procedure chirurgiche

  A queste si aggiungono una serie di applicazioni cliniche ausiliarie come, ad esempio, il prelievo di campioni, la manipolazione e i test dei tessuti e i servizi clinici correlati.

- *Robotica per la riabilitazione*: quest'area copre l'assistenza a seguito di lesioni post-operatorie o in cui l'interazione fisica diretta con un sistema robotico è in grado di favorire nel paziente il recupero di funzionalità parzialmente compromesse, o di agire in sostituzione delle funzionalità perdute.

  Le applicazioni riguardano principalmente protesi robotiche ed esoscheletri. Tali dispositivi possono essere utilizzati per l'assistenza, la riabilitazione, il potenziamento o la sostituzione nella deambulazione. I modelli più recenti prevedono dei rilevatori di impulsi elettrici che registrano gli impulsi neurali o neuromuscolari del paziente e li trasmettono a dispositivi automatici in grado di muovere la parte meccanica della protesi o dell'esoscheletro.

  Protesi robotiche ed esoscheletri possono essere utilizzati in ambito clinico o domestico, tuttavia gli operatori sanitari dovranno fornire le indicazioni per l'impostazione dei parametri e il monitoraggio dei progressi nella motilità del paziente.

- *Robotica di assistenza*: nell'ambito del processo di cura, la funzione primaria dei sistemi che appartengono a questa area è quella di fornire assistenza, sia ai *caregiver* e agli operatori sanitari, sia ai pazienti, in un contesto ospedaliero o in una struttura di assistenza specialistica.

  La robotica di assistenza è progettata per supportare lo svolgimento di attività di routine. Tali tecnologie trovano per lo più applicazione in strutture di assistenza specialistica e in strutture del servizio sanitario nazionale. Tuttavia, a seconda che si tratti dell'una o dell'altra, esistono differenze significative nella progettazione e nell'implementazione di tali sistemi. In un contesto di assistenza sanitaria specialistica, come un ospedale o una casa di cura per anziani, i robot saranno gestiti da uno staff professionale e dovranno conformarsi alle norme e agli standard di certificazione clinica e sanitaria. Tali sistemi fungono da ausilio per i dipendenti, e in particolare per gli operatori sanitari, nello svolgimento quotidiano delle proprie mansioni. Questo tipo di tecnologie consente ai *caregiver* di ridurre il lavoro fisico, come ad esempio il sollevamento dei pazienti, e dà loro assistenza nelle mansioni di routine.

Queste tre macro-aree sono caratterizzate dalla necessità di sistemi che garantiscano sicurezza e tengano conto delle esigenze cliniche dei pazienti. Tipicamente sono gestiti o istituiti da personale clinicamente qualificato.

L'uso della robotica e dei sistemi d'intelligenza artificiale in area medica rappresenta una sfida importante di ricerca per la sua natura multidisciplinare e comprende una vasta gamma di applicazioni. Queste sono illustrate dettagliatamente, nelle sezioni seguenti. A partire dallo studio *Robotics for healthcare: final report*[1], abbiamo cercato di fornire una classificazione delle numerose sottoaree della robotica clinica, di assistenza e riabilitativa, e alcuni esempi delle tecnologie e dei sistemi automatici in uso.

2.3 La robotica clinica

Questa macro-area copre la robotica per la chirurgia, e i sistemi per la diagnosi e i processi terapeutici ed è potenzialmente in grado di trovare applicazione per quasi tutte le patologie, da quelle cardiache, vascolari, ortopediche, a quelle oncologiche e neurologiche.

[1] butter2008robotics

2.3.1 La chirurgia robotica

Una delle principali attività del sistema sanitario è l'esecuzione di interventi medici, tra cui ad esempio interventi di chirurgia e interventi minori come biopsie e prelievi di tessuti. In questo settore la ricerca ha portato a grandi risultati e, non solo le tecnologie sviluppate iniziano a prendere in consegna alcuni dei compiti prima esclusivamente affidati al lavoro del chirurgo, ma consentono anche di compiere attività che il chirurgo da solo non è in grado di eseguire. Si pensi, per esempio, ad alcune caratteristiche sei sistemi robotici, quali l'altissima precisione, la resistenza e la ripetibilità. [2].

Dai più considerata una tecnologia emergente ed eccitante, è nata e si è sviluppata negli ultimi vent'anni [3]. La letteratura più recente è ricca di contributi specialistici sull'impiego della chirurgia robotica, con casistiche importanti e

[2]La precisione e la ripetibilità sono caratteristiche misurabili di *Robot* e sistemi d'intelligenza artificiale. Tali caratteristiche hanno un impatto diretto sull'efficacia del sistema durante l'esecuzione di un *task*. La ripetibilità identifica la capacità del sistema di realizzare ripetizioni dello stesso compito. Più precisamente misura la variabilità dei risultati ottenuti nella procedura assegnata. La precisione o accuratezza identifica la differenza tra il compito richiesto e il risultato ottenuto. A differenza della ripetibilità, per la precisione è necessario definire un sistema di riferimento assoluto a cui riferire l'errore. Inoltre, la precisione aumenta al diminuire della variabilità e quindi della ripetibilità.

[3]Per un'approfondimento sulla storia della robotica in chirurgia si leggano Kalan, S. et al. «History of robotic surgery». In: *Journal of Robotic Surgery* 4.3 (2010), pp. 141–147; Davies, B. «A review of robotics in surgery». In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 214.1 (2000), pp. 129–140. La storia della robotica in chirurgia ha inizio nel 1985 negli Stati Uniti con una macchina denominata PUMA 560 (Programmable Universal Manipulation Arm), della Westinghouse United e utilizzata per biopsie cerebrali di alta precisione e tre anni più tardi per la resezione transuretrale della prostata. Pochi anni dopo, furono sviluppate altre macchine con gli stessi scopi: il PROBOT, sviluppato dall'Imperial College di Londra e usato nell'urologia del Guy's and St Thoma's Hospital di Londra, e il ROBODOC della Integrated Surgical Systems, usato in ortopedia per la chirurgia del femore e del ginocchio, ed entrambi approvati dalla FDA negli Stati Uniti . All'inizio degli anni novanta l'esercito americano stabilì una collaborazione con lo Stanford Research Institute per portare attraverso la telechirurgia il soccorso chirurgico al soldato ferito in un luogo distante, con la speranza di diminuire la mortalità dei soldati in guerra, mediante ospedali chirurgici mobili collegati a distanza con un chirurgo operatore. Tuttavia, il progetto si limitò alla parte sperimentale sugli animali. Queste ricerche furono riprese in ambito civile con la realizzazione dell'AESOP, un braccio robotico capace di rispondere ai comandi vocali del chirurgo, per l'orientamento di una telecamera. Poco dopo nacquero due apparecchiature statunitensi per telechirurgia robotica: il Da Vinci e lo Zeus. Esse hanno di fatto creato la cosiddetta chirurgia robotica. Tra gli altri tentativi di robotica vanno ricordati quelli italiani e dei quali uno della Scuola Superiore Sant'Anna che ha proposto un prototipo miniaturizzato per colonscopia computerizzata e l'altro il TELELAP non ancora in commercio della SOFAR di Lodi che sta per produrre un sistema competitivo rispetto al DA VINCI.

studi comparativi rispetto, per esempio, alla chirurgia laparoscopica tradizionale, oltre a revisioni complete sull'argomento ed editoriali di approfondimento[4]. Questi sistemi sono anche in grado di operare in uno spazio molto più contenuto all'interno del corpo umano e per tale motivo sono, infatti, considerati un fattore molto importante per il futuro della chirurgia. I sistemi di chirurgia robotica permettono interventi chirurgici a cielo aperto, interventi di chirurgia mini-invasiva, tele-chirurgia e ancora la navigazione intra-operatoria e la simulazione chirurgica.

Possiamo distinguere cinque sotto-aree principali:

(I) Sistemi di assistenza per la microchirurgia

La microchirurgia è una branca della chirurgia che prevede l'applicazione di tecniche ad altissima precisione su superfici corporee molto ristrette e l'uso di speciali microscopi operatori. Oggi si applica oggi a numerose discipline mediche, tra cui la neurologia, la cardiologia, l'oftalmologia, l'ortopedia, la ginecologia, l'otorinolaringoiatria, la medicina maxillo-facciale, la pediatria, la chirurgia plastica ricostruttiva e la medicina estetica. Le attuali tecnologie non consentono di sostituire il chirurgo nell'esecuzione dell'intervento, in particolare per gli interventi sui tessuti molli, poiché i chirurghi riescono ad adattarsi più facilmente al comportamento del tessuto, ma permettono di assistere l'esecuzione dell'intervento per garantire un livello di precisione molto più elevato. Questi sistemi, infatti, migliorano la scalabilità e filtrano i naturali tremori del movimento. Sono dotati di interfacce utenti, sistemi di feedback tattili.

Di seguito si riportano alcuni esempi.

- ACRobot: sviluppato dalla Acrobot Company Ltd e utilizzato soprattutto nella microchirurgia ortopedica [5].

[4]Si vedano in proposito Giulianotti, P. C. et al. «Robotics in general surgery: personal experience in a large community hospital». In: *Archives of surgery* 138.7 (2003), pp. 777–784; Herron, D., Marohn, M. et al. «A consensus document on robotic surgery». In: *Surgical endoscopy* 22.2 (2008), pp. 313–325; Tang, V. C. e Stacey, N. C. «Robotic surgery». In: *Annals of The Royal College of Surgeons of England* 89.4 (2007), p. 447.

[5]Si vedano ad esempio Jakopec, M. et al. «The hands-on orthopaedic robot" Acrobot": Early clinical trials of total knee replacement surgery». In: *Robotics and Automation, IEEE Transactions on* 19.5 (2003), pp. 902–911; Cobb, J. et al. «Hands-on robotic unicompartmental knee replacement A PROSPECTIVE, RANDOMISED CONTROLLED STUDY OF THE ACROBOT SYSTEM». in: *Journal of Bone and Joint Surgery, British Volume* 88.2 (2006), pp. 188–197. Con particolare riferimento alla microchirurgia del ginocchio e a una rassegna su alcune delle tecnologie utilizzate, si veda Delp, S. L. et al. «Computer assisted knee replacement.» In: *Clinical orthopaedics and related research* 354 (1998), pp. 49–56.

- Steady Hand Robot: sviluppato dalla Johns Hopkins University [6].
- Da Vinci Surgical System: sviluppato dalla Intuitive Surgical Inc., è uno dei sistemi oggi più utilizzati in numerose branche della chirurgica [7].

(II) Sistemi per la chirurgia di precisione

Si tratta di procedure chirurgiche cd. *hard material*, come per esempio interventi ossei, fortemente predisposte all'assistenza da parte di sistemi robotici, anche grazie alla natura statica e alle capacità predittive.

Essi permettono di definire, pianificare e costruire modelli 3D del paziente molto precisi. Questi vengono poi perfettamente allineati al *robot* e al paziente in modo che quest'ultimo sia in grado di manipolare e posizionare strumenti come trapani, bisturi, seghe o aghi.

Di seguito si riportano alcuni esempi.

- ROBODOC: prodotto dalla Integrated Surgical Systems e utilizzato soprattutto in ortopedia per la chirurgia del femore e del ginocchio[8].
- Paky Needle Driver: sviluppato dalla Johns Hopkins University [9].

(III) Sistemi per la chirurgia mini-invasiva (MIS)

La chirurgia robotica mini-invasiva presenta numerosi vantaggi rispetto alla chirurgia mini-invasiva teleassistita classica. Il chirurgo deve operare

[6] Per approfondimenti Taylor, R. et al. «A steady-hand robotic system for microsurgical augmentation». In: *The International Journal of Robotics Research* 18.12 (1999), pp. 1201–1210Üneri, A. et al. «New steady-hand eye robot with micro-force sensing for vitreoretinal surgery». In: *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE. 2010, pp. 814–819; Abbott, J. J., Hager, G. D. e Okamura, A. M. «Steady-hand teleoperation with virtual fixtures». In: *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*. IEEE. 2003, pp. 145–151.

[7] Tra i molti, si veda per esempio Ballantyne, G. H. e Moll, F. «The da Vinci telerobotic surgical system: the virtual operative field and telepresence surgery». In: *Surgical Clinics of North America* 83.6 (2003), pp. 1293–1304; Maeso, S. et al. «Efficacy of the Da Vinci surgical system in abdominal surgery compared with that of laparoscopy: a systematic review and meta-analysis». In: *Annals of surgery* 252.2 (2010), pp. 254–262.

[8] Si veda per esempio Sugano, N. «Computer-assisted orthopedic surgery». In: *Journal of Orthopaedic Science* 8.3 (2003), pp. 442–448 e Lang, J. et al. «Robotic systems in orthopaedic surgery». In: *Journal of Bone & Joint Surgery, British Volume* 93.10 (2011), pp. 1296–1299.

[9] Per approfondimenti, Stoianovici, D. et al. «A modular surgical robotic system for image guided percutaneous procedures». In: *Medical Image Computing and Computer-Assisted Intervention—MICCAI'98*. Springer, 1998, pp. 404–410; Stoianovici, D. «URobotics—urology robotics at Johns Hopkins». In: *Computer Aided Surgery* 6.6 (2001), pp. 360–369.

attraverso incisioni molto piccole che non gli permettono di inserire le mani all'interno dell'incisione e neppure di vedere direttamente l'area su cui sta operando, deve quindi avvalersi di un endoscopio e di un display. In molti casi la chirurgia robotica mini-invasiva rende possibile una chirurgia altrimenti difficile o impossibile. Inoltre, elimina il tremore e aumenta i cosiddetti gradi di libertà degli strumenti operatori e delle loro estremità articolabili. Questi sistemi sono generalmente composti da un interfaccia-utente per la gestione degli strumenti operatori che non è posizionata sul tavolo operatorio. Il chirurgo può sedere in una posizione comododa e, attraverso l'interfaccia, controllare tutti gli strumenti avvalendosi di manigle ergonomiche, e visualizzare un'immagine 3D orientabile.

Di seguito si riportano alcuni esempi.

- Da Vinci Surgical System: sviluppato dalla Intuitive Surgical Inc., è uno dei sistemi più diffusi [10]. (Commercialmente disponibile).
- Zeus: sviluppato dalla Computer Motion [11].(Commercialmente disponibile).

Questo gruppo di sistemi si sovrappone spesso ai sistemi di microchirurgia, come ad esempio il sistema Da Vinci.

(IV) Sistemi Nanobot e Microbot

Le nanotecnologie applicate alla chirurgia permettono di intrudurre nel corpo del paziente sistemi di monitoraggio, diagnosi o intervento. Esse includono nanoparticelle a scopi diagnostici e di screening, sistemi di somministrazione di farmaci, applicazioni di terapia genica e i cd. nanorobot. I nanobot sono generalmente coordinati da un computer di bordo. La maggior parte di questi sistemi è in fase di ricerca e sviluppo, alcuni sono già stati testati. Un esempio è un sensore dotato di un interruttore di dimensione pari a 1,5 nanometri, in grado di contare molecole specifiche all'interno di un campione chimico. Alcune applicazioni molto interessanti potrebbero riguardare l'identificazione e la distruzione di cellule tumorali. I sistemi Nanobot [12] sono ancora molto poco sviluppati, mentre alcuni

[10] Per approfondimenti si veda Ballantyne e Moll, «The da Vinci telerobotic surgical system: the virtual operative field and telepresence surgery», cit.; Maeso et al., «Efficacy of the Da Vinci surgical system in abdominal surgery compared with that of laparoscopy: a systematic review and meta-analysis», cit.; Palep, J. H. et al. «Robotic assisted minimally invasive surgery». In: *Journal of Minimal Access Surgery* 5.1 (2009), p. 1.

[11] Per un confronto tra i due sistemi citati nella chirurgia laparoscopica, branca in cui il loro utilizzo è largamente diffuso, si legga Sung, G. T. e Gill, I. S. «Robotic laparoscopic surgery: a comparison of the da Vinci and Zeus systems». In: *Urology* 58.6 (2001), pp. 893–898

[12] Le cui dimensioni sono inferiori a 100 nm, 1 nm = 10-9 m.

sistemi Microbot sono già stati sviluppati e testati. Il potenziale di questi sistemi è enorme, potendo fornire alternative alla chirurgia o contribuire alla chirurgia mini-invasiva [13].

Di seguito si riportano alcuni esempi.

- Miniature in vivo robots: sviluppato dalla University of Nebraska Medical [14]. (In fase di sviluppo).
- Micro medical robot: sviluppato dalla Ritsumeikan University. (Prototipo).
- Swimming Micro Robot: sviluppato dalla Technion. (Prototipo).
- Endoscopic Micro-Capsules: sviluppate dalla Carnegie Mellon University. (In fase di sviluppo).

(V) Sistemi di Telechirurgia

Attraverso l'ausilio di mezzi telematici, la telechirurgia consente di eseguire interventi chirurgici a distanza: l'operatore, con l'ausilio di un monitor che gli consente l'osservazione continua del campo operatorio, esegue le varie manovre dell'intervento che, teletrasmesse, vengono ripetute sul paziente da un *robot*, con estrema precisione. Uno dei primi interventi chirurgici a distanza, caso noto come *Lindberg Operation*, è stato eseguito nel 2001 mentre il chirurgo era a New York e il paziente a Strasburgo [15]. L'avvento della realtà virtuale ha permesso il collegamento a distanza di sistemi di realtà virtuale portando la chirurgia robotica verso la telechirurgia virtualmente assistita. Il chirurgo e il *robot* sono entrambi dotati di sensori, come per esempio telecamere, microfoni e sensori pressori, ed effettori che permettono di riprodurre fedelmente e in tempo reale stimoli sensoriali e azioni. Il chirugo riceve a sua volta informazioni dal campo operatorio, come immagini stereoscopiche tridimensionali e stimoli pressocettivi, che gli permettono di operare avendo l'illusione di essere in sala operatoria. I sistemi di Telechirurgia sono simili ai sistemi utilizzati nella micro-chirurgia robotica e mini-invasiva.

[13] Per approfondimenti sul tema si veda Sitti, M. «Micro-and nano-scale robotics». In: *American Control Conference, 2004. Proceedings of the 2004.* Vol. 1. IEEE. 2004, pp. 1–8

[14] Dolghi, O. et al. «Miniature in vivo robot for laparoendoscopic single-site surgery». In: *Surgical endoscopy* 25.10 (2011), pp. 3453–3458

[15] Marescaux, J. et al. «Transatlantic robot-assisted telesurgery». In: *Nature* 413.6854 (2001), pp. 379–380; Marescaux, J. «Nom de code:«Opération Lindbergh»». In: *Annales de chirurgie*. Vol. 127. 1. Elsevier Masson. 2002, pp. 2–4

2.3.2 Robotica e sistemi d'intelligenza artificiale per la diagnostica

Il tema della prevenzione e della diagnosi sono al centro dei programmi di assistenza e prevenzione sanitaria di moltissimi paesi europei. Numerosi sono i programmi di screening per monitorare, per esempio, la salute dei giovani o il cancro al seno, così come lo stutio statistico delle correlazioni tra i problemi di salute e il background sanitario, lo stile di vita, le condizioni ambientali e di lavoro o precedenti problemi di salute. Per il riconoscimento precoce dei gruppi a rischio è necessario un processo che coinvolga il monitoraggio, lo screening e alla fine la diagnosi.

La robotica e sistemi d'intelligenza artificiale per uso diagnostico iniziano a svilupparsi intorno agli anni settanta. Si tratta per lo più di *robot* fisici o sistemi esperti di diagnosi automatica [16], in grado di acquisire dati attraverso il proprio sottosistema di conoscenza, sistemi di telediagnosi e strumenti, come i *robot* miniaturizzati, per la diagnosi interna, consentendo attività diagnostiche altrimenti non possibili.

Possiamo distinguere tre sotto-aree principali.

(I) Telediagnostica e monitoraggio

La telediagnostica rende possibile la diagnosi e la valutazione dello stato di salute dei pazienti a distanza. Si tratta di tecnologie che possono includere funzionalità di comunicazione, sistemi sensoriali, sistemi di gestione della conoscenza e i cd. *smart Human Machine Interface tools* (HMI). Rientrano in quest'area i sistemi di sensori automatici e di imaging che possonono essere teleoperati, sistemi per la consultazione a distanza e sistemi capaci di monitorare costantemente le condizioni del paziente, segnalando e trasmettendo informazioni in tempo reale.

Di seguito alcuni esempi.

- OTELO: sviluppato e prodotto dalla Laboratoire Vision & Robotique, IUT de Bourges. È un sistema di tele-ecografia mobile che utilizza un *robot* ultraleggero per la diagnosi in remoto dei pazienti [17]. (Progetto di sviluppo terminato)

[16] Salcudean, S. E. et al. «Medical Image Computing and Computer-Assisted Intervention – MICCAI'99: Second International Conference, Cambridge, UK, September 19-22, 1999. Proceedings». In: a cura di Taylor, C. e Colchester, A. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999. Cap. Robot-Assisted Diagnostic Ultrasound – Design and Feasibility Experiments, pp. 1062–1071Dario, P. e Bergamasco, M. «An advanced robot system for automated diagnostic tasks through palpation». In: *IEEE Transactions on Biomedical Engineering* 35.2 (feb. 1988), pp. 118–126

[17] Si veda Vieyres, P. et al. «A tele-operated robotic system for mobile tele-echography: The OTELO project». In: *M-Health*. Springer, 2006, pp. 461–473.

- Robot-Based Tele-Echography II: sviluppato dalla University Hospital of Grenoble, è un sisteme di tele-ecografia mobile [18]. (In fase di sperimentazione clinica)
- IWARD: è un sistema di monitoraggio sviluppato dal Fraunhofer (FhG) in collaborazione con altri partner europei [19].

(II) Sistemi intracorporei e Smart medical capsules

Generalmente, le diagnosi intracorporee avvengono mediante procedure considerate gravose per il paziente. L'uso della robotica combinata alla miniaturizzazione dei sensori ha portato allo sviluppo di capsule endoscopiche, miniaturizzate e wireless. Le cosiddette *smart medical capsules*, sviluppate nei primi anni duemila e approvate dalla FDA per il controllo delle malattie dell'apparato digerente, sono i primi dispositivi miniaturizzati e non invasivi[20]. Le attuali capsule endoscopiche sono dotate di telecamenre che permettono di visualizzare il tratto gastrointestinale. Per poter avere un controllo sulla posizione e l'orientamento della capsula, la ricerca si sta muovendo verso lo sviluppo di strumenti che permettano di telegestire le capsule ed eseguire biopsie dei tessuti.

Di seguito alcuni esempi:

- Microcapsule dotate di adesivi: sviluppate da un gruppo di ricerca dell'università di Pittsburgh, sono dotate di adesivi cosiddetti bio-inspired che permettono alla capsula di attaccarsi alle pareti intestinali, per la diagnosi dei disturbi gastrointestinali[21].
- PillCam COLON: sviluppate dalla Given Imaging e disponibili in commercio, sono un esempio di capsule intelligenti, utilizzate per

[18] Per approfondienti, Martinelli, T. et al. «Robot-based tele-echography clinical evaluation of the TER system in abdominal aortic exploration». In: *Journal of Ultrasound in Medicine* 26.11 (2007), pp. 1611–1616.

[19] Si veda Szecsi, T. et al. «Hospital robot module development in the iward project». In: *6th CIRP International Conference on Intelligent Computation in Manufacturing Engineering. Naples, Italy*. 2008.

[20] Cheung, E. et al. «A new endoscopic microcapsule robot using beetle inspired microfibrillar adhesives». In: *Advanced Intelligent Mechatronics. Proceedings, 2005 IEEE/ASME International Conference on*. IEEE. 2005, pp. 551–557

[21] Karagozler, M. E. et al. «Miniature endoscopic capsule robot using biomimetic micropatterned adhesives». In: *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*. IEEE. 2006, pp. 105–111.

colonscopie e per la diagnosi del cancro del colon-retto[22]

(III) Sistemi esperti e sistemi d'intelligenza artificiale per la diagnosi

Negli ultimi anni i sistemi di organizzazione e uso delle informazioni hanno avuto un largo sviluppo in medicina. Essa è ritenuta un campo di applicazione ideale dell' intelligenza artificiale, per le specifiche caratteristiche che la conoscenza medica assume in rapporto alla decisione clinica. La ricerca degli ultimi anni ha portato allo sviluppo di sistemi di supporto alla decisione di vario tipo e complessità. I più sofisticati sono i cosiddetti sistemi esperti, attraverso cui la conoscenza medica, opportunamente organizzata, è usata secondo le strategie definite dall'esperto per raggiungere uno specifico obiettivo. Una caratteristica proprietà dei sistemi esperti è la capacità di fornire adeguate spiegazioni sui procedimenti seguiti e sulle conclusioni raggiunte. Sono quindi un valido strumento per assistere la decisione clinica [23]. Poiché basati su una conoscenza altamente strutturata, che comprende riferimenti ai principali aspetti dell'azione medica, come per esempio selezione, associazione e organizzazione dei dati; identificazione e gerarchizzazione dei problemi; attivazione e verifica delle ipotesi; e giustificazione e spiegazione dei risultati, essi danno la possibilità di esplorare in modo intelligente la base di conoscenza da cui dipendono, di rendere trasparenti le strategie del ragionamento clinico su cui si basano e di giustificare le decisioni elaborate [24]. Le caratteristiche principali di questi sistemi sono (a) l'esplicito riferimento a un modello metodologico convalidato; (b) la facilità di archiviazione e recupero dei dati preesistenti; (c) la disponibilità di

[22]Per approfondimenti si vedanoEliakim, R. et al. «Evaluation of the PillCam Colon capsule in the detection of colonic pathology: results of the first multicenter, prospective, comparative study.» In: *Endoscopy* 38.10 (2006), pp. 963–970; Sieg, A., Friedrich, K. e Sieg, U. «Is PillCam COLON Capsule Endoscopy Ready for Colorectal Cancer Screening? A Prospective Feasibility Study in a Community Gastroenterology Practice». In: *The American journal of gastroenterology* 104.4 (2009), pp. 848–854

[23]Per approfondimenti sulle tecniche di intelligenza artificiale utilizzate, per esempio, nel monitoraggio di infezioni si veda Lamma, E. et al. «Artificial intelligence techniques for monitoring dangerous infections». In: *Information Technology in Biomedicine, IEEE Transactions on* 10.1 (2006), pp. 143–155

[24]I sistemi esperti più diffusi sono basati su un modello empirico della conoscenza; in altri casi i dati clinici sono elaborati mediante calcolo probabilistico, utilizzando le conoscenze derivate dall'analisi di grandi basi di dati preesistenti. Altri ancora si basano sull'uso di reti neurali, elaborando la decisione sulla base delle relazioni precedentemente osservate tra i dati clinici e le ipotesi possibili. Esistono anche sistemi di supporto alla decisione meno complessi, basati sul principio di presentare in modo fortemente interattivo la conoscenza necessaria a risolvere specifici problemi (computer-assisted instructions). Essi possono essere notevolmente utili al medico, in quanto capaci di guidare in modo metodologicamente corretto nella ricerca di soluzioni strettamente pertinenti alla situazione in esame: trovano applicazione non solo per facilitare la decisione in ambito professionale, ma anche in campo educativo.

aiuti in linea; (d) la giustificazione delle richieste e delle conclusioni raggiunte; (e) l'aggiornamento frequente delle basi di conoscenza; e infine (f) La verifica e la validazione clinica delle prestazioni.

Di seguito si riportano alcuni esempi.

- MYCIN: sviluppato dalla Standford University, è un sistema esperto per l'identificazione dei batteri causa di gravi infezioni, come la batteriemia e la meningite, e la raccomandazione di terapie antibiotiche e dosaggio in base al peso corporeo del paziente. Il sistema Mycin è stato utilizzato anche per la diagnosi di malattie della coagulazione del sangue. MYCIN non è mai stato utilizzato nella pratica clinica, ma i test hanno indicato che il sistema ha restituito soluzioni terapeutiche corrette, in circa il 69% dei casi, superando le prestazioni degli esperti in malattie infettive della facoltà di medicina di Standford [25].

- DNSEV: (Expert System for clinical result Validation), sviluppato dalla DIANOEMA SpA e dall'ospedale S. Orsola-Malpighi di Bologna, è utilizzato per migliorare la qualita' del processo di validazione eseguito dai laboratori di analisi biochimica [26].

- ESMIS: (Expert System for Microbiological Infection Surveillance), per migliorare la qualita' del processo di validazione eseguito dai laboratori di analisi microbiologica e per monitorare gli eventi infettivi all'interno di un ospedale, è stato sviluppato dalla DIANOEMA SpA e dall'ospedale S. Orsola-Malpighi di Bologna [27].

- DNTAO: (Expert System for supporting the Oral Anticoagulation Treatment) per il supporto ai medici, in campo ematologico, per le prescrizioni e visite per la Terapia Anticoagulante Orale. Anche questo sistema esperto è stato sviluppato dalla DIANOEMA SpA e dall'ospedale S. Orsola-Malpighi di Bologna [28].

[25] Per approfondimenti si vedano Buchanan, B. G., Shortliffe, E. H. et al. *Rule-based expert systems*. Vol. 3. Addison-Wesley Reading, MA, 1984; e Heckerman, D. E. e Shortliffe, E. H. «From certainty factors to belief networks». In: *Artificial Intelligence in Medicine* 4.1 (1992), pp. 35–52.

[26] Storari, S. et al. «Validation of biochemical laboratory results using the DNSev expert system». In: *Expert systems with applications* 25.4 (2003), pp. 503–515.

[27] Si vedano Lamma, E. et al. «An expert system for microbiological data validation and surveillance». In: *Medical Data Analysis*. Springer, 2001, pp. 153–160; Lamma et al., «Artificial intelligence techniques for monitoring dangerous infections», cit.

[28] Barbieri, B. et al. «An expert system for the oral anticoagulation treatment». In: *Innovations in Applied Artificial Intelligence*. Springer, 2005, pp. 773–782

- WATSON: è un sistema d'intelligenza artificiale di elaborazione del linguaggio naturale, information retrieval, rappresentazione della conoscenza, ragionamento automatico e tecnologie di apprendimento automatico nel campo del cd. *open domain question answering*, lanciata nel 2011 dall'IBM che nel 2013 ne ha annunciato la prima applicazione commerciale nella diagnosi e nel trattamento del cancro al polmone, al memorial Sloan-Kettering Cancer Center, in collaborazione con la società di assicurazione sanitaria WellPoint [29].

2.4 La robotica riabilitativa e protesica

La robotica riabilitativa si concentra sul trattamento di pazienti con una disabilità fisica e/o mentale. Essa è finalizzata al recupero funzionale attraverso l'uso di tecnologie assistive. Le patologie generalmente trattate riguardano ictus, sindromi dolorose muscolo-scheletriche, come mal di schiena, fibromialgia e trauma cranico, casi di amputazione e lesioni post-traumatiche e post-operatorie. Anche pazienti con malattie cardiache e polmonari possono beneficiare di trattamenti riabilitativi. La robotica riabilitativa può essere impiegata in ambiente clinico o domestico, in situazioni in cui il *robot* è in grado di supportare il paziente nell'esecuzione indipendente della terapia, soprattutto in caso di attività ripetitive. Le terapie per il trattamento di disabilità fisiche riguardano il recupero della coordinazione motoria e del sostegno muscolare [30]. Le terapie per le

[29] Upbin, B. «IBM's Watson Gets Its First Piece Of Business In Healthcare». In: *Forbes Tech* (2013); Cohn, J. «ntroduction to special issue: Robotic assistance in neuro-motor therapy». In: *Robotica* 21 (2003)

[30] Kazerooni, H. «Springer Handbook of Robotics». In: a cura di Siciliano, B. e Khatib, O. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. Cap. Exoskeletons for Human Performance Augmentation, pp. 773–793 Il tema degli esoscheletri non è una tematica molto recente. Per una revisione sullo stato dell'arte si veda Lo, H. S. e Xie, S. Q. «Exoskeleton robots for upper-limb rehabilitation: state of the art and future prospects». In: *Medical engineering and physics* 34.3 (2012), pp. 261–268. Già negli anni sessanta, furono realizzati i primi prototipi di esoscheletri meccanici, che portarono, in seguito, allo sviluppo degli attuali esoscheletri. Ricercatori statunitensi e della ex Iugoslavia furono tra i primi a sviluppare degli esoscheletri per il potenziamento delle prestazioni fisiche in ambito militare. In entrambi i casi, le difficoltà maggiori si sono riscontrate in relazione alla portabilità e alle interfacce con l'operatore. Verso la fine degli anni sessanta, il reparto della General Electric, con l'ausilio dell' U.S. Office of Naval Research, realizzò il primo prototipo di esoscheletro con attuatori idraulici. Questa macchina pesava circa 680 kg con trentasei gradi di libertà, per lo sviluppo della forza di braccia e gambe. Tale progetto, anche se solo come prototipo, ha ampiamente contribuito a sviluppare questa tecnologia, tanto che fu istituito il progetto EHPA (Exoskeletons for Human Performance Augmentation). Si veda Garcia, E., Sater, J. M. e Main, J. «Exoskeletons for Human Performance Augmentation (EHPA): A Program Summary.» In: *The Robotics Society of Japan* 20.8 (2002), pp. 822–826. Il progetto coinvolse numerose istituzioni e centri di sviluppo. Un esempio è l'esoscheletro Bleex sviluppato dall'Università di

disfunzioni mentali e cognitive sono per lo più basate su giochi di ruolo.

Possiamo distinguere alcune sotto-aree.

(I) *Sistemi di ausilio alla coordinazione motoria*

I danni al cervello o al sistema nervoso possono compromettere la coordinazione motoria. Un esempio molto noto è la riduzione o la perdita del controllo motorio a seguito di ictus. Le terapie sono finalizzate al ripristino delle prestazioni funzionali e al recupero delle funzionalità motorie nel cervello. I meccanismi che regolano il ripristino delle funzioni sulla base della plasticità cerebrale non sono ancora completamente conosciuti. Tuttavia, la concezione generale è che il movimento ripetuto della parte compromessa porti al ripristino del funzionamento cerebrale per il controllo del movimento [31]. Sono state sviluppate tecnologie robotiche per la riabilitazione degli arti superiori e inferiori. Tali sistemi intervengono attraverso il movimento guidato degli arti, per ottimizzare gli effetti terapeutici e funzionali. Generalmente, restituiscono un feedback al paziente in modo da poter regolare la forza, aumentando l'effetto della terapia e stimolando il paziente [32].

Di seguito si riportano alcuni esempi tra i più significativi.

- EC-GENTLE/S: nato da un progetto finanziato dalla Commissione Europea all'interno del quinto programma quadro (5PQ) e dotato

Berkeley. Esso è composto da due gambe motorizzate e da un telaio posteriore a cui è possibile posizionare diversi tipi di carichi. Questa struttura era in grado di riprodurre i movimenti di un essere umano di 75 kg, permettendogli di correre e camminare su diverse tipologie di terreno, grazie all'ausilio di cilindri idraulici. Vedi Zoss, A. B., Kazerooni, H. e Chu, A. «Biomechanical design of the Berkeley lower extremity exoskeleton (BLEEX)». in: *Mechatronics, IEEE/ASME Transactions on* 11.2 (2006), pp. 128–138. Successivamente furono portati avanti altri progetti dal Sarcos Research Corporation e dal MIT, che apportarono migliorie nel campo del carico sostenibile, dell'indossabilità e dei consumi energetici. Un altro progetto, che ha portato notevoli sviluppi nel campo degli esoscheletri, è HAL (Hybrid Assistive Leg) dell'Università di Tsukuba in Giappone, finalizzato al potenziamento e alla riabilitazione. Vedi Sankai, Y. «HAL: Hybrid assistive limb based on cybernics». In: *Robotics Research*. Springer, 2010, pp. 25–34

[31] Si veda Harwin, W. e Hillman, M. «Introduction». In: *Robotica* 21 (01 gen. 2003), pp. 1–1. Per una rassegna sull'uso di sistemi robotici a seguito di lesioni neurologiche si legga ad esempio Marchal-Crespo, L. e Reinkensmeyer, D. J. «Review of control strategies for robotic movement training after neurologic injury». In: *Journal of neuroengineering and rehabilitation* 6.1 (2009), p. 20

[32] Si veda Harwin, W. S., Patton, J. L. e Edgerton, V. R. «Challenges and Opportunities for Robot-Mediated Neurorehabilitation». In: *Proceedings of the IEEE* 94.9 (set. 2006), pp. 1717–1726

di un'interfaccia aptica sviluppata dalla FCS Control System per la riabilitazione degli arti superiori. [33].

- Lokomat: tecnicamente è un esoscheletro robotizzato, controllato elettronicamente con un sistema di allevio del peso e un tapis roulant. I supporti si applicano agli arti inferiori e forniscono un'assistenza diversificata alle gambe. Velocità, frequenza, lunghezza del passo, escursione delle articolazioni di ginocchio e anca, sono fra i parametri del cammino, modificabili nell'arco della riabilitazione. Il paziente è coinvolto in maniera attiva. Grazie alla realtà virtuale visibile su di uno schermo, un avatar procede in una distesa verde "guidato" dal paziente che, imbragato con il Lokomat, cammina sul tapis roulant. Compatibilmente alle condizioni di mobilità, il paziente può anche dirigere il suo avatar a destra o a sinistra muovendo le anche.[34]
- MIME (Mirror Image Movement Enhancer): è un sistema robotizzato per la riabilitazione del gomito e della spalla, sviluppato dal Dipartimento dei Veterans Affairs di Palo Alto negli Stati Uniti, in collaborazione con la Divisione di Medicina Fisica e Riabilitativa e il Dipartimento di Riabilitazione Funzionale dell'Università di Stanford [35].

Questi citati sono solo alcuni tra i più importanti sistemi di ausilio alla coordinazione motoria.

(II) *Sistemi per la terapia assistita dell'allenamento fisico*

[33] Per approfondimenti si leggano Loureiro, R. et al. «Upper Limb Robot Mediated Stroke Therapy—GENTLE/s Approach». In: *Autonomous Robots* 15.1 (2003), pp. 35–51; Coote, S. et al. «The effect of the GENTLE/s robot-mediated therapy system on arm function after stroke». In: *Clinical rehabilitation* 22.5 (2008), pp. 395–405; Amirabdollahian, F. et al. «Multivariate analysis of the Fugl-Meyer outcome measures assessing the effectiveness of GENTLE/S robot-mediated stroke therapy». In: *Journal of NeuroEngineering and Rehabilitation* 4.1 (2007), pp. 1–16

[34] Per approfondimenti si leggano: Mirbagheri, M. M. «Comparison between the therapeutic effects of robotic-assisted locomotor training and an anti-spastic medication on spasticity». In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. ago. 2015, pp. 4675–4678; Alcobendas-Maestro, M. et al. «Lokomat robotic-assisted versus overground training within 3 to 6 months of incomplete spinal cord lesion randomized controlled trial». In: *Neurorehabilitation and neural repair* 26.9 (2012), pp. 1058–1063

[35] Si vedano Lum, P. S. et al. «MIME robotic device for upper-limb neurorehabilitation in subacute stroke subjects: A follow-up study». In: *Journal of rehabilitation research and development* 43.5 (2006), p. 631; Burgar, C. G. et al. «Development of robots for rehabilitation therapy: the Palo Alto VA/Stanford experience». In: *Journal of rehabilitation research and development* 37.6 (2000), pp. 663–674

Tali sistemi sono finalizzati alle terapie di sostegno muscolare, attraverso la ripetizione di alcune attività motorie fondamentali, per cui non è necessaria una supervisione terapeutica continua, in ambiente clinico. Esse possono, quindi, essere svolte in ambiente domestico. I sistemi robotici per uso domestico sono generalmente semplici, poco complessi, non eccessivamente costosi e integrati con PC standard. Spesso si avvalgono di tecnologie provenienti da altri domini (ad esempio dall'industria dei videogiochi).

Si riportano di seguito alcuni esempi.

- Rutgers Ankle: è un sistema di riabilitazione ortopedica della caviglia, sviluppato dall'università di Rutgers [36], che permette al paziente di svolgere una serie di esercizi e controllare il movimento, attraverso l'interazione con un ambiente virtuale.
- BerkelBike: si tratta di una bicicletta reclinata che sfrutta la forza di braccia e gambe. In caso di parziale o tatle perdita dell'uso delle gambe, il sistema di propulsione permette di sostenere il movimento delle gambe attraverso la forza di braccia e spalle. In caso di paralisi, la bicicletta può essere integrata con un sistema di stimolazione elettrica funzionale (FES), che permette di coinvolgere le gambe nel movimento Berkelmans, R., Arns, M. e Duysens, J. «The development of a hybrid outdoor FES bike». In: *Proc. of the 8th Annual Conference of the International Functional Electrical Stimulation Society (IFESS 2003)*. 2003.

(III) *Sistemi per terapie mentali, cognitive e sociali*

L'interazione sociale può essere gravemente ostacolata dal deterioramento cognitivo e mentale. Tali sistemi sono progettati per interagire con gli esseri umani, simulando diversi tipi di comportamento sociale, come ad esempio la comunicazione e il gioco cooperativo. Uno dei vantaggi che derivano dall'uso terapeutico di questi sistemi risiede nel comportamento controllato e nella ripetizione di azioni da parte del *robot*. Le capacità comunicative di soggetti affetti da malattie mentali e/o cognitive è spesso molto problematica. Nella maggior parte dei casi, questi sistemi sono considerati come giocattoli o animali domestici meccanici AMBROSE, R. et al.

[36] Girone, M. et al. «Orthopedic rehabilitation using the" Rutgers ankle" interface». In: *Studies in health technology and informatics* (2000), pp. 89–95

«WTEC Panel Report on International Assessment of Research and Development in Robotics.» In: (2007). Il potenziale terapeutico è ampiamente riconosciuto in letteratura [37] e oggetto di ricerca.

Si riportano di seguito alcuni esempi.

- PARO: si tratta di un *robot* zoomorfo utilizzato come ausilio nella riabilitazione cognitiva [38].
- PlayROB: un sistema sviluppato dalla ARCS per bambini con grave disabilità motoria che permette loro di utilizzare giocattoli tradizionali [39].
- Leonardo: sviluppato dal MIT, è capace di esprimere una vasta gamma di espressioni facciali e corporee, monitorare visivamente il volto di utenti umani, rispondere al contatto fisico ed impegnarsi nell'apprendimento sociale dell'interazione [40].

(IV) *Sistemi di ausilio alla mobilità*

La mobilità è un bisogno sociale fondamentale per molte persone disabili. La sedia a rotelle tradizionale è uno degli strumenti più importanti

[37] Studi di riferimento sono quelli effettuati da Plaisant e collaboratori con i *Pet Robot*, in particolare si veda Plaisant, C. et al. «A storytelling robot for pediatric rehabilitation». In: *Proceedings of the fourth international ACM conference on Assistive technologies*. ACM. 2000, pp. 50–55; le sperimentazioni di Dautenhahn all'interno del progetto AURORA, si veda Dautenhahn, K. «Robots as social actors: Aurora and the case of autism». In: *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco*. Vol. 359. 1999, p. 374; e ancora Sony Aibo, sperimentato in Giappone presso una casa di cura per anziani affetti da demenza, nell'ambito di una terapia occupazionale, si veda Yonemitsu, S., Fujimoto, T. e Tamura, T. «Research for practical use of rehabilitation support equipment for severe dementia». In: *Gerontechnology* 2.1 (2002), p. 91. Sul potenziale terapeutico in bambini affetti da autismo si veda anche Ferrari, E., Robins, B. e Dautenhahn, K. «Therapeutic and educational objectives in robot assisted play for children with autism». In: *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*. IEEE. 2009, pp. 108–114

[38] Sulle potenzialità di PARO nella riabilitazione cognitiva di bambini affetti da sindrome autistica o con problemi nella sfera relazionale, si veda Palma, V. et al. *Paro therapy: potenzialità di un robot zoomorfo come mediatore sociale nel trattamento non farmacologico di bambini con sindrome autistica*

[39] In particolare: Kronreif, G. et al. «Robot assistance in playful environment-user trials and results». In: *Robotics and Automation, 2007 IEEE International Conference on*. IEEE. 2007, pp. 2898–2903; e Kronreif, G. et al. «Playrob-robot-assisted playing for children with severe physical disabilities». In: *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*. IEEE. 2005, pp. 193–196

[40] MIT Personal Robots Group (2014). http://robotic.media.mit.edu/projects/robots/leonardo/overview/overview.html. Si vedano anche CONTI, D. «SOCIALLY ASSISTIVE ROBOTICS UNA POSSIBILE UNIONE TRA ROBOTICA E PSICOLOGIA». in: (); Ingebretsen, M. «Where's My Personal Robot?» In: *Intelligent Systems, IEEE* 24.6 (2009), pp. 90–93

per la mobilità. Tuttavia, un significativo gruppo di persone non è in grado di utilizzare la sedia a rotelle tradizionale per potersi muovere in modo indipendente (alcune cause possone essere ad esempio, cecità, limitata potenza muscolare nelle braccia, deficit della coordinazione occhio-mano).

Lo scopo delle cd. *smart wheelchairs* è quello di garantire autonomia nella mobilità a coloro che non sono in grado di usare sedie a rotelle elettriche tradizionali. Attraverso una serie di sensori e tecnologie per la navigazione e il rilevamento di ostacoli, gli utenti sono in grado di manovrare in modo sicuro ed efficace la sedia a rotelle, mentre il sistema si occupa di controllarla. Simili sono i cd. *smart walkers*, capaci di sostenere individui con limitata capacità di stare in piedi, camminare o rimanere in equilibrio. Tra i sistemi di ausilio alla mobilità troviamo anche esoscheletri, indossabili da persone disabili che permettono il movimento attraverso sensori EMG e protesi intelligenti per i casi di perdita di un arto. Queste protesi imitano le funzionalità motorie attraverso muscoli artificiali, articolazioni o parti dello scheletro. Rispetto alle cd. protesi passive, il controllo della protesi avviene attraverso il sistema nervoso del paziente che beneficia anche del ripristino della sensazione dell'arto, riuscendo a regolarne meglio l'azione.

Di seguito sono riportati alcuni esempi:

- HAL5 Hybrid Assistive Limb: un esoscheletro sviluppato dalla CYBERDYNE e utilizzato da pazienti paraplegici [41].
- IBot: sedia a rotelle intelligente prodotta dalla società americana Indipendence Now [42]
- C-leg knee joint: protesi intelligente prodotta dalla Otto Bock [43]
- McKibben muscle, protesi intelligente sviluppata dalla Shadow Robot Company [44]

[41] Per approfondimenti si leggano Kawamoto, H. et al. «Power assist method for HAL-3 using EMG-based feedback controller». In: *Systems, Man and Cybernetics, 2003. IEEE International Conference on*. Vol. 2. IEEE. 2003, pp. 1648–1653; Suzuki, K. et al. «Intention-based walking support for paraplegia patients with Robot Suit HAL». in: *Advanced Robotics* 21.12 (2007), pp. 1441–1469

[42] Per approfondimenti Arthanat, S., Desmarais, J. M. e Eikelberg, P. «Consumer perspectives on the usability and value of the iBOT® wheelchair: findings from a case series». In: *Disability and Rehabilitation: Assistive Technology* 7.2 (2012), pp. 153–167

[43] Si vedano Kastner, J., Nimmervoll, R. e Wagner, I. «What are the benefits of the Otto Bock C-leg? A comparative gait analysis of C-leg, 3R45 and 3R80». In: *MEDIZINISCH ORTHOPADISCHE TECHNIK* 119 (1999), pp. 131–137; Orendurff, M. S. et al. «Gait efficiency using the C-Leg». In: *Journal of rehabilitation research and development* 43.2 (2006), p. 239

[44] Per approfondimenti Tondu, B. e Lopez, P. «The McKibben muscle and its use in actuating robot-arms showing similarities with human arm behaviour». In: *Industrial Robot: An International Journal* 24.6 (1997), pp. 432–439

2.4.1 La robotica di assistenza

La robotica di assistenza copre i servizi di assistenza professionale [45]. Nonostante la tendenza crescente sia quella di fornire assistenza ai pazienti in ambiente domestico, il più a lungo possibile, la maggior parte di queste attività avviene in istituti di cura come ad esempio ospedali e case di riposo. In questa sezione ci occuperemo principalmente della robotica di assistenza rivolta ai cd. *caregiver* professionali, all'interno di ambienti istituzionali [46]. Le attività svolte da questi soggetti comprendono, per esempio, il trasferimento e il sollevamento dei pazienti, i processi logistici, l'assistenza nello svolgimento di attività quotidiane come l'igiene, l'alimentazione e il monitoraggio dei pazienti.

Possiamo distinguere quattro sotto-aree principali.

(I) *Sistemi di ausilio ad attività di natura logistica*

Un'ampia parte delle attività di assistenza professionale è di natura logistica. Esse comprendono, per esempio, attività di cura personale, fornitura di farmaci e cibo ai pazienti. Lo sviluppo di sistemi intelligenti ha permesso di automatizzare attività tradizionalmente svolte manualmente, migliorando la sicurezza e l'affidabilità e riducendo il costo del lavoro. Si tratta principalmente di *robot* che svolgono attività di: (a) pulizia in ambiente ospedaliero [47]; (b) distribuzione automatica dei farmaci [48]; (c) distribuzione automatizzata dei pasti; (d) gestione automatizzata di magazzini.

Di seguito sono riportati alcuni esempi:

- Care-O-bot: sviluppato dal Fraunhofer Institute for Manufacturing Engineering and Automation, è un *robot* capace di assistere attivamente lo svolgimento di attività quotidiane in ambiente ospedaliero

[45] I servizi di assistenza professionale sono un importante istituzione sociale. Nei paesi del Nord Europa, come i Paesi Bassi, la Svezia e il Regno Unito, la forza lavoro impiegata in quest'area è pari al 8-10% di tutti i lavoratori. Gli esperti stimano che circa il 70% delle persone di età superiore ai 70 anni non sono in grado di effettuare almeno una o due attività di routine quotidiane senza supporto (Eurofound, 2006).In Europa, secondo gli ultimi dati disponibili, il numero di *robot* impiegati nel sostegno dei servizi di assistenza professionale è aumentato nell'arco di tre anni, dal 2002 al 2005, da 12.400 a 25.500 (European Robotics Research Network (EURON), Technology Roadmap). Veruggio, G. «The EURON Roboethics Roadmap.» In: *Humanoids*. Citeseer. 2006, pp. 612–617

[46] Della robotica di assistenza individuale, finalizzata all'autonomia e all'indipendenza degli individui, ci siamo occupati nella sezione precedente.

[47] Alcuni esempi sono i *robot* per la pulizia delle sale operatorie o delle stanze dei pazienti.

[48] Tali sistemi tengono traccia dei farmaci, li distribuiscono ai pazienti, monitorandone la *compliance*

e domestico, ma anche ad esempio all'interno di strutture turistiche recettive [49]. (Disponibile in commercio)

- TUG: è un sistema, sviluppato dalla Aethon, che si occupano di trasporto di materiali come, ad esempio, cibo, farmaci, biancheria e immondizia [50]. (Disponibile in commercio)

Generalmente, i sistemi utilizzati in questa sotto-area si basano su tecnologie sviluppate e utilizzate anche in ambiti diversi da quello sanitario.

(II) *Sistemi di monitoraggio dei pazienti*

Una delle attività infermieristiche più comuni consiste nel monitorare i pazienti e assistere i medici nei loro incontri quotidiani con i pazienti stessi. Le tecnologie sviluppate in questo ambito supportano gli infermieri, in ambiente ospedaliero, case di cura e riposo, ad esempio, nel monitorare a distanza i pazienti. Possono essere utilizzati anche fungendo da mezzo, per i pazienti, per consultare medici, infermieri e personale sanitario, senza che essi siano fisicamente presenti nello stesso luogo. Si tratta principalmente di sistemi che svolgono attività di: (a) monitoraggio dei pazienti; (b) ausilio per il consulto virtuale di personale medico.

I primi possono essere molto semplici o particolarmente sofisticati. Fanno parte dei primi i sistemi che monitorano i segni vitali del paziente, come ad esempio l'assenza di movimento e altri segnali facilmente rilevabili, e li trasmettono al personale medico. Si tratta di tecnologie che possono essere utilizzate sia in ambiente professionale che domestico. I sistemi più avanzati sono, invece, in grado di eseguire in remoto analisi sofisticate delle condizioni del paziente e trasmetterle al medico.

Il secondo tipo di sistemi consentono al paziente di consultare a distanza il medico, fisicamente non presente nello stesso luogo in cui si trova il paziente, ad esempio tramite connessione audio-video.

Di seguito sono riportati alcuni esempi.

[49] Nel 2015, la quarta generazione Care-O-bot ha ricevuto ilRed Dot Award: Product Design. Per approfondimenti sull'uso di Care-O-bot in ambito sanitario, si leggano ad esempio: Schaeffer, C. e May, T. «Care-o-bot-a system for assisting elderly or disabled persons in home environments». In: *Assistive technology on the threshold of the new millenium* (1999); Nejat, G., Sun, Y. e Nies, M. «Assistive robots in health care settings». In: *Home health care management and practice* 21.3 (2009), pp. 177–187

[50] Per approfondimenti, si veda Niechwiadowicz, K. e Khan, Z. «Robot based logistics system for hospitals-survey». In: *IDT Workshop on Interesting Results in Computer Science and Engineering*. Citeseer. 2008; Bloss, R. «Mobile hospital robots cure numerous logistic needs». In: *Industrial Robot: An International Journal* 38.6 (2011), pp. 567–571

- RP7: prodotto dalla InTouch Health è un sistema di cd. presenza remota; permette al medico seduto di fronte all'unità di controllo TotalView di connettersi tramite internet ad un *Robot* e interagire con pazienti, familiari, personale medico, dovunque si trovi. Permette di aumentare le possibilità di cura e gestione del paziente. L'utilizzo di connessioni wireless all'interno dell'ospedale permette a RP-7 di muoversi liberamente in ogni reparto con l'aiuto di alcuni sensori; l'utilizzo di telecamere ad alta risoluzione garantisce immagini molto nitide e ingrandite del paziente; webcam posizionate sulle unità di controllo permettono al malato di vedere il medico durante le visite mentre sistemi di microfoni e amplificatori rendono possibile il dialogo diretto.[51]. (Disponibile in commercio)
- Mir-H: sviluppato dalla MOST I TECH, consente la comunicazione a distanza tra medico e paziente, in ambiente domestico. (Disponibile in commercio).

(III) *Sistemi di ausilio per attività di assistenza fisica*

Prendersi cura dei pazienti comprende ad esempio attività di pulizia e igiene dei pazienti e degli ambienti. Spesso questo tipo di attività richiede il sollevamento di pesi o l'esercizio di una notevole forza fisica. Per alleviare lo sforzo fisico del personale infermieristico e prevenire infortuni connessi al lavoro, posso essere utilizzate tecnologie in grado di coadiuvare il personale o addirittura sostituire gli infermieri nello svolgimento di compiti specifici. Questo tipo di sistemi deve essere in grado di relazionarsi e gestire con la massima cura i pazienti. Si possono distinguere (a) sistemi di supporto e assistenza al personale infermieristico per il sollevamento dei pazienti; (b) sistemi autonomi disollevamento dei pazienti [52].

Di seguito sono riportati alcuni esempi.

- RI-Man: sviluppato dalla RIKEN, è un *robot* interattivo, umanoide; è dotato di morbidi sensori tattili areali che misurano la grandezza e

[51] Per approfondimenti si leggano Agarwal, R. et al. «The RoboConsultant: telementoring and remote presence in the operating room during minimally invasive urologic surgeries using a novel mobile robotic interface». In: *Urology* 70.5 (2007), pp. 970–974; Marttos, A. et al. «Usability of telepresence in a level 1 trauma center». In: *Telemedicine and e-Health* 19.4 (2013), pp. 248–251

[52] Questi sistemi sono considerati socialmente meno accettabili, a causa dei problemi di sicurezza e accettazione da parte del paziente

la posizione della forza di contatto, consentendogli di interagire fisicamente e in sicurezza con gli esseri umani. È in grado di elaborare informazioni audio, video, e olfattive [53].

- ETL-Humanoid: sviluppato dall'Università di Tokyo, è un *robot* umanoide interattivo, molto versatile e capace di svolgere molteplici attività di assistenza. [54].
- C-PAM: sviluppato dalla Daihen Co. Ltd., è un sistema molto semplice ed efficace per spostare un paziente da un letto all'altro [55].

L'introduzione di questi sistemi nel sistema sanitario è una grande sfida. L'accettazione sociale da parte di operatori e pazienti sembra essere migliore nei paesi asiatici e ancora piuttosto bassa in Europa.

(IV) *Sistemi di ausilio ad attività paramediche*

Il lavoro degli operatori sanitari prevede lo svolgimento di un numero elevato di attività ricorrenti e ripetitive, come ad esempio l'analisi di campioni corporei, la misurazione della pressione sanguinea e la medicazione di ferite, che tuttavia richiedono grande attenzione a causa dei potenziali rischi per la salute dei pazienti e degli standard di qualità richiesti.

Un esempio è:

- PERROB: sviluppato dalla Vision Dynamics, è in grado di medicare e applicare bendaggi.

La figura 2.1 fornisce una mappa concettuale sintetica delle aree e sotto-aree della robotica e dei sistemi di automazione nell'e-health, così come sono state presentate.

Dall'esame delle tecnologie, in uso o di prossima introduzione, nel settore sanitario è emersa la difficoltà di ricomprenderle tutte all'interno di un'unica categoria concettuale, capace di definirne le caratteristiche in modo univoco. Esse

[53] Per approfondimenti si veda Onishi, M. et al. «Generation of human care behaviors by human-interactive robot RI-MAN». in: *Robotics and Automation, 2007 IEEE International Conference on*. IEEE. 2007, pp. 3128–3129; Mukai, T. et al. «Development of the tactile sensor system of a human-interactive robot RI-MAN». in: *IEEE Transactions on Robotics* 24.2 (2008), pp. 505–512

[54] Per un approfondimento sulle caratteristiche e sulle abilità del *robot* si legga Nagakubo, A., Kuniyoshi, Y. e Cheng, G. «Etl-humanoid-a high-performance full body humanoid system for versatile actions». In: *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*. Vol. 2. IEEE. 2001, pp. 1087–1092; Nagakubo, A., Kuniyoshi, Y. e Cheng, G. «The ETL-Humanoid system—a high-performance full-body humanoid system for versatile real-world interaction». In: *Advanced robotics* 17.2 (2003), pp. 149–164

[55] Si veda Wang, H. e Kasagami, F. «A patient transfer apparatus between bed and stretcher». In: *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 38.1 (2008), pp. 60–67

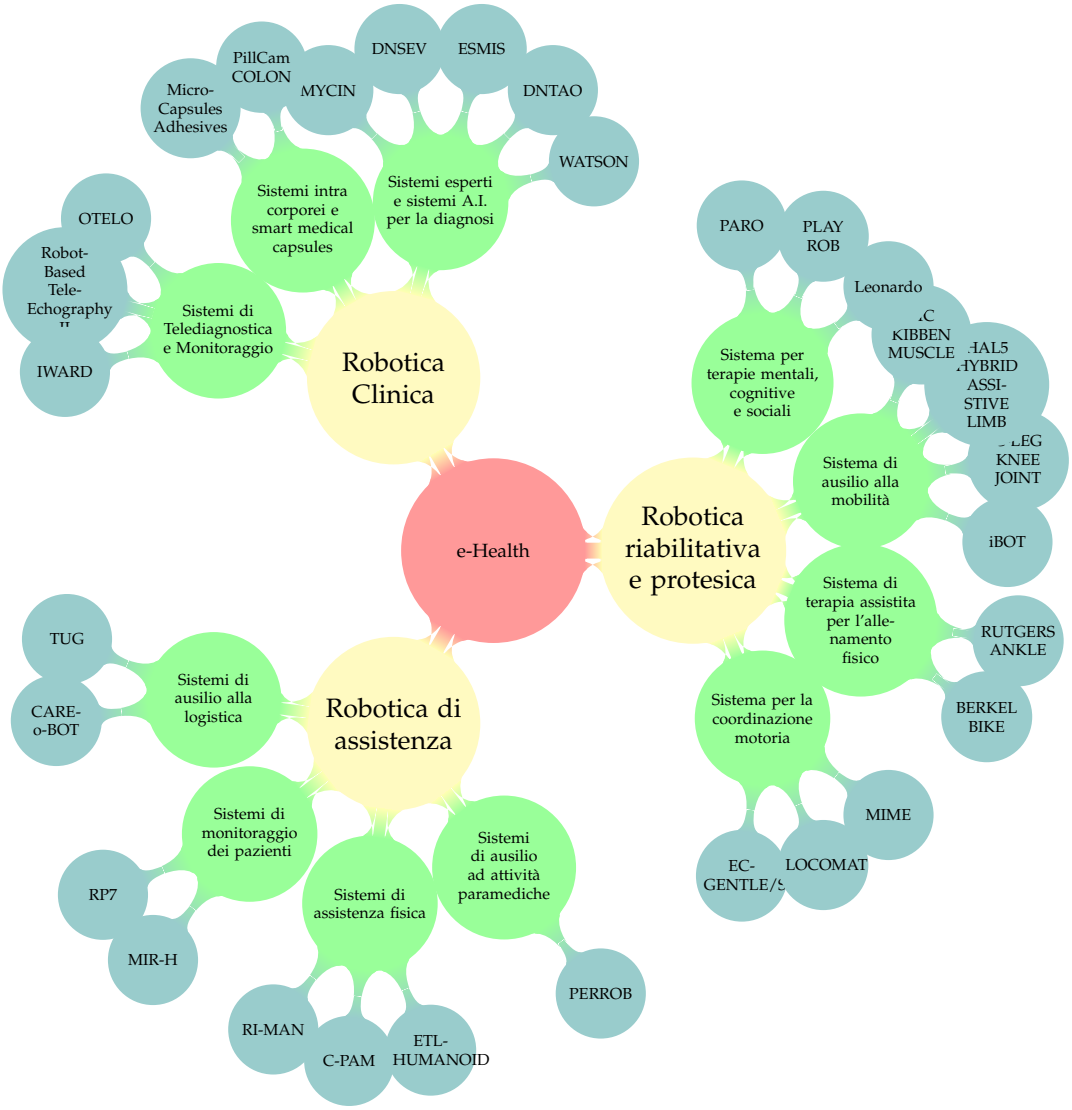


FIGURA 2.1: Mappa delle aree e dei sistemi di automazione nell'e-Health

hanno proprietà estremamente eterogenee e differenti tra loro, pur essendo accomunate da un elevato grado di automazione. Alcune di esse possono certamente essere ricomprese nella nozione di robot, inteso come sistema dotato di alcune funzioni essenziali, quali la capacità di agire su stimoli ambientali in combinazione con rilevamento e ragionamento logico, la cui funzione principale risiede nell'automazione del lavoro fisico, come per esempio i robot chirurgici. Tuttavia restano fuori da tale definizione tutti i sistemi privi di caratteristiche meccatroniche, come per esempio sistemi esperti e altri sistemi d'intelligenza artificiale. Inoltre, la maggior parte dei sistemi d'intelligenza artificiale è in grado di operare autonomamente senza l'ausilio di esseri umani, nello svolgimento dei propri compiti, per esempio Watson, mentre la maggior parte dei sistemi robotici in area sanitaria, per esempio il Da Vinci, hanno bisogno di un operatore umano per poter espletare tali compiti. Alcune tecnologie si caratterizzano principalmente come strumenti di ausilio per l'operatore umano, altri, pur mantenendo tale caratteristica, si atteggiano come agenti che operano autonomamente all'interno di un ambiente predeterminato.

Tali proprietà non possono che riflettersi sia sui livelli di automazione dei singoli sistemi, sia sui profili di responsabilità che derivano dal loro impiego. Tali aspetti saranno approfonditi nei capitoli successivi trattando il tema della responsabilità.

2.5 I sistemi socio-tecnici

La teoria dei sistemi socio-tecnici (STS) nasce agli inizi degli anni cinquanta nell'ambito di una ricerca condotta dal Tavistock Institute di Londra, sulla riorganizzazione del lavoro all'interno dell'industria mineraria inglese [56]. Il termine fu coniato per descrivere i sistemi che implicano una complessa interazione tra l'uomo, le macchine e gli aspetti ambientali del sistema lavoro. A partire dall'analisi di come la mecanizzazione avesse abbassato la produttività dei lavoratori, Eric Trist, allora ricercatore al Tavistock, argomentò che i sistemi produttivi hanno aspetti non solo tecnici, ma anche umani e sociali, collegati e interconnessi tra loro, e dalle porprietà di questa interconnessione, ovvero dal modo in cui si specifica la natura del sistema, si detrmina la perfomance del sistema nel suo complesso [57].

Il concetto di "sistema", alla base del lavoro dei ricercatori del Tavistock Institute, era stato sviluppato dal biologo austriaco Ludwig Von Bertalanffy (1950),

[56] Per approfondimenti si legga Emery, F. e Trist, E. «Sistemi socio-tecnici». In: *Progettazione e sviluppo delle organizzazioni* (1974).

[57] Ropohl, G. «Philosophy of socio-technical systems». In: *Techné: Research in Philosophy and Technology* 4.3 (1999), pp. 186–194

fondatore della teoria generale dei sistemi, il cui articolo *"Open systems in physics and biology"* era da poco stato pubblicato [58]. Secondo Von Bertalanffy, tutti i sistemi viventi sono sistemi aperti, identità complesse di unità in interazione, a loro volta impegnate in un continuo scambio di energia e materia con l'ambiente. L'interazione ininterrotta tra sistema e ambiente, intesa come flusso ininterrotto di input-output, rende il sistema un'entità complessa in continua evoluzione, caratterizzata da equilibri dinamici[59]. Diversamente, i sistemi chiusi si caratterizzano per l'assenza di scambi vitali con l'ambiente.

.

2.5.1 Le origini e il lavoro dei ricercatori del Tavistock Institute di Londra

Dopo la seconda guerra mondiale, un gruppo di ricerca installato a Londra, conquistò grande notorietà nel campo della teoria delle organizzazioni, diffondendo la nozione di sistema socio-tecnico. Uno dei primi studi basati su questa formula fu condotto da Eric Trist e Ken Bamforth, a seguito delle difficoltà incontrate in quel momento, in Gran Bretagna, dai responsabili della modernizzazione produttiva delle miniere di carbone.

La ricerca paragonava due tipi differenti di organizzazione lavorativa. In particolare, i ricercatori del Tavistock Institute confrontarono il metodo tradizionale di lavoro, cosiddetto *"shortwall"*, e il nuovo metodo *"longwall"*, tecnologicamente avanzato, che avrebbe dovuto portare a un aumento della produzione giornaliera.

Secondo il metodo tradizionale, gli operai venivano divisi in piccole squadre che si occupavano di tutte le fasi di produzione: preparazione, estrazione e trasporto. Il carbone veniva quindi tagliato a mano e accumulato negli spazi liberi. Veniva caricato su carrelli o su un nastro trasportatore per essere portato in superficie e, infine, tutto il materiale veniva trasportato all'esterno.

Il metodo delle pareti lunghe, che aveva preso piede a seguito della meccanicizzazione delle tecniche di estrazione, permetteva - a parità di condizioni tecniche - di riprodurre nelle miniere i modelli di coordinamento e di direzione in uso nell'organizzazione industriale del lavoro alla catena di montaggio. I lavoratori venivano quindi assegnati ad un unico compito, in un'unica postazione in parete lunga, senza alcuna opportunità di rotazione nel lavoro.

[58] Von Bertalanffy, L. et al. «The theory of open systems in physics and biology». In: *Science* 111.2872 (1950), pp. 23–29.

[59] Von Bertalanffy, L. «The history and status of general systems theory». In: *Academy of Management Journal* 15.4 (1972), pp. 407–426:412.

Tra le variabili sottoposte ad analisi da Trist e i suoi collaboratori vi erano le unità operative, i sistemi di rifornimento, la percezione dei ruoli lavorativi, i sistemi di ricompensa e i bisogni psicologici dei lavoratori.

In un primo articolo del 1951 Trist e Bamforth mostrarono come l'adozione del metodo *longwall* non fosse appropriata al lavoro nelle miniere di carbone [60]. L'organizzazione messa in opera era infatti molto più rigida di quanto potesse sopportare un lavoro sottoposto per definizione a incessanti irregolarità, come per esempio cambiamenti nella natura della vena da sfruttare o nella resistenza del terreno. Il risultato era che i minatori, destinati in linea di principio a compiti rigidamente definiti e coordinati tra loro, dovevano continuamente intervenire per cercare di far fronte agli imprevisti. Il metodo della parete lunga aveva creato forti differenze di status tra i minatori che si occupavano delle diverse fasi di produzione, non era stato in grado di autoregolarsi, aveva portato a negoziazioni separate con la direzione e a criteri differenti per la determinazione del salario[61].

In questo studio, divenuto rapidamente un punto di riferimento per tutti gli specialisti, Trist e Bamforth delinearono alcune idee fondamentali.

In primo luogo, il modello burocratico centralistico non è sinonimo di alcuna garanzia di efficienza, anche nel caso in cui segua a un programma di modernizzazione. Nel caso delle miniere di carbone l'adozione di questo modello si era anzi rivelata anti-produttiva. Si trattava, secondo Trist e Bamforth, di sostituire la parcellizzazione, la completa etero-direzione e l'isolamento della mansione fordista con un'organizzazione del lavoro in gruppi, cui fosse attribuita la responsabilità di una porzione di attività di dimensioni significative. L'esecuzione di questo modello comporta scelte di divisione del lavoro tra i componenti del gruppo e scelte relative ai ritmi di lavoro. Scelte che il gruppo di lavoro è in grado di assumere in maniera semi-autonoma e senza pregiudicare l'efficienza complessiva, a condizione che i singoli componenti il gruppo siano formati allssumere più compiti e ruoli in modo da poter ruotare tra gli stessi, in funzione delle esigenze di produzione. Un principio di design basato sulla ridondanza delle funzioni, piuttosto che delle parti tende a sviluppare più competenze nell'individuo aumentando le capacità di risposta dell'intero gruppo.

In secondo luogo, due differenti ordini di variabili, tecniche e sociali, concorrono in pari misura a definire il sistema produttivo. Gli esseri umani sono quindi complementari e non subordinati alle macchine, con risorse professionali e sociali da sviluppare all'interno di sistemi cooperativi. La tecnologia disponibile non impone quindi un solo modello organizzativo. È possibile scegliere

[60] Trist, E. e Bamforth, K. «Some social and psychological consequences of the Longwall method». In: *Human relations* 4.3 (1951), pp. 3–38.

[61] ibid.

tra diversi modelli di organizzazione del lavoro quello più adatto a conciliare le esigenze tecniche con quelle del sistema. Infine, ogni organizzazione aziendale è un "sistema aperto", così come definito da von Bertalanffy, e l'equilibrio va ricercato nell'interscambio con il contesto.

Per la prima volta, attraverso il concetto di sistema socio-tecnico, l'organizzazione di una fabbrica non fu più vista come un sistema rigidamente determinato dalla tecnologia, ma come un sistema dinamico e variabile in base alla soluzione organizzativa più adeguata. Trist sviluppò il concetto di sistema socio-tecnico fino a farne uno schema capace di investire l'intera strategia manageriale. In un ambiente organizzativo che diventa sempre più complesso, turbolento e impredicibile [62], la necessità di assicurare affidabilità è di essenziale importanza e implica che tutte le parti di un'organizzazione siano interrelate non solo tra loro, ma anche con l'esterno, esaminando i cambiamenti che si verificano ed elaborando un approccio adattivo per rispondere a tutte quelle situazioni critiche che risultano inaspettate e imprevedibili [63].

In un celebre articolo del 1976, Cherns, che al tempo lavorava al Tavistock, riassunse i principi chiave per il design e la realizzazione di una struttura socio-tecnica[64]:

- Principio 1: Compatibilità.

  Il processo di design organizzativo deve essere compatibile con gli obiettivi. Se lo scopo è realizzare una struttura democratica, in grado di adattarsi al cambiamento e di utilizzare le capacità creative degli individui, deve essere messo in atto un processo democratico: la partecipazione degli individui al processo di riorganizzazione del lavoro sarà quindi imprescindibile;

- Principio 2: Specifiche critiche minime.

  Nulla più di ciò che è essenziale deve essere specificato. Tale principio riguarda l'allocazione critica dei compiti in base ai ruoli e alla specificazione degli obiettivi e delle metodologie più adatte per raggiungerli. Se è vero che l'eccessiva specificazione dei compiti, da un lato, riduce l'incertezza, portando alcuni vantaggi, dall'altro, riduce il grado di apertura

[62] Si vedano Emery, F. E., Trist, E. L. et al. «The causal texture of organizational environments». In: *Human relations* 18.1 (1965), pp. 21–32; e Davis, L. E. «Evolving alternative organization designs: their sociotechnical bases». In: *Human Relations* 30.3 (1977), pp. 261–273.

[63] Si veda Clark, J. V. e Krone, C. G. «Towards an overall view of organizational development in the early seventies». In: *Thomas, JM/Bennis, WG: The Management of change and conflict. Harmondsworth. S. 284f* (1972).

[64] Cherns, A. «The Principles of Sociotechnical Design1». In: *Human relations* 29.8 (1976), pp. 783–792.

all'innovazione e le possibilità di successo dell'organizzazione. Tale successo dipende dalle idee e dalle abilità messe in atto dall'organizzazione e dai singoli;

- Principio 3: Criterio sociotecnico.

  Si tratta di un principio legato al concetto di varianza. Esso identifica la deviazione rispetto a regole e standard attesi. Si tratta di eventi non programmati e inaspettati, come per esempio l'interruzione di un sistema o di una macchina. Se tali deviazioni non possono essere previste ed eliminate, devono essere controllate e gestite nel punto più vicino possibile alla loro origine. Problemi di questo tipo dovrebbero essere risolti dal gruppo che li sperimenta e non da altri gruppi, come per esempio un gruppo di vigilanza. Il processo di identificazione e comprensione della varianza associata allo svolgimento di un determinato compito richiede una lunga analisi e sforzi per poterla codificare in maniera corretta sulla base dell'aspetto tecnico e sociale;

- Principio 4: Principio multifunzionale.

  La ridondanza delle funzioni è necessaria per la capacità di adattamento e apprendimento. Affinchè i gruppi siano flessibili e in grado di rispondere ai cambiamenti, è necessaria la combinazione di funzioni e la varietà di competenze all'interno del gruppo. L'organizzazione quindi deve essere considerata un organismo complesso, costituito da differenti elementi (umani e tecnologici) che sono chiamati ad interagire e collaborare;

- Principio 5: Linea di confine.

  Ogni organizzazione è chiamata a delineare i propri confini. Essi dovrebbero facilitare la condivisione di conoscenze ed esperienze. Secondo Miller le organizzazioni solitamente definiscono i propri confini, così come i gruppi di persone e le attività che sono chiamate a svolgere, sulla base

di uno dei tre criteri seguenti: tecnologia [65], territorio [66] e tempo [67]. Essi dovrebbero essere presenti ove vi sia una discontinuità naturale - tempo, cambiamento tecnologico e territorio - nel processo di lavoro. Vi sono confini quando le attività lavorative passano da un gruppo ad un altro e sono necessarie nuove attività e competenze. tutti i gruppi dovrebbero imparare reciprocamente, nonostante la presenza di confini.

- Principio 6: Flusso di informazioni.

  Le informazioni devono giungere dove è necessaria un'azione. Esse devono essere sempre disponibili in qualunque luogo e in qualunque momento per poter essere utilizzabili, per esempio per attività di controllo o di routine.

- Principio 7: Supporto.

  Sulla base dei principi espressi, e in particolare del sesto, è necessario adottare una nuova filosofia improntata alla collaborazione e al supporto tra le diverse unità. Anche i singoli individui dovranno adottare un nuovo comportamento orientato all'interazione costante. Quindi, se il sistema è composto da gruppi diversi, tutti gli individui appartenenti alle unità, così come il management, dovranno interagire e collaborare. È necessario eliminare ogni status di differenziazione tra il management e i lavoratori, per enfatizzare la collaborazione e il supporto e far sì che si diffonda nell'organizzazione;

- Principio 8: Design e valori umani.

[65] Il raggruppamento sulla base della tecnologia è tipico delle organizzazioni in cui vengono utilizzate tecnologie differenti, poiché esse saranno suddivise in base al tipo e alla funzione per e poste sotto la supervisione di esperti specializzati in un certo ambito.

[66] Un esempio di suddivisione sulla base del territorio è rappresentato dalle aziende che, pur avendo dei distaccamenti sul territorio, localizzano tutte le attività nella medesima area, per riduure i tempi di realizzazione dei prodotti. Questa suddivisione può determinare difficoltà per la condivisione delle informazioni, della conoscenza e per l'apprendimento. È necessario saper gestire le relazioni tra dipartimenti ed unità e tra questi e l'organizzazione nel suo complesso. Il controllo delle attività di ciascun dipartimento diventa responsabilità dei membri dell'unità stessa e la supervisore si focalizza sull'attività e non sull'intera unità. Il team di lavoro ha le capacità per poter gestire tutte le funzioni, prendere decisioni e coordinare attività. In circostanze ottimali i gruppi di lavoro possono acquisire anche completa autonomia nel gestire i propri confini e fare del supervisore una risorsa di cui usufruire per il conseguimento di determinati obiettivi;

[67] Miller, E. J. «Technology, territory, and time.» In: *Human Relations* 12 (1959), pp. 245–272

Riguardo ai valori umani e al design organizzativo Thorsrud individua sei principi chiave [68]: 1) la realizzazione del lavoro richiede il pieno coinvolgimento del personale così da poter contare su elementi innovativi; 2) sviluppo di capacità di learning continuativo e costante; 3) individuazione di aree di decision making; 4) gradi di supporto e riconoscimento del singolo nel contesto lavorativo; 5) possibilità di correlazione tra ciò che l'individuo fa e ciò che produce in termini di impatto sulla propria vita sociale; 6) sviluppo di un atteggiamento ottimistico nei confronti del lavoro in termini di vantaggi per il proprio futuro;

- Principio 9: Incompletezza, evoluzione e apprendimento continuo.

  Il design è un processo iterativo. Il processo di ridefinizione dell'organizzazione è continuo, richiede costanti interventi e presenta delle continue possibilità di miglioramento e di implementazione.

2.5.2 Gli sviluppi della teoria socio-tecnica e il concetto di automazione

Tra gli anni settanta e ottanta, furono condotti numerosi studi sulla teoria dei sistemi socio-tecnici in moltissimi paesi europei, come Svezia, Norvegia, Germania e Paesi Bassi[69]. Anche i primi esperimenti di adozione di questo approccio risalgono agli stessi anni [70]. Le teorie sociotecniche erano ormai condivise e accettate nelle nazioni più industrializzate e all'interno di organizzazioni di natura diversa, dal settore manifatturiero a quello dei servizi. Era ormai maturata la consapevolezza che fattore umano e fattore tecnico non potevano più essere considerati come aspetti operanti in direzione opposta. Il design organizzativo

[68] Thorsrud, E. «Policy-making as a learning process in working life». In: *Working life* (1981), pp. 313–327

[69] Per un approfondimento si legga Mumford, E. «The story of socio-technical design: Reflections on its successes, failures and potential». In: *Information Systems Journal* 16.4 (2006), pp. 317–342.

[70] Tra i pionieri nell'applicazione della teoria socio-tecnica possiamo ricordare la Volvo, in Svezia. Nel corso degli anni settanta l'azienda automobilistica decise di rimuovere il tradizionale sistema di produzione, sostituendolo con dei gruppi di lavoro il cui compito era assemblare un'intera automobile, lavorare, coordinarsi e controllare l'intera attività lavorativa pur essendo distanti sia in termini di spazio, sia temporalmente [71]. Questo fu reso possibile grazie ad eccellenti sistemi informativi che garantivano la trasmissione di informazioni e il raggiungimento di tutti gli individui appartenenti al gruppo. In Italia i primi esperimenti furono fatti dall'Olivetti, con la realizzazione delle "isole di lavoro", un modello in cui i cicli di lavoro si ricompongono in fasi molto più lunghe e vengono affidati a singoli operai o a team di lavoro con responsabilità di controllo, programmazione e manutenzione, prima separate.Butera, F. e De Witt, G. «Valorizzare il lavoro per rilanciare l'impresa». In: *La storia delle isole di produzione alla Olivetti negli anni'70* (2011).

della gestione di sistemi complessi era diventato un argomento di ampio studio. Il lavoro sugli incidenti e i sistemi di sicurezza aveva abbandonato la visione secondo cui gli incidenti erano il risultato dell'errore umano o del malfunzionamento di una macchina, singolarmente considerati, e il modello elaborato da Jim Reason, cosìddetto "Swiss Cheese" era diventato prevalente[72]. Così gli incidenti sono considerati come il risultato di una complessa concatenazione di eventi, a livelli differenti, capace di rompere le difese e le barriere erette per evitare eventi catastrofici. Questo modello, adottato ancora oggi in molti settori come per esempio l'aviazione e la sanità, ha contribuito a spiegare il funzionamento dei sistemi complessi. Anche l'automazione, davanti alle scoperte tecnologiche e all'aumento della complessità dei sistemi ingegneristici, inizia ad essere vista secondo un'ottica sistemica.

Nei primi anni sessanta, in un lavoro dal titolo *Vers l'automatisme social? Problèmes du travail et de l'automatisation*, Naville definisce l'automazione non come un concetto di natura puramente tecnica, ma come un'organizzazione avanzata dove ogni tecnologia è un sistema di concetti. Le realizzazioni tecniche sarebbero quindi effetti e risultati[73]. Questa visione si pone in netto contrasto con il concetto di automazione come tecnologia, come sostenuto da Brith e Crossman [74], secondo cui l'automazione sostituirebbe sempre più lo svolgimento di funzioni che dovrebbero appartenere agli esseri umani, mediante il controllo automatico dei processi. Secondo l'approccio socio-tecnico di Naville, l'automazione è invece un un sistema tecnico, organizzativo, flessibile e capace di controllo. Il sistema nel suo complesso è il risulato di singole macchine (automazione come sviluppo tecnico), capaci di sostituire il lavoro umano (automazione come tecnologia) e integrate in un unico sistema di controllo (automazione come integrazione): il sistema rivelerebbe capacità di apprendimento, di evoluzione e di creazione, divenetando quindi anche autoreferenziale e al tempo stesso flessibile, capace

[72] Reason, J. T. *Managing the risks of organizational accidents*. Vol. 6. Ashgate Aldershot, 1997. Sul modello elaborato da Reason si veda la sezione 2.7.

[73] Naville, P. *Vers l'automatisme social?: problèmes du travail et de l'automation*. Gallimard, 1963.

[74] Bright, J. R. *Automation and management*. Division of Research, Graduate School of Business Administration, Harvard University Boston, 1958 e Crossman, E. *Automation and skill, dsir, Problems of Progress in Industry (9), London. Reprinted in Edwards et Lees F.(eds), The Human Operator in Process Control*. 1974.Secondo questi autori l'automazione è una particolare tecnologia che permette di sostituire, mediante il controllo automatico dei processi, funzioni che dovrebbero appartenere all'uomo. Butera, inButera, F. «Note sulla storia dell'automazione. Dall'impatto sociale dell'automazione alla progettazione congiunta di tecnologia, organizzazione e sviluppo delle persone». In: *STUDI ORGANIZZATIVI* (2014), nota che una tale visione dell'automazione avrebbe poco a che vedere con le proprietà costruttive delle macchine stesse, focalizzandosi al contrario sulle loro prestazioni. Questo approccio, nonostante offra un notevole contributo alle analisi del progresso tecnologico, sarebbe insoddisfacente per ciò che riguarda la descrizioni dell'unità tecnico-organizzativa e l'analisi delle ragioni dello sviluppo tecnico.

di evolversi e di adattarsi all'ambiente. L'automazione diventa così un fenomeno complesso di natura tecnologica economica, organizzativa e sociale, che ha per oggetto la gestione e l'evoluzione di complessi sistemi tecnico-organizzativi che realizzano processi produttivi[75]. Oggi, anche l'ingegneria sociale ha accolto l'idea che l'automazione e i sistemi ingegneristici complessi, non sono costituiti da sole macchine ma, dipendono in modo critico dal fattore umano.

2.6 L'interazione uomo-macchina

I gravi problemi connessi all'automazione e i numerosi errori legati all'interazione uomo-macchina sono da tempo documentati in letteratura[76] e sono stati spesso associati a deficit di vigilanza, mancanza di cura e perdita di una chiara e corretta percezione di ciò che accade o può accadere nell'immediato futuro, cd. *Situation Awareness*(SA), da parte degli operatori umani[77].

Un fattore chiave capace di contribuire in modo rilevante ai problemi di performance degli esseri umani, che operano e interagiscono con sistemi automatici complessi, riguarda le cosiddette performance *out-of-the-loop* (OOTL)[78]. I problemi di performance *out-of-the-loop* sono caratterizzati da una minore capacità dell'operatore umano di intervenire e di assumere il controllo manuale di sistemi complessi e di *control loops* automatici. È possibile identificare due classi principali di problemi. La prima è relativa alla difficoltà degli operatori umani, con

[75]Così Butera, F. *Il castello e la rete. Impresa, organizzazioni e professioni nell'Europa degli anni'90*. Vol. 1. FrancoAngeli, 2005. L'autore giunge a questa conclusione dopo aver analizzato la definizione di automazione secondo quattro concezioni diverse: l'automazione come tipo particolare di sviluppo tecnico, l'automazione come tecnologia, l'automazione come forma d'integrazione della produzione e dell'impresa e l'automazione come sistema socio-tecnico capace di autoregolazione e di adattamento.

[76]Cfr. Moray, N., Inagaki, T. e Itoh, M. «Adaptive automation, trust, and self-confidence in fault management of time-critical tasks.» In: *Journal of Experimental Psychology: Applied* 6.1 (2000), p. 44; Sarter, N. B. e Woods, D. D. «How in the world did we ever get into that mode? Mode error and awareness in supervisory control». In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1 (1995), pp. 5–19.

[77]Si vedano gli studi condotti da Endsley, M. R. «The application of human factors to the development of expert systems for advanced cockpits». In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 31. 12. SAGE Publications. 1987, pp. 1388–1392; Endsley, M. R. e Kiris, E. O. «The out-of-the-loop performance problem and level of control in automation». In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.2 (1995), pp. 381–394; Parasuraman, R. et al. «Adaptive function allocation reduces performance cost of static automation». In: *7th International Symposium on Aviation Psychology*. DTIC Document. 1993, pp. 37–42; Parasuraman, R. e Riley, V. «Humans and automation: Use, misuse, disuse, abuse». In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39.2 (1997), pp. 230–253.

[78]Kessel, C. J. e Wickens, C. D. «The transfer of failure-detection skills between monitoring and controlling dynamic systems». In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 24.1 (1982), pp. 49–60.

funzioni di controllo, di rilevare errori di sistema e di intervenire manualmente in caso di guasti o malfunzionamenti[79]. Al tempo necessario affinché l'operatore rilevi la presenza di un errore si aggiunge un periodo di tempo significativo affinché comprenda lo stato del sistema e si riorganizzi per agire in maniera appropriata. Questo ritardo è in grado di incidere sia sulla possibilità che l'operatore intervenga, sia sull'efficacia delle azioni intraprese. Wickens e Kessel hanno condotto studi di laboratorio, dimostrando tempi più lunghi per la ripresa del normale funzionamento del sistema e una scarsa precisione e accuratezza di intervento e risposta degli operatori davanti a eventi critici[80]. Per superare i problemi creati da approcci storicamente centrati sulla tecnologia e l'automazione, è stata proposta una filosofia dell'automazione *human-ceneterd* che prevede un metodo di progettazione centrato sull'utente e partecipativo. Questo tipo di approccio è ormai altamente diffuso in molti domini e in particolare nell'aviazione e nell'healthcare[81]. L'obiettivo di un design centrato sull'essere umano è quello di facilitare le funzioni di controllo e gestione del sistema da parte dell'operatore umano. Si tratta quindi di destinare sia agli esseri umani, sia ai sistemi automatici, i compiti a loro più adatti, per ottenere la migliore combinazione tra controllo automatico e controllo umano[82]. Come postulato da Sheridan e Verplanck, l'automazione non è "tutto o niente", non si tratta di automatizzare del tutto o solo parzialmente una certa attività o un dato compito, ma di stabilirne il livello di automazione[83]. L'assegnazione dei compiti dovrebbe essere assegnata in modo da ottenere un lavoro di squadra tra esseri umani e sistemi automatici[84] Questo tipo di approccio permette di evitare che l'operatore umano svolga un lavoro frammentato e difficile da gestire. Così, negli ultimi decenni sono state sviluppate tassonomie dei livelli di automazione per attività psicomotorie e cognitive, per chiarire il range di opzioni tra automazione e assenza di automazione. Alcuni autori hanno chiarito che la decisione su cosa automatizzare non può

[79] In questo senso Wickens, C. D. e Kessel, C. «The effects of participatory mode and task workload on the detection of dynamic system failures». In: *Systems, Man and Cybernetics, IEEE Transactions on* 9.1 (1979), pp. 24–34.

[80] ibid.; Kessel e Wickens, «The transfer of failure-detection skills between monitoring and controlling dynamic systems», cit.

[81] Si vedano Billings, C. E. «Human-centered aircraft automation: A concept and guidelines». In: (1991); Degani, A. *Taming HAL: Designing interfaces beyond 2001*. Palgrave Macmillan, 2004; Parasuraman, R. E. e Mouloua, M. E. *Automation and human performance: Theory and applications*. Lawrence Erlbaum Associates, Inc, 1996.

[82] Sheridan, T. B. «Task analysis, task allocation and supervisory control». In: *Handbook of human-computer interaction* (1997), pp. 87–105.

[83] Sheridan, T. B. e Verplank, W. L. *Human and computer control of undersea teleoperators*. Rapp. tecn. DTIC Document, 1978

[84] Endsley, M. R. «Automation and situation awareness». In: *Automation and human performance: Theory and applications* (1996), pp. 163–181

essere semplicemente basata su un'allocazione delle funzioni per sostituzione[85]. Questo approccio, ormai superato, è stato applicato in passato anche attraverso le liste MABA-MABA (Men Are Better At – Machines Are Better At) elaborate da Paul Fitts nei primi anni cinquanta per il settore aeronautico[86]. Le scelta di avvalersi di una tassonomia dei livelli di automazione (LOAT-Level Of Automation Taxonomy), come strumento per l'analisi giuridica della responsabilità, risiede nella possibilità di:

(1) identificare il livello di automazione di una tecnologia, con riferimento a specifiche funzioni cognitive;

(2) determinare l'esatta divisione dei compiti tra uomo e macchina;

(3) investigare il tema della responsabilità associandola al livello di automazione e alla divisione dei compiti.

In particolare, la LOAT permette di definire il modo in cui le attività sono distribuite tra l'essere umano e la macchina e il grado di coinvolgimento di entrambi nel sistema[87], così da poter allocare in modo più corretto la responsabilità in caso di eventi dannosi.

2.7 LOAT: Tassonomia dei livelli di automazione

Sono state proposte numerose tassonomie e sistemi di classificazione dei livelli di automazione (LOA). Lo schema di classificazione più antico è quello proposto da Sheridan e Verplank alla fine degli anni settanta[88]. Esso prevede una scala di dieci punti, che rappresentano livelli crescenti di automazione. Al livello più basso della scala, l'essere umano prende decisioni e agisce senza alcuna assistenza. Al livello più alto, la macchina prende decisioni e agisce in completa autonomia, come mostrato nella tabella 2.1.

Tuttavia, la scala usata in questa prima tassonomia si limita a determinare una serie di punti distinguibili, lungo livelli continui di automazione, principalmente in relazione alle funzioni di output del processo decisionale e della scelta

[85] In questo senso Hollnagel, E. «From function allocation to function congruence». In: *Coping with computers in the cockpit(A 00-40958 11-54), Aldershot, United Kingdom and Brookfield, VT, Ashgate Publishing, 1999,* (1999), pp. 29–53.

[86] Fitts, P. M. «Human engineering for an effective air-navigation and traffic-control system.» In: (1951). Questi elenchi si basano sull'idea che, dato un insieme di attività preesistenti si decide quali di queste rendere automatiche, considerando i punti di forza di esseri umani e macchine

[87] Endsley, «Automation and situation awareness», cit.

[88] Sheridan e Verplank, *Human and computer control of undersea teleoperators*, cit.

TABELLA 2.1: Levels of Automation of decision and action selection - Sheridan and Verplanck (1978)

Automation Level	Automation Description
1	The computer offers no assistence: human must take all decisions and actions
2	The computer offers a complete set of decisions and actions alternatives, or
3	narrows the selection down to a few, or
4	suggest one alternative and
5	executes that suggestion if the human approves, or
6	allows the human a restricted time to veto before automatic execution, or
7	executes automatically, then necessarily informs human and
8	informs the human only if asked, or
9	informs the human only if it, the computer decides to.
10	The computer decides everything and acts autonomously, ignoring the human

del tipo di azione da mettere in atto. Mancava una specificazione dettagliata delle funzioni di input, relativa all'acquisizione delle informazioni.

Negli stessi anni Endsley e Kaber[89] costruiscono una tassonomia su dieci livelli, applicabile a una vasta gamma di settori e tipi di attività, riportata nella tabella 2.2. La loro tassonomia comprende quattro funzioni generiche, comparabili a quelle successivamente elaborate da Parasuraman, Sheridan, e Wickens, dove a ogni livello corrisponde una funzione o una combinazione di funzioni sia dell'essere umano, sia del sistema automatico.

TABELLA 2.2: Endsley and Kaber's (1999) - LOA taxonomy for human–computer performance in dynamic, multitask scenarios.

	Roles			
Level of Automation	Monitoring	Generating	Selecting	Implementing
(1) Manual Control	Human	Human	Human	Human
(2) Action support	Human/Computer	Human	Human	Human/Computer
(3) Batch processing	Human/Computer	Human	Human	Computer
(4) Shared control	Human/Computer	Human/Computer	Human	Human/Computer
(5) Decision support	Human/Computer	Human/Computer	Human	Computer
(6) Blended decision making	Human/Computer	Human/Computer	Human/Computer	Computer
(7) Rigid system	Human/Computer	Computer	Human	Computer
(8) Automated decision making	Human/Computer	Human/Computer	Computer	Computer
(9) Supervisory control	Human/Computer	Computer	Computer	Computer
(10) Full automation	Human	Computer	Computer	Computer

[89]Endsley, M. R. «Level of automation effects on performance, situation awareness and workload in a dynamic control task». In: *Ergonomics* 42.3 (1999), pp. 462–492

Un passo decisivo è stato fatto da Parasuraman, Sheridan, e Wickens[90], che, sulla base della scala elaborata da Sheridan e Verplank, hanno introdotto l'idea di associare i livelli di automazione ai tipi di funzione, come riportato nella figura 2.2. Le funzioni sono modellate sulla base di quattro fasi distinte, necessarie affinché un essere umano sia in grado di processare informazioni e agire: (1) acquisizione dei dati, (2) analisi delle informazioni, (3) decisione e scelta dell'azione, e infine (4) attuazione. Le quattro fasi possono essere tradotte in funzioni equivalenti del sistema e ogni funzione può essere automatizzata secondo livelli differenti di automazione.

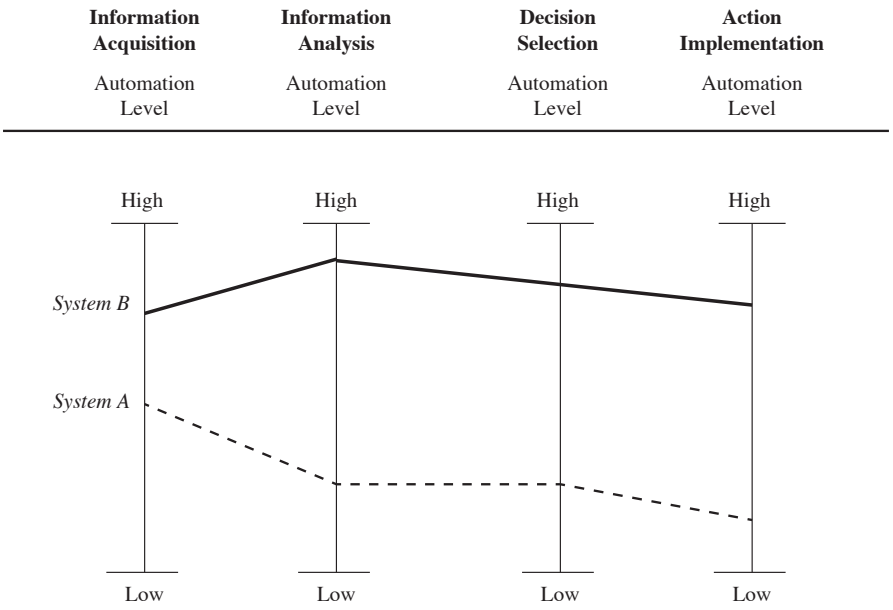


FIGURA 2.2: A model for Types and Levels of Automation proposed by Parasuraman, Sheridan and Wickens (2000)

Più di recente, e all'interno di un progetto SESAR (Single European Sky ATM

[90]Parasuraman, R., Sheridan, T. B. e Wickens, C. D. «A model for types and levels of human interaction with automation». In: *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 30.3 (2000), pp. 286–297.

Research)[91] nel settore aeronautico, chiamato "Good Practices for HP Automation Support", e sulla base dei modelli precedenti, Save e Feuerberg [92] hanno presentato una nuova tassonomia dei livelli di automazione. Per ogni funzione sono stati sviluppati differenti livelli di automazione. La tassonomia -riportata nella tabella 2.3- si presenta suddivisa in quattro colonne, corrispondenti alle quattro funzioni generiche elaborate da Parasuraman, Sheridan, e Wickens. Ogni funzione presenta un diverso numero di livelli di automazione: cinque per l'acquisizione dei dati e l'analisi delle informazioni, sei per la decisione e scelta dell'azione e infine otto per l'attuazione.

TABELLA 2.3: The Level of Automation Taxonomy (LOAT)-Save e Feuerberg (2012)

A	**B**	**C**	**D**
Information Acquisition	**Information Analysis**	**Decision and Action Selection**	**Action Implementation**
A0 Manual Information Acquisition	**B0 Working Memory Based Information Analysis**	**C0 Human Decision Making**	**D0 Manual Action and Control**
The human acquires relevant information on the process s/he is following without using any tool.	The human compares, combines, and analyses different information items regarding the status of the process s/he is following by way of mental elaborations. S/he does not use any tool or support external to her/his working memory.	The human generates decision options, selects the appropriate ones and decides all actions to be performed.	The human executes and controls all actions manually.
A1 Artefact-Supported information Acquisition	**B1 Artefact-Supported Information Analysis**	**C1 Artefact-Supported Decision Making**	**D1 Artefact-Supported Action Implementation**

[91] SESAR è un progetto a cui collaborano la Commissione Europea ed Eurocontrol, insieme ad oltre cento partner del settore. L'obiettivo è modernizzare l'ATM, il sistema di gestione del traffico aereo.

[92] Save, L., Feuerberg, B. e Avia, E. «Designing human-automation interaction: a new level of automation taxonomy». In: *2012), Proc. Human Factors of Systems and Technology* (2012)

The human acquires relevant information on the process s/he is following with the support of low-tech non-digital artefacts.	The human compares, combines, and analyses different information items regarding the status of the process s/he is following utilising paper or other non-digital artefacts.	The human generates decision options, selects the appropriate ones and decides all actions to be performed utilising paper or other non-digital artefacts.	The human executes and controls actions with the help of mechanical non-software based tools.

A2 Low-Level Automation Support of Information Acquisition	**B2 Low-Level Automation Support of Information Analysis**	**C2 Automated Decision Support**	**D2 Step-by-step Action Support:**
The system supports the human in acquiring information on the process s/he is following. Filtering and/or highlighting of the most relevant information are up to the human.	Based on user's request, the system helps the human in comparing, combining and analysing different information items regarding the status of the process being followed.	The system proposes one or more decision alternatives to the human, leaving freedom to the human to generate alternative options. The human can select one of the alternatives proposed by the system or her/his own one. The system assists the operator in performing actions by executing part of the action and/or by providing guidance for its execution. However, each action is executed based on human initiative and the human keeps full control of its execution.	

A3 Medium-Level Automation Support of Information Acquisition	**B3 Medium-Level Automation Support of Information Analysis**	**C3 Rigid Automated Decision Support**	**D3 Low-Level Support of Action Sequence Execution**
The system supports the human in acquiring information on the process s/he is following. It helps the human in integrating data coming from different sources and in filtering and/or highlighting the most relevant information items, based on user's settings.	Based on user's request, the system helps the human in comparing, combining and analysing different information items regarding the status of the process being followed. The system triggers visual and/or aural alerts if the analysis produces results requiring attention by the user.	The system proposes one or more decision alternatives to the human. The human can only select one of the alternatives or ask the system to generate new options.	The system performs automatically a sequence of actions after activation by the human. The human maintains full control of the sequence and can modify or interrupt the sequence during its execution.

A4 High-Level Automation Support of Information Acquisition	**B4 High-Level Automation Support of Information Analysis**	**C4 Low-Level Automatic Decision Making**	**D4 High-Level Support of Action Sequence Execution**
The system supports the human in acquiring information on the process s/he is following. The system integrates data coming from different sources and filters and/or highlights the information items which are considered relevant for the user. The criteria for integrating, filtering and highlighting the relevant information are predefined at design level but visible to the user.	The system helps the human in comparing, combining and analysing different information items regarding the status of the process being followed, based on parameters pre-defined by the user. The system triggers visual and/or aural alerts if the analysis produces results requiring attention by the user.	The system generates options and decides autonomously on the actions to be performed. The human is informed of its decision.	The system performs automatically a sequence of actions after activation by the human. The human can monitor all the sequence and can interrupt it during its execution.

A5 Full Automation Support of Information Acquisition	**B5 Full Automation Support of Information Analysis**	**C5 High-Level Automatic Decision Making**	**D5 Low-Level Automation of Action Sequence Execution**
The system supports the human in acquiring information on the process s/he is following. The system integrates data coming from different sources and filters and/or highlights the information items which are considered relevant for the user. The criteria for integrating, filtering and highlighting the relevant info are predefined at design level and not visible to the user (transparent to the user in Computer Science terms).	The system performs comparisons and analyses of data available on the status of the process being followed based on parameters defined at design level. The system triggers visual and/or aural alerts if the analysis produces results requiring attention by the user.	The system generates options and decides autonomously on the action to be performed. The human is informed of its decision only on request.	The system initiates and executes automatically a sequence of actions. The human can monitor all the sequence and can modify or interrupt it during its execution.
		C6 Full Automatic Decision Making	**D6 Medium-Level Automation of Action Sequence Execution**
		The system generates options and decides autonomously on the action to be performed without informing the human. (Note that this level is always connected to some kind of ACTION IMPLEMENTATION, at an automation level not lower than D5.)	The system initiates and executes automatically a sequence of actions. The human can monitor all the sequence and can interrupt it during its execution.

	D7 High-Level Automation of Action Sequence Execution
	1|The system initiates and executes a sequence of actions. The human can only monitor part of it and has limited opportunities to interrupt it.
	D8 Full Automation of Action Sequence Execution
	1|The system initiates and executes a sequence of actions. The human cannot monitor nor interrupt it until the sequence is not terminated.

Questa tassonomia sarà utilizzata nell'ultimo capitolo per l'analisi di uno scenario, definendo in che modo attività e compiti sono distribuiti tra uomo e macchina e analizzando le possibili responsabilità degli attori coinvolti.

2.8 Oltre l'errore umano: il rischio e la modellazione degli incidenti

Oggi, una delle sfide di carattere cruciale che si pone alle società contemporanee riguarda l'identificazione, la valutazione, la riduzione e il controllo dei rischi correlati alle moderne tecnologie. Si tratta, evidentemente, di una sfida non riducibile al mero livello tecnico e, d'altra parte, neppure alla produzione e all'attuazione di determinate normative. All'interno di organizzazioni e sistemi complessi è necessario svolgere un riesame del concetto di rischio e di gestione del rischio, secondo una visione socio-tecnica, nel cui ambito i processi di comunicazione tra i soggetti e l'interazione con le tecnologie coinvolte, rivestono un ruolo centrale. La crescita della complessità delle organizzazioni e delle moderne tecnologie ne ha reso più opaco il funzionamento, aumentando il rischio di errori e possibili incidenti. Per lungo tempo, incidenti e danni sono stati spiegati e ricondotti ad un fallimento della tecnologia o ad un errore umano. Tuttavia,

errori e danni raramente sono riconducibili ad una singola causa, umana o tecnologica, ma derivano da molteplici eventi, che entrando in relazione tra loro, causano il verificarsi di un evento avverso. Se è vero quindi che quest'ultimo è spesso attivato dall'errore di un operatore, sia egli un pilota, un medico o un tecnico di una centrale, è ugualmente vero che spesso l'errore si innesta in un sistema organizzativo caratterizzato da criticità latenti, attivate dall'errore umano. Il verificarsi di un danno è riconducibile non solo alla violazione di norme o protocolli ma anche, per esempio, per il rispetto di regole talvolta fallaci o non adatte alla complessità del compito. Una visione socio-tecnica del rischio, degli incidenti e quindi del danno, necessita di un'analisi che tenga conto dell'interazione tra i processi sociali, culturali, tecnologici, organizzativi e interorganizzativi che sono alla base dell'incidente. Uno dei primi a spingersi verso un approccio sistemico dell'errore fu James Reason, professore di psicologia a Manchester, attraverso lo studio della natura dell'errore umano nelle organizzazione complesse e i meccanismi cognitivi ad esso sottesi. Questo modello è utile ad individuare e diagnosticare gli errori all'interno dei sistemi socio-tecnici complessi e fondato sulla distinzione tra errori latenti ed errori attivi [93].

1. Errori latenti:

   Gli errori latenti o patogeni, possono convivere con un'organizzazione e rimanere nascosti all'interno del sistema, di per sé incapaci di manifestare una sintomatologia o un danno conclamato. Tuttavia, in connessione con altri fattori eziologici e in presenza di condizioni facilitanti, possono dare origine a un evento patologico. Tra gli errori patogeni, si annoverano: (a) gli errori legati alle tecnologie, come per esempio errori di progettazione, cattiva manutenzione, insufficiente addestramento all'uso; (b) gli errori gestionali, ad esempio la non corretta distribuzione dei carichi di lavoro o una pressione temporale eccessiva; (c) le carenze di *leadership*, ad esempio l'inadeguata motivazione del personale e la mancata chiarezza di compiti, obiettivi e responsabilità. Tutti questi fattori sono in grado di influire sul verificarsi di errori, violazioni e incidenti. Il riconoscimento e l'eliminazione di questi errori riduce la probabilità che si verifichi un danno.

2. Errori attivi.

   Gli errori attivi sono i più semplici da individuare, poiché sono i fattori che scatenano l'evento indesiderato. Si collocano a livello di persone e, quindi, il loro riscontro coincide spesso con l'identificazione di una responsabilità individuale. Tuttavia, fermando l'analisi a questo livello si rischia di

[93] Reason, *Managing the risks of organizational accidents*, cit.

> concludere che la rimozione o la punizione del responsabile possa evitare il ripetersi dell'evento. Al contrario, è possibile che la causa generatrice risieda in scelte organizzative o decisioni manageriali sbagliate. Se così fosse, è altamente probabile che individui diversi, o anche lo stesso individuo, a distanza di tempo e nelle medesime condizioni, ripropongano lo stesso tipo di errore.

Il verificarsi di un incidente è rappresentato da Reason attraverso il grafico del modello *Swiss Cheese*. Questo modello descrive le organizzazioni come una serie di fette di formaggio svizzero che scivolano l'una sull'altra, mediante moti continui. I buchi nelle fette di formaggio identificano le falle nelle barriere difensive che un'organizzazione può mettere in atto per impedire il verificarsi di eventi avversi, come per esempio strumenti tecnologici o procedure e protocolli. Lo spostamento delle fette può determinare l'allineamento casuale dei buchi. Se tutti i buchi si allineassero, la traiettoria delle opportunità dell'incidente potrebbe compiersi.

La letteratura scientifica affronta questo tema ormai da molti anni [94]. L'attenzione alla genesi e alla dinamica incidentale inizia a spostarsi su come gli aspetti organizzativi, per esempio i sistemi di coordinamento e controllo, la formazione degli operatori e i processi di comunicazione, l'adozione di protocolli e standard inadatti, siano determinanti per gli incidenti. Questi fattori organizzativi rimangono latenti, manifestando i loro effetti nel tempo e predisponendo l'organizzazione agli incidenti, con evidenti conseguenze sul concetto di responsabilità.

Come vedremo affrontando il tema della responsabilità, in sistemi altamente complessi, non solo definire la responsabilità individuale diviene piuttosto problematico ma, si rischia di attribuire la responsabilità non tanto ai responsabili dell'evento dannoso, quanto a coloro che ereditano i difetti presenti nel sistema, generati da attori e unità organizzative distanti nel tempo e nello spazio. L'errore e quindi l'evento dannoso è spesso organizzativamente costruito da strutture, pratiche e processi complessi.

2.9 Il sistema socio-tecnico sanità

Il sistema sanitrio nazionale è costituito dal complesso di funzioni, strutture, servizi e attività destinati alla promozione, al mantenimento e al recupero della

[94] Si vedano i contributi di Perrow, C. *Normal accidents: Living with high risk technologies*. Princeton University Press, 2011 in relazione al rischio e alle tecnologie Bucchi, M. «M. Catino," Da Chernobyl a Linate. Incidenti tecnologici o errori organizzativi?", 2002». In: *Rassegna Italiana di Sociologia* 45.2 (2004), pp. 299–300.

salute fisica e psichica della popolazione, senza distinzioni di condizioni individuali o sociali e secondo modalità che assicurino l'eguaglianza dei cittadini nei confronti dei servizi[95]rtItalia. «Legge 23 dicembre 1978, n. 833: Istituzione del Servizio Sanitario Nazionale». In: *Gazzetta Ufficiale* 360 (), art.1.. Compongono il sistema il Ministero della salute, che coordina il piano sanitario nazionale, ferme le competenze costituzionalmente garantite delle Regioni; una serie di enti e organi di livello nazionale, quali per esempio il Consiglio Superiore di Sanità (CSS) e l'Istituto Superiore di Sanità (ISS), l'Agenzia Nazionale per i Servizi Sanitari Regionali (Age.na.s.) e l'Agenzia Italiana del Farmaco (AIFA). Troviamo poi i servizi sanitari regionali, che a loro volta comprendono: (a) le regioni e le province autonome di Trento e Bolzano, (b) le aziende sanitarie locali (ASL) e le aziende ospedaliere (AO), attraverso cui le regioni e le province autonome assicurano l'assistenza sanitaria. A loro volta le aziende sanitarie locali comprendeno al proprio interno numerosi sottoinsiemi, come per esempio servizi di emergenza pre-ospedaliera, ospedali- che a loro volta si dividono in dipartimenti, reparti, servizi, divisioni, gruppi e comitati- poliambulatori, farmacie, laboratori, ditte specializzate, agenzie governative e associazioni di pazienti. Ognuno di questi sottoinsiemi presenta un proprio organigramma ed è caratterizzato da obiettivi, aspetti culturali e norme di comportamento, risorse finanziarie, tecniche e umane differenti.

Il presupposto da cui partire è che le aziende sanitarie sono sistemi complessi, che generano a loro volta, problemi complessi che richiedono necessariamente risposte complesse. Prevenire gli errori significa ridisegnare il sistema e i processi di lavoro, per renderli più sicuri. La letteratura è unanime nell'indicare l'esigenza di un approccio sistemico all'errore e al rischio clinico. Come affermato dall'Institute of Medicine (IOM), nel celebre documento *To err is human*, "la natura decentralizzata e frammentata del sistema di assistenza sanitaria contribuisce alle condizioni di insicurezza del paziente e rappresenta un ostacolo agli sforzi per migliorare la sicurezza. E' necessario un approccio olistico e sistemico che non può focalizzarsi su una singola soluzione"[96]. L'organizzazione è quindi un sistema costituito da componenti di natura diversa, umane, tecnologiche e procedurali, che interagiscono tra loro. Occorre superare una concezione dell'errore umano che vede l'attore potenzialmente libero di agire e di violare o meno le regole del sistema, e diventa necessario guardare all'interazione tra l'attore e il sistema, per capire le ragioni che lo hanno indotto a commettere l'errore.

Il comportamento umano e la ricerca del colpevole smettono di avere un ruolo centrale, per fare spazio alle condizioni in cui si verifica l'errore e quindi

[95] a

[96] Kohn, L. T., Corrigan, J. M., Donaldson, M. S. et al. *To err is human: building a safer health system. A report of the Committee on Quality of Health Care in America, Institute of Medicine*. 2000.

alla ricerca delle cause di fallimento del sistema. Secondo Reason non potendo cambiare gli esseri umani è necessario intervenire cambiando le condizioni in cui essi lavorano[97]. L'errore diviene fonte di apprendimento, per evitare il ripetersi delle circostanze che lo hanno prodotto. Promuovere quindi la cultura dell'imparare dall'errore e non nasconderlo diventa la strategia vincente come dimostrano esperienze già maturate in altri contesti.

All'interno dell'organizzazione sanitaria, la maggior parte dei problemi non esiste in modo isolato ma è correlato a tutti gli aspetti del sistema. Per risolvere eventuali problemi specifici, è necessario considerarli in una prospettiva più ampia e come parte di un insieme coerente. Il pensiero sistemico, e socio-tecnico in particolare, orientato al riconoscimento di schemi e interrelazioni tra esseri umani, tecnologia e ambiente, è applicato ai problemi industriali e gestionali da molto tempo; tuttavia solo recentemente è stato applicato ai problemi organizzativi in campo sanitario.

Le strutture sanitarie sono sistemi adattativi complessi [98], che si specificano in assetti organizzativi molteplici e interconnessi e che tendono a ridefinire costantemente il rapporto tra sistema e ambiente (centro/periferia, sistema dell'assistenza ospedaliera e delle cure primarie, sistema dei professionisti clinici e dei professionisti convenzionati, strutture accreditate e apparati regionali), dovendo assorbire una crescente domanda di forme assistenziali innovative e ad elevato contenuto tecnologico. In altri termini, si tratta di sistemi che devono rispondere alla complessità con un più alto livello di organizzazione. Le strutture sanitarie pubbliche sono insiemi molto complessi da governare il cui ultimo fine è la salute e, come tutte le strutture produttive complesse, devono essere in grado di garantire un livello elevato di efficienza. Così, un'azienda sociosanitaria può essere definita in molti modi, a seconda dei punti di vista e delle scuole di pensiero. In una visione socio-tecnica, essa è un'organizzazione inserita in un ambiente di riferimento da cui acquisisce risorse (input) e su cui opera restituendo prestazioni (output)[99]. Un organizzazione in cui esseri umani, tecnologie e infrastrutture sono in posizione di interdipendenza e interagiscono per la realizzazione di un fine comune. Un sistema aperto, caratterizzato da componenti miste sia di tipo tecnologico che umano-comportamentale, normativo e procedurale, che realizza scambi di risorse, persone e informazioni in una continua interazione con l'ambiente esterno.

Shortell e Kaluzny individuano alcune caratteristiche distintive del sistema Sanità, rispetto ad altre organizzazioni, tra cui il carattere altamente variabile e

[97] Reason, J. *Human error*. Cambridge university press, 1990.

[98] Wright, J., Hill, P. e Favaretti, C. *La governance clinica*. McGraw-Hill, 2005.

[99] Serpelloni, G. e Simeoni, E. «Principi sull'organizzione dell'azienda socio-sanitaria pubblica». In: *Quality management* (1995).

complesso, e la natura urgente e non dilazionabile di una parte consistente del lavoro; l'alta interdipendenza delle attività lavorative, che richiedono un alto grado di coordinamento tra i diversi gruppi professionali; e una ridotta tolleranza all'errore e all'ambiguità [100]. Inoltre, negli ultimi anni il sistema sanità è stato investito da una serie di dinamiche, riconducibili a fattori ambientali esterni, che hanno inciso in modo significativo sulle caratteristiche generali e organizzative dell'intero sistema, tra cui (a) cambiamenti istituzionali, (b) modifiche del quadro normativo, (c) dinamiche socio-economiche e infine (d) un dirompente processo di innovazione tecnologica.

Si tratta quindi di un sistema aperto, caratterizzato da un'organizzazione dinamica, eteroreferenziale, perchè orientato al raggiungimento di un fine ultimo, la salute, altamente flessibile e capace di rispondere bene in contesti altamente instabili.

Abbiamo avuto modo di vedere come la teoria socio-tecnica abbia portato al superamento di regole univoche di corretta organizzazione, per arrivare a una visione di coerenza tra le variabili e gli elementi organizzativi e tra gli stessi con la situazione ambientale, tanto esterna quanto interna al sistema. Le variabili a cui dobbiamo riferirci hanno natura profondamente diversa [101]. Possiamo distinguere due ordini principali di variabili capaci di influire sul sistema sanità: (1) di contesto e (2) ambientali, esterne al sistema organizzativo e legate ai fenomeni sociali, economici e culturali.

1. Variabili di contesto.

   Sono interne al sistema organizzativo e comprendono:

   (a) Variabili individuali.

   Riguardano le caratteristiche di coloro che operano all'interno del sistema e giocano un ruolo fondamentale nei processi di cambiamento organizzativo. Le variabili individuali comprendo al loro interno sistemi di valori e processi di comportamento. Ne fanno parte: le caratteristiche professionali e le abilità degli operatori; i bisogni e i desideri;la percezione individuale dei fenomeni; gli aspetti motivazionali; e infine gli aspetti decisionali.

   (b) Variabili sociali.

[100] Shortell, S. M. e Kaluzny, A. D. *Health care management: a text in organization theory and behavior*. Albany, New York: Delmar Thomson Learning, 1988., 2013.

[101] Per approfondimenti sui diversi tipi di variabili organizzative si veda per esempio Rebora, G. «Organizzazione aziendale». In: *Teorie e strumenti per l'analisi e la progettazione, Carocci Editore* (1998).

Sono costituite tipicamente dall'insieme delle relazioni interpersonali che si manifestano all'interno di un sistema e rilevano per il suo comportamento. Esse riguardano il significato delle relazioni sociali interne ed esterne al sistema. Ne fanno parte: le relazioni tra i membri del sistema; l'organizzazione in gruppi di lavoro; i processi decisionali e operativi dei gruppi; il grado di collaborazione o conflitto all'interno del gruppo e tra gruppi diversi.

(c) Variabili tecniche.

Scaturiscono dalle modalità operative e applicative di cui si avvale il lavoro umano. Ne fanno parte: macchine, impianti, attrezzature e tecnologie; i modi in cui il lavoro si svolge; le tecniche applicate; la complessità e la rigidità delle tecniche applicate.

(d) Variabili organizzative.

Realizzano le relazioni di connessione fra gli elementi del sistema, definendone specifici attributi, funzionali ai risultati organizzativi. Ne fanno parte: la struttura organizzativa. Essa identifica l'articolazione formale delle responsabilità organizzative ed economiche; i meccanismi e i sistemi operativi e di controllo. Si tratta delle regole formali di funzionamento dell'organizzazione, dei sistemi informativi, di comunicazione, e di valutazione e dei meccanismi di supervisione diretta, adattamento reciproco, standardizzazione dei processi e delle conoscenze, piani e programmi e ruoli di coordinamento;

(e) Variabili istituzionali.

Sono variabili che condizionano il campo di variabilità del sistema considerato. Ne fanno parte: le finalità istituzionali; l'assetto giuridico; la combinazione dei processi produttivi; e le strategie di sviluppo dell'azienda.

2. Variabili ambientali.

Sono esterne al sistema organizzativo e relative ad aspetti socio-economici, giuridici e culturali dell'ambiente in cui il sistema e la singola organizzazione operano. Ne fanno parte: (a) la legislazione, (b) le competenze tecniche e professionali disponibili, (c) i livelli retributivi correnti sul mercato, e infine (d) le innovazioni tecnologiche

2.10 Il rischio clinico

La strada per migliorare la qualità e la sicurezza delle prestazioni sanitarie si interseca con quella della consapevolezza e prevenzione degli errori e di eventuali

danni alla salute. Le organizzazioni sanitarie, come da tempo fanno le industrie, sono chiamate ad analizzare gli eventi avversi utilizzando tecniche di indagine rigorose per rimuovere gli errori di sistema, alla base di tali eventi.

I primi studi sul rischio clinico sono partiti dall'esame di eventi conseguenti a trattamento medico da cui è derivata disabilità o prolungamento del ricovero ospedaliero[102], ma è solo con la pubblicazione del rapporto *To err is human*[103] che il tema dell'errore umano in medicina diviene centrale per studiosi, professionisti e istituzioni, alimentando un'area di ricerca finalizzata ad analizzare il rapporto tra tecnologie e gestione del rischio [104].

In Italia, il Ministero della Salute ha recentemente emanato un decreto Ministeriale, intitolato *Risk Management in Sanità. Il problema degli errori* [105], in cui si analizza in modo approfondito il tema del rischio clinico, fornendo al tempo stesso una raccolta di informazioni e raccomandazioni, utili a tutti gli operatori del settore sanitario. A livello di singole aziende sanitarie sono state costituite Unità di Gestione del Rischio (URG), gruppi interdisciplinari e multiprofessionali preposti al coordinamento delle attività di identificazione del rischio clinico ed all'analisi e alla programmazione di interventi migliorativi. Inoltre, come previsto dalla *Carta della sicurezza nell'esercizio della pratica medica e assistenziale*, va ribadito il diritto alla sicurezza inteso come "diritto a entrare in relazione con un professionista o una struttura che garantisca al paziente modalità organizzative e comportamenti professionali, in grado di tenere sotto controllo i rischi e di ridurre al minimo il verificarsi di errori, nel corso dei trattamenti medici e assistenziali[106].

Come abbiamo osservato all'inizio di questo capitolo *robot* e sistemi d'intelligenza artificiale sono strumenti ormai necessari e indispensabili allo svolgimento dell'attività medica.

Il problema del danno al paziente è un rischio generalizzato, che coinvolge i rapporti tra tutte le componenti del sistema sanità. Il problema del rischio clinico richiede un profondo cambiamento culturale di tutti gli attori coinvolti nel processo, attraverso l'adozione di una prospettiva che coinvolga, non solo la gestione degli eventi sfavorevoli, ma anche la gestione del rischio stesso. È

[102] Association, C. M., Association, C. H. et al. *Report on the medical insurance feasibility study*. California Medical Association, 1977

[103] Kohn, Corrigan e Donaldson, *To err is human: building a safer health system. A report of the Committee on Quality of Health Care in America, Institute of Medicine*, cit.

[104] Esteves, J. e Joseph, R. C. «A comprehensive framework for the assessment of eGovernment projects». In: *Government information quarterly* 25.1 (2008), pp. 118–132.

[105] Salute, M. della. «Risk management in Sanità». In: *Il problema degli errori. Allegato* 4 (2004).

[106] Salute, M. della. *Sicurezza dei pazienti e gestione del rischio clinico: Manuale per la formazione degli operatori sanitari*. 2007.

necessario quindi abbandonare una visione dell'errore sanitario come colpa individuale meritevole di punizione, per arrivare a un concetto di falla all'interno di un sistema organizzativo complesso, da studiare, rimediare e prevenire.

Così, analizzando la genesi di un danno e prendendo in considerazione tutti gli eventi, gli errori e i deficit che lo hanno generato, ci si accorge che solo in una percentuale di casi minore esso è dovuto ad un errore isolato, umano o tecnologico. Più spesso il danno, è il frutto di una concatenazione di errori ed eventi, capace di superare le difese dell'organizzazione, mentre l'errore umano o tecnologico non sono che l'ultimo anello causale della catena.

Da queste considerazioni e sulla base del principio ippocratico del *primum non nocere*, nasce il modello di Reason secondo cui gli errori sono conseguenze dell'agire organizzato e non causa del fallimento del sistema[107]. Il modello elaborato da Reason e fondato sulla distinzione tra errori attivi ed errori latenti, rimane molto attuale e utile ad individuare e diagnosticare gli errori in ambito sanitario. Come abbiamo avuto modo di vedere, gli errori latenti possono convivere con un'organizzazione senza emergere, anche per lunghi periodi, incapaci di per sé di provocare un danno conclamato, ma capaci di dare origine a un evento patologico in concomitanza con altri fattori eziologici, o in condizioni facilitanti. Il loro riconoscimento riduce la possibilità di errore e del verificarsi di un danno.

Charles Vincent indica in maniera ancor più dettagliata dal micro-livello al macro-livello i fattori latenti che determinano le condizioni di insicurezza del sistema, in ambito clinico[108]. In particolare possiamo distinguere errori:

- Relativi al paziente: complessità, gravità delle condizioni, linguaggio e comunicazione, personalità e fattori sociali;
- Individuali e dello staff: disegno e chiarezza del compito, disponibilità / uso protocolli, aiuto nel prendere decisioni;
- Relativi al team: comunicazione verbale e scritta, supervisione, ricerca d'aiuto e leadership nel team;
- Relativi all'ambiente di lavoro: competenze dello staff, carichi di lavoro, organizzazione dei turni, design, disponibilità e manutenzione delle tecnologie, supporto amministrativo e ambiente fisico;
- Relativi all'organizzazione e al management: risorse e vincoli finanziari, struttura organizzativa, politiche, standard, obiettivi, cultura sicurezza e priorità;

[107] Reason, *Managing the risks of organizational accidents*, cit.
[108] Vincent, C. *Patient safety*. John Wiley & Sons, 2011

- Relativi al contesto istituzionale: contesto economico e normativo, potere esecutivo a livello di SSN, collegamento con organizzazioni esterne.

Gli errori attivi costituiscono il fattore che ha scatenato l'evento indesiderato. Si collocano a livello delle persone, per cui il loro riconoscimento coincide in genere con l'identificazione di una responsabilità individuale.

Il rischio clinico può essere arginato attraverso iniziative di *risk management*, messe in atto a livello nazionale, regionale, aziendale e di singola struttura sanitaria. Una gestione del rischio efficace avviene attraverso (a) una prima fase di conoscenza e analisi dell'errore, per esempio attraverso sistemi di report; (b) l'individuazione e la correzione delle cause dell'errore, mediante strumenti di analisi come la Root Causes Analysis[109], riconosciuta come uno degli strumenti di analisi reattiva più efficaci e adattabili anche al contesto sanitario e considerata dalla Joint Commission on Accreditation of Healthcare Organization (JCAHO) lo strumento elettivo per l'analisi dei cd. eventi sentinella; (c) il monitoraggio delle misure messe in atto per la prevenzione dell'errore; e infine (d) l'implementazione e il sostegno attivo delle soluzioni proposte[110].

2.10.1 Modelli di analisi

Possiamo distinguere due tipi di analisi del rischio clinico: l'analisi proattiva e l'analisi reattiva. Ognuna di esse prevede uno specifico sistema di ricerca, raccolta ed esame dei dati. Nell'ambito della gestione del rischio, all'interno di una struttura sanitaria possono essere utilizzati entrambi gli approcci[111].

1. L'analisi proattiva.

   Essa mira all'individuazione e alla correzione preventiva di eventuali falle e criticità all'interno di un sistema, rispetto al verificarsi di un errore. Per individuare i punti critici e progettare sistemi sicuri è necessario scomporre in fasi e analizzare l'intero processo. L'analisi di processo è una metodologia integrata di tipo sia qualitativo che quantitativo. Il processo viene scomposto in macro-attività a loro volta analizzate in base a tutti i singoli compiti che devono essere portati a termine affinché l'attività sia conclusa con successo. Per ogni singolo compito si presumono gli errori potenziali

[109] Booster, C. «ROOT CAUSE ANALYSIS». in: *Joint Commission Perspectives on Patient Safety* 3.5 (2003).

[110] Salute, «Risk management in Sanità», cit.

[111] idem, *Sicurezza dei pazienti e gestione del rischio clinico: Manuale per la formazione degli operatori sanitari*, cit. Per approfondimenti in tema di analisi e gestione del rischio clinico si veda il Decreto Ministeriale 5 Marzo 2003, elaborato dalla Commissione Tecnica sul Rischio Clinico, idem, «Risk management in Sanità», cit.

e si stimano le probabilità che si verifichino e la gravità del danno. Successivamente, si valuta il grado di accettabilità del rischio e si pianificano le attività di intervento. Maggiore è la complessità di analisi, maggiore sarà la complessità del sistema applicativo e il tempo e le risorse necessarie [112]. In particolare, l'attività di analisi e valutazione del rischio può essere distinta in quattro fasi principali:

- Fase 1: Analisi dei processi e delle attività.
  L'analisi avviene attraverso la descrizione sistematica dello svolgimento delle principali attività dei processi di cura.
- Fase 2: Identificazione delle situazioni pericolose e dei modi di errore possibili.
  Questa fase prevede l'analisi delle singole attività; l'identificazione delle situazioni pericolose, fonte di possibili errori; l'identificazione dei modi di errore associati a ciascuna situazione pericolosa evidenziata, sulla base di una classificazione standardizzata dei modi di errore.
- Fase 3: Stima della probabilità di occorrenza dell'errore e della gravità del danno.
  La stima della probabilità di accadimento del singolo errore associato a una specifica situazione pericolosa. La stima può essere quantitativa o qualitativa. Nel primo caso si ricorre ai dati statistici riportati in letteratura, mentre la stima qualitativa viene fatta attraverso le valutazioni del personale di reparto, generalmente utilizzando una scala standardizzata di giudizi. Allo stesso modo, la stima del danno può essere svolta su base quantitativa o qualitativa.
- Fase 4: Valutazione del grado di accettabilità del rischio.
  Le stime effettuate nella fase precedente vengono collocate all'interno di una matrice di rischio, per determinare il grado di priorità d'intervento, sui singoli modi di errore, su specifiche situazioni pericolose o su parti del processo.

Le varie fasi di analisi di processo vengono condotte con diversi strumenti e di ampia applicazione, prevalentemente al di fuori dell'ambiente sanitario. Tra questi merita di essere segnalato il *Failure Mode Effects and Critical Analysis* (FMECA). È una tecnica previsionale utilizzata da molti anni negli Stati Uniti, in campo missilistico e di strumentazione elettronica e in Italia

[112] ibid.; Reason, J. «Combating omission errors through task analysis and good reminders». In: *Quality and Safety in Health Care* 11.1 (2002), pp. 40–44.

dalla FIAT, IVECO e dal Comitato Elettronico Italiano. Prevede considerazioni preventive di possibili errori e guasti, che portano alla valutazione del progetto e delle alternative, alla previsione di prove e controlli, e infine all'esplicitazione di un riferimento con cui confrontare il prodotto reale. Questo metodo è stato adattato alla realtà sanitaria[113] ed è in uso in un progetto sperimentale nella regione Emilia-Romagna [114].

2. L'analisi reattiva.

   Essa consiste nella ricerca delle cause del danno partendo dall'errore attivo che lo ha generato, per poi individuare le cause profonde e organizzative che hanno permesso il verificarsi del danno. Questa analisi a posteriori si basa sulla segnalazione obbligatoria o volontaria degli errori attraverso sistemi di *incident reporting*. Si tratta di uno degli strumenti indicati più frequentemente come base per una corretta gestione dei rischi, sia in ambito sanitario[115], sia in altri settori come per esempio nell'areonautica[116]. Il sistema italiano è obbligatorio a norma di un decreto ministeriale del 2009 e di un accordo del 2008 tra le regioni e il governo centrale. Il sistema di segnalazione obbligatoria è incentrato su incidenti molto gravi. Il personale sanitario segnala gli incidenti, scegliendo la categoria adeguata da una lista di 16, a livello regionale.

   In generale, la raccolta strutturata delle segnalazioni permette di costruire una base di dati da analizzare, per predisporre strategie e azioni correttive per prevenire l'accadimento del danno in futuro. L'analisi avviene attraverso dei *report standard* con indicazione predefinita delle informazioni. Per ogni evento è necessario indicare:

   a Elementi anagrafici.

     Ne fanno parte i dati relativi all'unità operativa e all'operatore, i dati relativi al paziente, le circostanze dell'evento (luogo, data e ora in cui si è verificato), e il tipo di prestazione.

   b Elementi oggettivi descrittivi.

[113] DeRosier, J. et al. «Using health care failure mode and effect analysis™: the VA National Center for Patient Safety's prospective risk analysis system». In: *The Joint Commission Journal on Quality and Patient Safety* 28.5 (2002), pp. 248–267

[114] ASR, R. E. R. *FMEA-FMECA analisi dei modi di errore/guasto e dei loro effetti nelle organizzazioni sanitarie*. 2002.

[115] Salute, «Risk management in Sanità», cit.

[116] Aviation Safety Reporting System (ASRS). Federal Aviation Administration. http://www.asias.faa.gov.

Rientrano in questa categoria il tipo di evento, la descrizione dell'evento e le indicazione dei fattori che possono aver contribuito a che si verificasse.Tra questi ultimi si possono distinguere fattori legati al paziente, fattori legati al personale e fattori legati al sistema.

c Elementi di valutazione dell'evento.
È possibile distiguere gli eventi in: potenziali ed effettivi.

- Eventi potenziali.
  Sono eventi potenzialmente pericolosi o causativi di un danno
  - non occorsi, cosìddetti *near miss* o *close call*. Identificano situazioni in cui un errore stava per essere commesso, ma non si è verificato per motivi fortuiti, o per l'intervento di meccanismi di contenimento; oppure
  - occorsi ma intercettati, prima di poter sfociare in un danno. Sono i cosìddetti *no harm events*, situazioni in cui l'errore è accaduto, ma senza che si verificassero conseguenze negative per il paziente.
- Eventi avversi.
  I cosìddetti *adverse events* identificano una situazione da cui un paziente ha ricevuto un danno, a seguito di un intervento, o di un omesso intervento, sanitario.

Ad oggi non esistono criteri nazionali univoci per definire i livelli di gravità di un evento. In linea generale è sicuramente possibile classificare un evento come: (1) grave: evento che causa morte, danni o invalidità permanete al paziente; (2) medio: evento che comporta un'invalidità temporanea, un cospicuo aumento dei giorni di degenza; (3) lieve: evento che provoca al paziente solo disturbi temporanei e limitati.

Ogni sistema è libero di modellare la gravità del danno adottando criteri differenti. La figura 2.4 mostra, a titolo di esempio, la scheda di segnalazione spontanea degli eventi redatta dall'agenzia sanitaria e sociale della regione Emilia-Romagna.

L'efficacia dell'*incident reporting* è riconosciuta a livello internazionale e documentata in letteratura da molti anni[117]. Introdotto per la prima volta dalla NASA[118] per migliorare la sicurezza aerea, in qualità di sistema confidenziale, volontario e non punitivo di segnalazione di eventi

[117] Liang, B. A. e Storti, K. «Creating problems as part of the" solution": the JCAHO Sentinel Event Policy, legal issues, and patient safety.» In: *Journal of health law* 33.2 (1999), pp. 263–285; National Reporting and Learning System: National Patient Safety Agency. http://npsa.nhs.uk.

[118] Aviation Safety Reporting System (ASRS). Federal Aviation Administration. http://www.asias.faa.gov

da parte di piloti e controllori di volo, *l'incident reporting* è oggi adottato dai sistemi sanitari di moltissimi paesi, con l'obiettivo di migliorare la sicurezza delle prestazioni. Caratteristiche peculiari del sistema di *incident reporting* sono: la base obbligatoria o volontaria della segnalazione, il carattere confidenziale/anonimo di raccolta dei dati; l'assenza di sanzioni; la riservatezza dei dati raccolti che sono destinati al *risk manager* aziendale e non all'utente o alla cartella clinica; l'obbligo di rimuovere i riferimenti all'autore della segnalazione, prima dell'invio al data base e di distruggere la scheda di raccola dei dati[119].

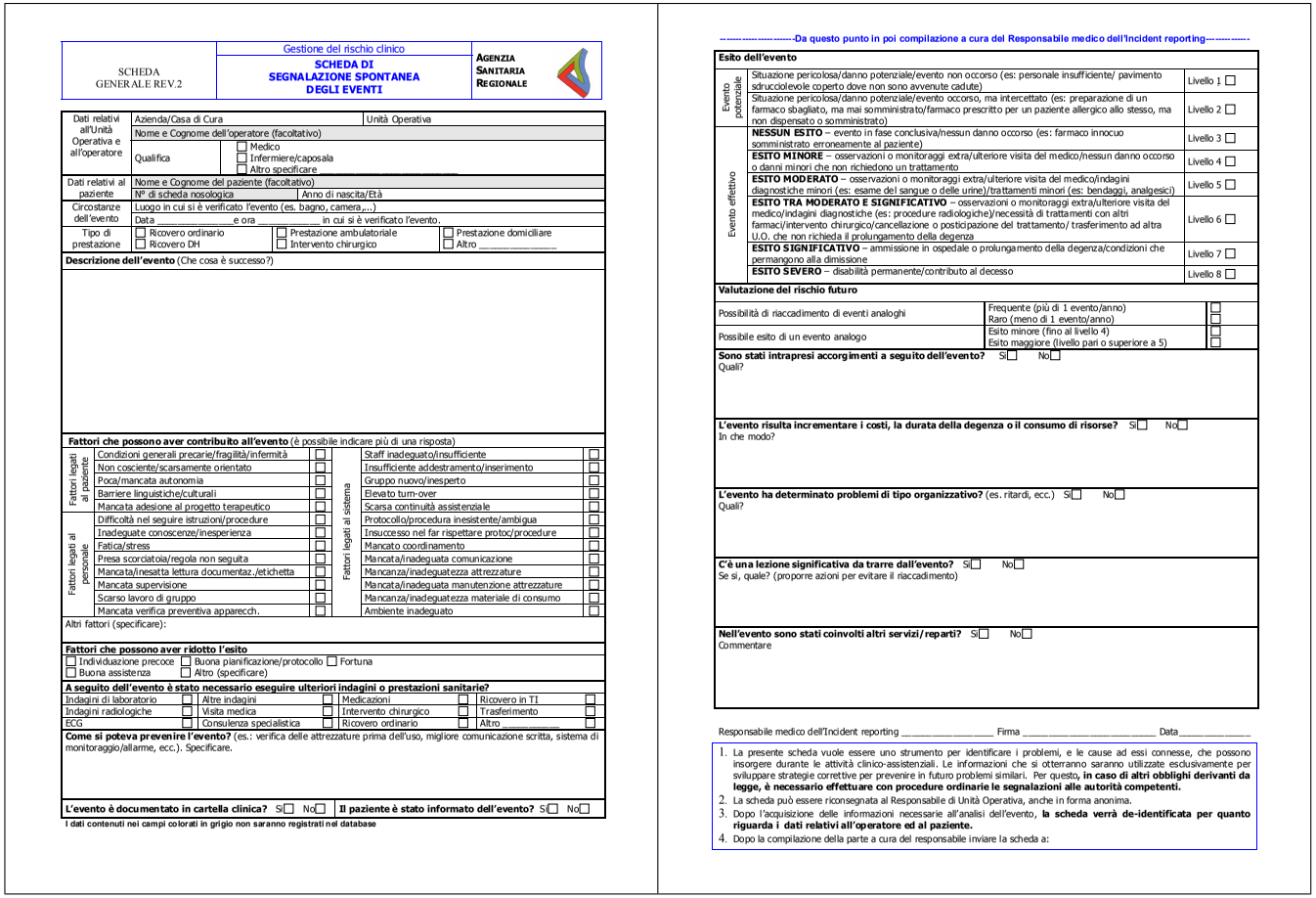
L'esempio più rilevante di applicazione del sistema in ambito sanitario è l'*Australian Incident Monitoring System* (AIMS), introdotto nel 1996 in Australia e in Nuova Zelanda[120]. L'AIMS nasce su iniziativa dell'Australian Patient Safety Foundation, organizzazione no profit attiva dal 1987 nella prevenzione dei danni in ambito sanitario e nel 2001 il sistema conteneva circa 50.000 segnalazioni [121]. Tra le esperienze più significative riportate in letteratura possiamo segnalare quella dell'Azienda Sanitaria Regione Emilia-Romagna che, nel 2003, ha sperimentato un programma di *incident reporting* ispirato al modello australiano, in cinque Aziende sanitarie territoriali e ospedaliere (Aziende AUSL di Modena e Reggio-Emilia, Aziende ospedaliere di Parma e Bologna, Istituti Ortopedici Rizzoli di Bologna), per 39 unità operative e raccogliendo 403 segnalazioni in tre mesi di sperimentazione[122]. Fanno da freno alla diffusione e al successo delle iniziative, la scarsa conoscenza della materia e la mancanza di una legislazione garante di depenalizzazione nei confronti degli attori.

[119] Cinotti, R., Basini, V. e Di Denia, P. «Il sistema di incident reporting nelle organizzazioni sanitarie». In: *Collana Dossier* 86 (2003).

[120] Australian Incident Monitoring System. Australian Patient Safety Foundation. http://www.apsf.net.au.

[121] Runciman, W. «Lessons from the Australian Patient Safety Foundation: setting up a national patient safety surveillance system—is this the right model?» In: *Quality and Safety in Health Care* 11.3 (2002), pp. 246–251.

[122] Vedi Fig.1. Cinotti, Basini e Di Denia, «Il sistema di incident reporting nelle organizzazioni sanitarie», cit.



SCHEDA GENERALE REV.2	Gestione del rischio clinico **SCHEDA DI SEGNALAZIONE SPONTANEA DEGLI EVENTI**	**AGENZIA SANITARIA REGIONALE**

Dati relativi all'Unità Operativa e all'operatore	Azienda/Casa di Cura	Unità Operativa
	Nome e Cognome dell'operatore (facoltativo)	
	Qualifica	☐ Medico ☐ Infermiere/caposala ☐ Altro specificare _____
Dati relativi al paziente	Nome e Cognome del paziente (facoltativo)	
	N° di scheda nosologica	Anno di nascita/Età
Circostanze dell'evento	Luogo in cui si è verificato l'evento (es. bagno, camera,...)	
	Data _____e ora _____ in cui si è verificato l'evento.	
Tipo di prestazione	☐ Ricovero ordinario ☐ Ricovero DH	☐ Prestazione ambulatoriale ☐ Intervento chirurgico
		☐ Prestazione domiciliare ☐ Altro _____

Descrizione dell'evento (Che cosa è successo?)

Fattori che possono aver contribuito all'evento (è possibile indicare più di una risposta)

Fattori legati al paziente	Condizioni generali precarie/fragilità/infermità	☐	Fattori legati al sistema	Staff inadeguato/insufficiente	☐
	Non cosciente/scarsamente orientato	☐		Insufficiente addestramento/inserimento	☐
	Poca/mancata autonomia	☐		Gruppo nuovo/inesperto	☐
	Barriere linguistiche/culturali	☐		Elevato turn-over	☐
	Mancata adesione al progetto terapeutico	☐		Scarsa continuità assistenziale	☐
Fattori legati al personale	Difficoltà nel seguire istruzioni/procedure	☐		Protocollo/procedura inesistente/ambigua	☐
	Inadeguate conoscenze/inesperienza	☐		Insuccesso nel far rispettare protoc/procedure	☐
	Fatica/stress	☐		Mancato coordinamento	☐
	Presa scorciatoia/regola non seguita	☐		Mancata/inadeguata comunicazione	☐
	Mancata/inesatta lettura documentaz./etichetta	☐		Mancanza/inadeguatezza attrezzature	☐
	Mancata supervisione	☐		Mancata/inadeguata manutenzione attrezzature	☐
	Scarso lavoro di gruppo	☐		Mancanza/inadeguatezza materiale di consumo	☐
	Mancata verifica preventiva apparecch.	☐		Ambiente inadeguato	☐

Altri fattori (specificare):

Fattori che possono aver ridotto l'esito

☐ Individuazione precoce ☐ Buona pianificazione/protocollo ☐ Fortuna
☐ Buona assistenza ☐ Altro (specificare)

A seguito dell'evento è stato necessario eseguire ulteriori indagini o prestazioni sanitarie?

Indagini di laboratorio	☐	Altre indagini	☐	Medicazioni	☐	Ricovero in TI	☐
Indagini radiologiche	☐	Visita medica	☐	Intervento chirurgico	☐	Trasferimento	☐
ECG	☐	Consulenza specialistica	☐	Ricovero ordinario	☐	Altro _____	☐

Come si poteva prevenire l'evento? (es.: verifica delle attrezzature prima dell'uso, migliore comunicazione scritta, sistema di monitoraggio/allarme, ecc.). Specificare.

L'evento è documentato in cartella clinica? Si☐ No☐ **Il paziente è stato informato dell'evento?** Si☐ No☐

I dati contenuti nei campi colorati in grigio non saranno registrati nel database

----------Da questo punto in poi compilazione a cura del Responsabile medico dell'Incident reporting----------

Esito dell'evento

Evento potenziale	Situazione pericolosa/danno potenziale/evento non occorso (es: personale insufficiente/ pavimento sdrucciolevole coperto dove non sono avvenute cadute)	Livello 1 ☐
	Situazione pericolosa/danno potenziale/evento occorso, ma intercettato (es: preparazione di un farmaco sbagliato, ma mai somministrato/farmaco prescritto per un paziente allergico allo stesso, ma non dispensato o somministrato)	Livello 2 ☐
Evento effettivo	**NESSUN ESITO** – evento in fase conclusiva/nessun danno occorso (es: farmaco innocuo somministrato erroneamente al paziente)	Livello 3 ☐
	ESITO MINORE – osservazioni o monitoraggi extra/ulteriore visita del medico/nessun danno occorso o danni minori che non richiedono un trattamento	Livello 4 ☐
	ESITO MODERATO – osservazioni o monitoraggi extra/ulteriore visita del medico/indagini diagnostiche minori (es: esame del sangue o delle urine)/trattamenti minori (es: bendaggi, analgesici)	Livello 5 ☐
	ESITO TRA MODERATO E SIGNIFICATIVO – osservazioni o monitoraggi extra/ulteriore visita del medico/indagini diagnostiche (es: procedure radiologiche)/necessità di trattamenti con altri farmaci/intervento chirurgico/cancellazione o posticipazione del trattamento/ trasferimento ad altra U.O. che non richieda il prolungamento della degenza	Livello 6 ☐
	ESITO SIGNIFICATIVO – ammissione in ospedale o prolungamento della degenza/condizioni che permangono alla dimissione	Livello 7 ☐
	ESITO SEVERO – disabilità permanente/contributo al decesso	Livello 8 ☐

Valutazione del rischio futuro

Possibilità di riaccadimento di eventi analoghi	Frequente (più di 1 evento/anno) Raro (meno di 1 evento/anno)	☐ ☐
Possibile esito di un evento analogo	Esito minore (fino al livello 4) Esito maggiore (livello pari o superiore a 5)	☐ ☐

Sono stati intrapresi accorgimenti a seguito dell'evento? Si☐ No☐
Quali?

L'evento risulta incrementare i costi, la durata della degenza o il consumo di risorse? Si☐ No☐
In che modo?

L'evento ha determinato problemi di tipo organizzativo? (es. ritardi, ecc.) Si☐ No☐
Quali?

C'è una lezione significativa da trarre dall'evento? Si☐ No☐
Se si, quale? (proporre azioni per evitare il riaccadimento)

Nell'evento sono stati coinvolti altri servizi/reparti? Si☐ No☐
Commentare

Responsabile medico dell'Incident reporting _____ Firma _____ Data_____

1. La presente scheda vuole essere uno strumento per identificare i problemi, e le cause ad essi connesse, che possono insorgere durante le attività clinico-assistenziali. Le informazioni che si otterranno saranno utilizzate esclusivamente per sviluppare strategie correttive per prevenire in futuro problemi similari. Per questo, **in caso di altri obblighi derivanti da legge, è necessario effettuare con procedure ordinarie le segnalazioni alle autorità competenti.**
2. La scheda può essere riconsegnata al Responsabile di Unità Operativa, anche in forma anonima.
3. Dopo l'acquisizione delle informazioni necessarie all'analisi dell'evento, **la scheda verrà de-identificata per quanto riguarda i dati relativi all'operatore ed al paziente.**
4. Dopo la compilazione della parte a cura del responsabile inviare la scheda a:

TABELLA 2.4: Scheda di *incident resporting*- Regione Emilia-Romagna

2.10.2 Alcune considerazioni sul sistema di gestione del rischio clinico

La riduzione dei danni derivanti dai processi di cura procede di pari passo al riconoscimento e alla prevenzione degli errori. Uno dei principali limiti del sistema di gestione del rischio clinico è quello di non riuscire a fornire analisi quantitative, in grado di garantire la rappresentatività del campione, variabile in base alla capacità degli operatori del settore di riconoscere gli eventi e alla volontà di renderli pubblici. Talvolta è molto difficile identificare anche le dimensioni del contesto cui riportare il numero di segnalazioni [123]. Un elemento critico, spesso sollevato in ambito sanitario e relativo al sistema di *Incident Reporting*, siano essi volontari o obbligatori, è l'incapacità di raggiungere a breve termine i propri obiettivi [124]. Si sottolinea come l'unico esempio di programma che abbia diminuito in modo incisivo il verificarsi di incidenti sia stato il National Nosocomial Infection Survey, che ha ridotto del 32% le infezioni ospedaliere nelle strutture degli Stati Uniti che lo hanno applicato per intero [125]. Tale incapacità nel raggiungimento degli obiettivi sembra essere riconducibile ai sistemi di *Reporting* utilizzati, raramente semplici da utilizzare, ambigui nella definizione degli eventi accaduti e di quelli da segnalare, spesso molto costosi in termini di tempo e denaro e a rischio di aumentare il contenzioso tra ente ospedaliero e paziente[126]. La capacità dei sistemi di *Reporting* di collezionare eventi avversi sembra, infatti, estremamente variabile, attestandosi intorno al 5-30% degli eventi che potrebbero essere segnalati. Inoltre una quota delle segnalazioni risulta inutilizzabile per incompletezza dei dati o difetti della scheda di raccolta [127].

La segnalazione degli eventi avversi, in qualsiasi campo e attività umana, e in particolare in quello sanitario, sia il risultato di un atto spontaneo o obbligatorio per legge societaria, aziendale, statale o religiosa, è realmente praticabile solo qualora l'organizzazione che la richiede dimostri di saperne cogliere la connotazione positiva. Un esempio positivo è certamente l'esperienza dell'*Aviation Safety Reporting System*, che colleziona più di 30.000 reports all'anno[128].

[123] ibid.

[124] Leape, L. L. «Reporting of adverse events». In: *N Engl J Med*. Citeseer. 2002.

[125] HALEY, R. W. et al. «The efficacy oe infection surveillance and control programs in preventing nosocomial infections in us hospitals». In: *American journal of epidemiology* 121.2 (1985), pp. 182–205.

[126] Rosenthal, J., Riley, T. e Booth, M. *State reporting of medical errors and adverse events: results of a 50-state survey*. National Academy for State Health Policy, 2000; Leape, «Reporting of adverse events», cit.

[127] ibid.

[128] Connell, L. «Statement before the subcommittee on oversight and investigations, Committee on Veterans' Affairs». In: *Washington, DC: US House of Representatives* (2000).

Infine, è certamente auspicabile un cambiamento culturale sul concetto di errore in sanità. Secondo la Società Italiana per la Qualità dell'assistenza Sanitaria (SIQuAS-VRQ), è necessario «riconoscere nell'errore un'opportunità di apprendimento e miglioramento, contrastando l'attuale prevalente atteggiamento punitivo, che è uno dei principali motivi del fallimento delle politiche e delle strategie per la sicurezza nei sistemi sanitari. L'atteggiamento punitivo ostacola la segnalazione degli eventi avversi e dei *near misses* impedendone di fatto la segnalazione "libera da rimprovero", in assenza di una politica coerente all'interno dell'organizzazione per la gestione confidenziale dei dati. Per questo motivo la legislazione vigente italiana va urgentemente aggiornata, sul modello di quanto fatto dai Governi australiano e danese che vietano di utilizzare i dati delle segnalazioni sugli eventi avversi ed i *near misses* a scopi giudiziari, a salvaguardia del principio generale del segreto professionale» [129].

[129] SIQuAS-VRQ. Raccomandazioni sulla gestione del rischio clinico per la sicurezza dei pazienti. Milano 15 maggio 2006, (Raccomandazione 8). http://www.siquas.it.

Responsabilità penale e sistemi di automazione

3.1 Verso una visione socio-tecnica della responsabilità

Quando all'interno di un sistema socio-tecnico vengono introdotte tecnologie ad elevata automazione, inevitabilmente, la maggiore complessità del sistema porta con se un aumento della complessità dei compiti[1]. Da un lato, l'innovazione tecnologica è in grado di aumentare le prestazioni complessive del sistema, coadiuvando e migliorando il lavoro degli esseri umani e talvolta permettendo di intervenire laddove prima non era possibile - si pensi ai vantaggi apportati dai sistemi di chirurgia robotica mini-invasiva, che rendono oggi possibile una chirurgia prima difficile o impossibile - dall'altro raramente riducono la complessità dei compiti e delle funzioni svolte dagli esseri umani.

La crescente complessità dei compiti e delle funzioni di controllo, si riverbera inevitabilmente sulla responsabilità, e in particolare sulla difficoltà di collegare in modo efficace la responsabilità ai compiti e ai ruoli all'interno dell'organizzazione.

In un contesto di questo tipo, può essere utile definire in modo preciso l'allocazione dei compiti, così da garantire una corrispondenza adeguata tra ruoli e responsabilità. Tuttavia, una mappatura univoca di compiti e responsabilità può non essere sufficiente a garantire la sicurezza e l'efficienza all'interno del sistema. Essa può talvolta rivelarsi inadeguata, specie all'interno di sistemi che non possono essere del tutto specificati e in cui è necessario stabilire compiti relativamente ampi e sovrapposti, talvolta ridondanti, e la cui realizzazione richiede una cooperazione flessibile tra gli agenti coinvolti[2]. All'interno di sistemi socio-tecnici molto complessi, come il sistema sanità, l'interazione uomo-macchina è

[1] Woods, D. D. *Behind human error*. Ashgate Publishing, Ltd., 2010; Perrow, *Normal accidents: Living with high risk technologies*, cit.

[2] Hollnagel, E. «The Human in Control: Modelling What Goes Right Versus Modelling What Goes Wrong». In: *Human Modelling in Assisted Transportation*. Springer, 2011, pp. 3–7.

spesso ricorsiva, dando vita ad un Sistema Cognitivo Integrato[3]. La nozione di Sistema Cognitivo Integrato ha origine nelle teorie del controllo di supervisione[4]. Secondo tali teorie, l'operatore umano rimane al vertice del sistema e può sostituirsi ad esso, quando una particolare situazione lo richieda. Si pensi ai protocolli d'intervento in telechirurgia, che richiedono la presenza di un secondo chirurgo in sala operatoria che possa sostituirsi al robot in caso di necessità; o nei casi di chirurgia robotica, alla presenza di un équipe formata all'assistenza dell'intervento robotico, sempre presente in sala operatoria, o ancora all'uso di sistemi esperti per la diagnostica.

A ciò si aggiunga che i sistemi presi in considerazione in questo lavoro, sono intimamente connessi al cd. rischio da ignoto tecnologico. Tale espressione identifica un contesto di incertezza scientifica, in cui "le modalità del calcolo del rischio, come sono state sinora definite dalla scienza e dalle istituzioni legali, collassano."[5]. Secondo una parte della dottrina, negli odierni sistemi socio-tecnici, il rischio di incidente diventa non solo ineliminabile, ma addirittura "normale"[6].

Tali caratteristiche del sistema possono avere un notevole impatto sulla ripartizione delle responsabilità. In tema di responsabilità penale legata all'uso di sistemi di automazione e d'intelligenza artificiale, all'interno del sistema sanità, sono state evidenziate alcune aree di particolare interesse, che saranno oggetto di trattazione nelle sezioni successive. In un ambiente così caratterizzato, è molto importante avere presente che i sistemi ad elevata automazione, e in particolare i sistemi d'intelligenza artificiale, possono essere considerati come strumenti nelle mani degli operatori umani e, al tempo stesso, come agenti autonomi, che operano all'interno del sistema, contribuendo attivamente allo svolgimento di compiti e talvolta sostituendosi, in tutto o in parte, all'operatore umano.

In particolare saranno analizzate: (a) la responsabilità penale da prodotto difettoso, con riferimento alla responsabilità per danno da dispositivi medici difettosi; (b) la responsabilità medica, connessa all'uso di tecnologie ad alto rischio, nelle ipotesi di danni derivanti dall'uso di dispositivi difettosi; e (c) la responsabilità penale legata all'uso di sistemi d'intelligenza artificiale, caratterizzati da un elevato grado di autonomia e i modelli elaborati dalla dottrina per regolare tale fenomeno. Analizzando le tecnologie in uso nel settore sanitario, è emersa

[3] In questo senso, Hollnagel, E. *Cognitive reliability and error analysis method (CREAM)*. Elsevier, 1998.

[4] Sheridan, T. B. «Human supervisory control of robot systems». In: *Robotics and Automation. Proceedings. 1986 IEEE International Conference on*. Vol. 3. IEEE. 1986, pp. 808–812.

[5] Beck, «La società del rischio, trad. it», cit., 29.

[6] Perrow, C. *Normal Accidents: Living with High Risk Technologies (Updated)*. Princeton University Press, 1999. Nello stesso senso Centonze, *La normalità dei disastri tecnologici: il problema del congedo dal diritto penale*, cit.

la difficoltà di ricomprenderle tutte all'interno di un'unica categoria concettuale, capace di definirne le caratteristiche in modo univoco. Esse hanno proprietà estremamente eterogenee e differenti tra loro, pur essendo accomunate da un elevato grado di automazione. Abbiamo evidenziato come alcune si caratterizzino principalmente come strumenti di ausilio per l'operatore umano, mentre altre, pur mantenendo tale caratteristica, si atteggino come agenti che operano autonomamente all'interno del sistema. Tali proprietà non possono che riflettersi sia sui livelli di automazione dei singoli sistemi, sia sui profili di responsabilità che derivano dal loro impiego.

Prima di procedere con l'analisi della responsabilità da prodotto difettoso, è indispensabile dedicare una parte della trattazione alla qualificazione giuridica dei sistemi di automazione nell'e-health e alla disciplina applicabile, europea e nazionale. Come vedremo, essa, da un lato, è strettamente legata alla responsabilità da prodotto difettoso, e al tempo stesso è in grado di incidere sui profili della responsabilità del medico, per danni al paziente nelle ipotesi di malfunzionamento del sistema.

3.2 La qualificazione giuridica dei sistemi di automazione nell'e-Health

In questa sezione, si cercherà di delineare una prima parte dell'ambito normativo in cui si collocano i sistemi di automazione utilizzati in campo medico, e le conseguenze di tale assetto regolatore sul regime della responsabilità.

In primo luogo è necessario chiarire se i sistemi robotici e d'intelligenza artificiale utilizzati nell'area di riferimento, rientrino o meno nella definizione di dispositivo medico (DM) e debba quindi ritenersi applicabile la relativa disciplina giuridica.

Al fine di chiarire cosa siano i dispositivi medici[7], è opportuno richiamare le definizioni comprese nelle due principali direttive che regolano il settore: la direttiva 90/385/CEE sui dispositivi medici impiantabili attivi[8], e la direttiva 93/42/CEE sui dispositivi medici in genere[9], rispettivamente recepite nel nostro ordinamento dal d. lgs. 14 dicembre 1992, n. 507, come da ultimo modificato

[7] In questa sede, verranno presi in considerazione i "dispositivi medici" diversi dai dispositivi medico-diagnostici in vitro regolati dalla direttiva 98/79/CE, trasposta nell'ordinamento nazionale italiano con D. Lgs. 8 settembre 2000, n. 332 Lgs, D. «settembre 2000, n. 332». In: *Attuazione della direttiva* 98 (8), p. 79.

[8] ex art.1 DIRECTIVA. «90/385/CEE e successiva modifica 2007/47/CE». in: *Dispositivo medico classe II a* (1990)

[9] ex art.1 DIRECTIVA. «93/42/CEE e successiva modifica 2007/47/CE». in: *Dispositivo medico classe II a* (2005).

dal d. lgs. 25 gennaio 2010, n. 37, e dal decreto legislativo 24 febbraio 1997, n. 46 e successive modifiche. Esse, seppure con alcune differenze, definiscono i dispositivi medici come una categoria di prodotti, strumenti, apparecchi, impianti, sostanze, software o altro, destinati ad essere impiegati nell'uomo, o sull'uomo, a scopo di diagnosi, prevenzione, controllo o terapia, attenuazione o compensazione di ferite o handicap, ma anche di studio, sostituzione o modifica dell'anatomia o di un processo fisiologico, o di controllo del concepimento[10]. Un prodotto può essere considerato dispositivo medico se svolge una delle funzioni previste nella definizione, attraverso una modalità d'azione che non sia farmacologica, immunologica o metabolica, pur potendo essere coadiuvato - nello svolgimento della sua funzione - da una o più di tali modalità.

La destinazione d'uso del prodotto deve essere in ogni caso caratterizzata come finalità medica. Le precisazioni sulla cosiddetta "medical purpose" sono contenute nella linea guida comunitaria MEDDEV 2.1/1, ex art. 1.1 lett. b), in cui si legge"*The medical purpose is assigned to a product by the manufacturer. The manufacturer determines through the label, the instruction for use and the promotional material related to a given device its specific medical purpose. As the directive aims essentially at the protection of patients and users, the medical purpose relates in general to finished products regardless of whether they are intended to be used alone or in combination...*"[11].

Sulla base di tali definizioni, è possibile fare alcune prime considerazioni. Il software informatico, rientra nella definizione di dispositivo medico, e più precisamente di dispositivo medico software (DMS), e dovranno applicarsi tutte le disposizioni europee e nazionali previste in materia. Una prima riflessione riguarda la destinazione d'uso del prodotto che deve necessariamente caratterizzarsi come finalità medica. Secondo la direttiva MEDDEV appena citata, la

[10] In particolare, l'articolo 2 della Direttiva 93/42/CE definisce dispositivo medico "qualunque strumento, apparecchio, impianto, software, sostanza o altro prodotto, utilizzato da solo o in combinazione compreso il software destinato dal fabbricante ad essere impiegato specificatamente con finalità diagnostiche, terapeutiche e necessario al corretto funzionamento del dispositivo, destinato dal fabbricante ad essere impiegato sull'uomo a fini di diagnosi, prevenzione, controllo, terapia o attenuazione di una malattia; diagnosi, controllo, terapia, attenuazione o compensazione di una ferita o di un handicap; studio, sostituzione o modifica dell'anatomia o di un processo fisiologico; intervento sul concepimento, la cui azione principale voluta nel o sul corpo umano non sia conseguita con mezzi farmacologici né immunologici né mediante metabolismo, ma la cui funzione possa essere assistita da questi mezzi".

[11] Si noti che le linee guida MEDDEV sono documenti non vincolanti che però sono elaborati a seguito di una intensa consultazione a livello comunitario tra le varie parti interessate: AC, Commissione Europea, rappresentanze dell'industria di settore e altri stakeholders. esse riflettono la posizione condivisa tra rappresentanti delle predette parti interessate. Per i soli dispositivi medici impiantabili attivi valgono le definizioni contenute nell'art. 1, c. 2 del D. Lgs. 507/92, come da ultimo modificato dal D. Lgs. 25 gennaio 2010, n. 37.

finalità medica è assegnata ad un prodotto dal fabbricante. Il produttore determina attraverso l'etichetta, le istruzioni e il materiale informativo, la specifica finalità medica. Tuttavia, come abbiamo avuto modo di notare, analizzando le aree e le applicazioni della robotica e dei sistemi d'intelligenza artificiale in campo medico, una larga parte di essi, già in uso o in fase di sperimentazione, non sono dispositivi nati, progettati e fabbricati per finalità primariamente mediche e avranno quindi bisogno di essere adeguate, secondo le norme specifiche di tale settore.

Una volta stabilito che i sistemi oggetto della presente trattazione, a certe condizioni, rientrano nella definizione di dispositivo medico, è necessario verificare se al software sia applicabile la disciplina sulla responsabilità da prodotto difettoso.

Il regime di responsabilità per danni derivanti da software difettoso, varia a seconda che lo si consideri un prodotto o un servizio.

In giurisprudenza, la classificazione di un software come prodotto o servizio dipende da elementi quali, per esempio, il grado di integrazione e interdipendenza tra software e hardware all'interno di un singolo prodotto tecnologico.

Nel 2009 la Commissione Europea, seppur in riferimento a un settore diverso da quello medico, aveva annunciato che un'area prioritaria di intervento dell'UE era senza dubbio l'estensione dei principi e delle norme a tutela dei consumatori alle licenze di prodotti come il software scaricato, ad esempio, per la protezione antivirus e per i giochi. L'idea alla base di tale scelta risiedeva nella convinzione che una maggiore responsabilità per i produttori di software e per le compagnie che offrono servizi digitali, avrebbe portato ad una maggiore scelta da parte dei consumatori. Anche se questa proposta politica non è più stata sviluppata, l'organizzazione del commercio degli sviluppatori software, ha aderito ad una qualificazione del software come prodotto, nei casi di rapporti business-to-consumer. Ciò aumenterebbe la tutela dei consumatori, concedendo agli acquirenti gli stessi diritti che avrebbero in caso di acquisto di beni fisici, e in particolare, di ottenere un prodotto funzionante nel rispetto di condizioni commerciali eque. A ciò si aggiunga che l'applicabilità al software della normativa sulla responsabilità per prodotti difettosi ha visto contrapporsi la tesi di coloro che ritenevano il software un prodotto, poiché l'oggetto della prestazione di un servizio è un prodotto, all'opinione di chi distingueva il software standard da quello personalizzato, definendo il primo come un prodotto e il secondo come un servizio. La commissione CEE, espressamente interrogata sulla questione con un'interpellanza parlamentare, ha chiarito che la direttiva sui prodotti difettosi non si applica esclusivamente ai prodotti fabbricati in serie ma anche a quelli di tipo artistico o artigianale, ivi compresi i programmi per elaboratore

personalizzati[12].

3.3 La responsabilità da prodotto difettoso

I DMS, e in generale i sistemi automatici, sono per lo più composti da una combinazione di hardware e software.

Com'è noto, l'uso di sistemi software implica sempre la possibilità di un malfunzionamento. Garantire l'assenza di errori all'interno di un sistema software non è un compito elementare. Il Software Engineering Institute stima che un ingegnere esperto di software produce circa un errore per ogni cento righe di codice. Sulla base di questa stima, anche se la maggior parte dei bug possono sembrare modesti, ed esistono metodi di sviluppo e programmazione in grado di ridurre al minimo gli errori, in un codice di un milione di linee, di base, sono presenti, nel corso di un ciclo di vita tipico del programma, circa mille bug. Un bug è generalmente riconducibile ad un errore umano, progettuale, di analisi o di implementazione, durante le varie fasi di sviluppo del sistema software. Questo fenomeno è noto in letteratura come "metamorfismo dell'errore"[13]. Secondo questa interpretazione, un errore umano, ad esempio commesso in fase di analisi progettuale del software, si evolve, durante il ciclo di sviluppo del programma in uno specifico errore del sistema, durante il funzionamento. La presenza di un bug nel sistema è a sua volta in grado di innescare un'ampia varietà di effetti di propagazione. Mentre alcuni bug producono effetti lievi sulla funzionalità di un programma e possono quindi rimanere silenti per un lungo periodo di tempo, altri possono causare il blocco o il congelamento del programma stesso.

Medici e operatori sanitari, dovranno essere pronti a mitigare le conseguenze di un malfunzionamento del software ed essere in grado di gestire manualmente le attività ad alta priorità, e al tempo stesso attuare una politica di prevenzione che mitighi i possibili danni alla salute dei pazienti, provocati da possibili cedimenti del sistema.

L'allocazione delle responsabilità, in caso di un malfunzionamento dovrà tenere conto di eventuali responsabilità (a) dei programmatori, a titolo di dolo, ad esempio quando gli errori o le omissioni siano state intenzionali, ovvero, come più comunemente accade, a titolo di colpa, nella triade classica indicata dall'articolo 43 c.p. della negligenza, imperizia e imprudenza, ovvero per inosservanza

[12] La risposta della commissione CEE del 15 novembre 1988 è stata pubblicata sulla GUCE n. 114 del 1989, 42. Per ulteriori approfondimenti in tema di applicabilità al software della responsabilità da prodotto difettoso, si veda, Finocchiaro, G. *I contratti ad oggetto informatico*. Cedam, 1993, passim.

[13] Padhy, N. *Artificial intelligence and intelligent systems*. Vol. 337. Oxford University Press Oxford, 2005

di leggi, regolamenti, ordini o discipline ; (b) delle strutture sanitarie nel caso in cui abbiano omesso di aggiornare correttamente il sistema o non abbiano previsto le attività di manutenzione prescritte, o ancora abbiano adottato protocolli non idonei a garantire un livello di sicurezza adeguato; e infine (c) di medici e operatori sanitari, nel caso in cui non siano stati colposamente in grado di intervenire per evitare il danno. Il numero e la complessità dei ruoli e dei compiti coinvolti nello sviluppo, nell'implementazione e nell'utilizzo di un software rende l'attribuzione della responsabilità una questione problematica, per cui è spesso molto difficile individuare esattamente dove si è verificato l'errore, quali sono i soggetti responsabili, e in quale misura.

Nonostante la responsabilità civile abbia acquisito un ruolo centrale e assorbente nella disciplina del danno da prodotto difettoso, lasciando il tema sostanzialmente inesplorato sul versante della tutela penalistica, autorevole dottrina[14] ha più volte manifestato la necessità di apprestare una tutela penale contro il danno da prodotto. Infatti, "il diritto penale non si atteggia come un edificio dogmatico costruito sopra categorie immutabili: esso dimostra di possedere una rilevante capacità di adattamento, sovente camaleontica, che trova, però, un limite insuperabile nel rispetto dovuto ai principi di rango superiore."[15]. Il problema della responsabilità penale ha avuto così un'emersione esclusivamente prasseologica, manifestandosi solo in presenza di violazioni produttive di gravi lesioni a beni giuridici di natura individuale e collettiva[16].

[14]Cfr. Piergallini, C. *Danno da prodotto e responsabilità penale: profili dommatici e politico-criminali*. Giuffrè, 2004; Paliero, «L'autunno del patriarca. Rinnovamento o trasmutazione del diritto penale dei codici», cit., 1239 e ss., il quale aveva affrontato, con grande lucidità, il problema della responsabilità penale relativo alle attività produttive ad altissimo grado di pericolosità, pur tuttavia con esclusione di ogni riferimento ai danni derivanti da prodotti difettosi; Bricola, F. «Responsabilità penale per il tipo e per il modo di produzione». In: *La responsabilità dell'impresa per i danni all'ambien-te e ai consumatori, Milano* 87 (1978), 101 e ss.; Azzali, G. «La responsabilità penale del produttore per danni alla salute». In: *Rivista trimestrale di diritto penale dell'economia* (1991), p. 848, 850 e ss.; Stortoni, L. «Angoscia tecnologica ed esorcismo penale». In: *Riv. it. dir. proc. pen*. Vol. 71. 2004, 71 e ss.. Di opinione contraria, Stella, «Giustizia e modernità», cit., 481 e ss.. L'autore conclude per la radicale inammissibilità del ricorso al diritto penale per fronteggiare i rischi della modernità - in particolare con riferimento alla responsabilità da prodotto - in considerazione dell'impossibilità della "prova particolaristica della causalità individuale" in tali vicende processuali. Persino sul terreno della responsabilità civile la tutela delle vittime risulterebbe inadeguata o solo parziale, come insegnerebbe la dominante prassi giurisprudenziale nordamericana, ove pure non sono mancati tentativi di flessibilizzazione mediante ricorso a concetti probabilistici di causalità, seppur minoritari. Riprendendo proposte avanzate dalla dottrina nordamericana, la soluzione indicata dall'Autore consiste nell'elaborazione legislativa di un nuovo modello amministrativo che si sostituisca al modello tradizionale di responsabilità civile.

[15]Piergallini, *Danno da prodotto e responsabilità penale: profili dommatici e politico-criminali*, cit., p.606.

[16]I casi Contergan, relativo alla messa in commercio di un preparato (talidomide) ingerito da donne gestanti che poi partorirono, in gran parte figli con malformazioni congenite, e Sveso, v.

Si tratta dell'imputazione delle lesioni di beni giuridici provocate dalla produzione industriale, a fronte di cui le categorie classiche del diritto penale si rivelerebbero superate sotto alcuni aspetti qualificanti.

Per dovere di completezza, prima di analizzare nel dettaglio come la materia del danno da prodotto difettoso si intersechi con le disposizioni in materia di dispositivi medici, si ritiene opportuno un breve cenno, alle fondamentali categorie del diritto penale classico che, una volta entrate in contatto con il fenomeno del danno da prodotto, entrano in crisi.

La dottrina è unanime[17] nell'identificare quattro istituti della dogmatica penalistica più immediatamente messi alla frusta nell'ambito del danno da prodotto, quale nuovo territorio del sistema penale: (1) la fisionomia del rischio e l'accertamento causale; (2) la struttura del tipo, con particolare riferimento alla ricostruzione della posizione di garanzia sul tessuto del reato omissivo improprio; (3) il problema della responsabilità plurisoggettiva, relativo all'individuazione dei soggetti a cui imputare la responsabilità nell'ambito di organizzazioni produttive complesse; e infine (4) la struttura e i contenuti del giudizio di colpevolezza. Passiamoli brevemente in rassegna.

1. La fisionomia del rischio e l'accertamento causale.

   Una delle questioni centrali che la giurisprudenza[19] ha dovuto affrontare riguarda la possibilità di ricostruire il nesso di causalità tra danni e produzione, distribuzione, messa in circolazione e utilizzo dei prodotti. In particolare, il paradigma condizionalistico logico-deduttivo, fondato sulle leggi di copertura universali o statistiche, fatica a trovare spazio in tale contesto, per un duplice ordine di ragioni. Una prima difficoltà nella ricostruzione del nesso causale si rinviene nella presenza di un *deficit* conoscitivo sui meccanismi che hanno prodotto l'evento, che mal si prestano ad una spiegazione di tipo deterministico, incentrata sulla causa necessaria e sufficiente e rimandano, invece, a modelli caratterizzati da una rete causale, incentrata sul concetto di causazione multipla[21]. Tali difficoltà sono aggravate dalla peculiare forma diacronica di manifestazione del danno. Durante il periodo di latenza del danno la situazione di rischio rimane

Cass. 23 maggio 1986, in Cass. pen. Mass. ann. , 1998, 1250 e ss.; e Bricola, «Responsabilità penale per il tipo e per il modo di produzione», cit., 101 e ss., ne rappresentano i più drammatici esempi.

[17] Tra gli altri, Bernardi, A. «La responsabilità da prodotto nel sistema italiano: profili sanzionatori». In: *Rivista Trimestrale di Diritto Penale dell'Economia. Padova* 1-2 (2003), pp. 1–45, 22; CASTRONUOVO, D. «Responsabilità da prodotto e struttura del fatto colposo». In: *Riv. it. dir. proc. pen*. Vol. 301. 2005, 305; [18], 1226.

[19] Per un'analisi dettagliata dei procedimenti sul "Contergan" (talidomide), sulle "macchie cutanee blu", e sul "Lederspray" si vedanoPagliaro, A. «Il diritto penale tra norma e società. Scritti 1956-2008». In: *Giuffrè* II (2009), 755 e ss.; [20], 50 e ss.

[21] Vineis, P. *Modelli di rischio: epidemiologia e causalità*. Einaudi, 1990, 13 e ss.

quiescente. È frequente che tra l'immissione del prodotto sul mercato e l'evento dannoso intercorra un lasso di tempo rilevante, dando vita ad un imprecisato arco temporale in cui il rischio, conseguente all'uso del prodotto, "coltiva" la lesione, secondo una sorta di processo d'incubazione. In tal modo, il periodo di latenza funge da anticamera di possibili ipotesi causali alternative o addizionali[22]. La difficoltà di reperire leggi in grado di spiegare gli eventi dannosi ha portato la giurisprudenza a privilegiare l'insorgenza del danno corroborata dall'assenza di cause concorrenti, operando un'inversione di metodo nella ricerca del nesso causale e facendo confluire l'*explanandum* all'interno dell'*explanans*[23]. In tal modo, la causa tende a confondersi con l'effetto, a cui si riconosce un elevato valore inferenziale.

2. La struttura del tipo.

   Sul piano della struttura del tipo, le decisioni giurisprudenziali, sembrano delineare una sorta di fungibilità tra condotta attiva e condotta omissiva[24], concentrando la propria attenzione sul contenuto della posizione di garanzia, presupposto del reato omissivo improprio. In particolare, nella posizione di garanzia da cd. ingerenza[25], a differenza delle classiche posizioni di garanzia individuate ex art. 40 cp., la responsabilità per omissione non deriva da un dovere di controllo previsto *ex lege* o *ex contractu*, ma da un'azione pericolosa precedente, individuata nella produzione e messa in

[22]Così Paliero, «L'autunno del patriarca. Rinnovamento o trasmutazione del diritto penale dei codici», cit., 1239 e ss.; Aleo, S. «Causalità, complessità e funzione penale». In: *Per un'analisi funzionalistica della responsabilità penale* (2003), 59 e ss.

[23]Sul funzionamento del paradigma nomologico nella ricostruzione della causalità si vedaStella, F. *Leggi scientifiche e spiegazione causale nel diritto penale*. Vol. 22. Giuffrè, 1990, 67 e ss.

[24]Peculiare in questo senso, il procedimento tedesco relativo al Lederspraysfall, un particolare spray per pelli, sospettato di provocare danni alla salute dei consumatori. I giudici del BGH, stravolgendo l'impianto penalistico classico, individuarono nel prodotto la causa del danno, sulla base dell'esistenza di una pluralità di eventi dannosi, temporalmente coicidenti con l'uso dello spray, sull'identità dei danni alla salute riscontrati, e sull'inesistenza di ipotesi causali alternative. Inoltre, i responsabili dell'azienda produttrice furono condannati sia per il reato di lesioni omissive colpose- per non aver provveduto a ritirare il prodotto dal mercato- sia per il reato di lesioni commisive dolose, per aver deciso in una riunione colleggiale di non ritirare il prodotto dal mercato e di continuare quindi la sua messa in circolazione). Così Piergallini, *Danno da prodotto e responsabilità penale: profili dommatici e politico-criminali*, cit., 50 e ss.,

[25]ibid., 99, secondo cui da una concezione psicologica del dolo si transiterebbe, per questa via, ad una concezione tipologica, orientata a stereotipare alcune forme di intenzionalità facilitando l'ascrizione normativa della responsabilità. Secondo l'Autore "il risultato di un simile percorso è un inesorabile scolorimento del dolo nella colpa".

circolazione di prodotti difettosi o pericolosi. L'obbligo di "seguire" il prodotto, e in caso disporne il ritiro, non esiste *ab origine*, ma dipende esclusivamente e direttamente dal comportamento pregiudiziale del soggetto agente che fa insorgere il rischio. La dottrina ha notato come la cd. posizione da ingerenza sia di dubbia ammissibilità, poiché determina l'insorgere di responsabilità per il mero aumento del rischio e non, come dovrebbe, per il mancato controllo di una ben individuata fonte di pericolo, con tutte le relative ripercussioni sul parametro della percepibilità del rischio.

3. La responsabilità plurisoggettiva.

   Nell'area del danno da prodotto, la realizzazione normale della condotta è plurisoggettiva, mentre le fattispecie classiche sono *naturaliter* monosoggettive. La responsabilità plurisoggettiva è riconducibile alla frammentazione dei centri decisionali e del processo di formazione della volontà sociale[26]. Sorge il problema di come distribuire la responsabilità tra coloro che hanno partecipato alla realizzazione del prodotto rivelatosi difettoso. Le difficoltà nel determinare i criteri di allocazione delle responsabilità, nelle organizzazioni complesse, risultano ancor più accentuate dalla difficoltà di risalire alla stessa azione dannosa di un soggetto determinato. A questo proposito la dottrina si è spinta fino a parlare di anonimia del danno[27]. È evidente come, in relazione allo svolgimento di attività riconducibili alla produzione industriale e alla grande distribuzione, caratterizzate da una suddivisione in fasi dell'attività e da una frammentazione di competenze, risulti spesso difficile identificare con certezza l'autore o gli autori materiali della condotta che ha causato il danno. Tale situazione di incertezza, è certamente aggravata dalla manifestazione diacronica del vizio e del danno potenziale che ne deriva. È indubbio che i produttori di beni di largo consumo, siano presenti sul mercato nella forma di impresa collettiva, e come è già stato notato, le organizzazioni complesse si caratterizzano per una frammentazione dei centri decisionali e un decentramento verticale e orizzontale dei poteri, con tutte le problematiche che ne derivano in sede di ripartizione delle responsabilità, tra persone fisiche dipendenti ed ente, e dei soggetti gravati da posizioni di garanzia. Tuttavia, in questa sede preme sottolineare come, la qualificazione in termini di responsabilità plurisoggettiva del fenomeno del danno da prodotto non coincida esclusivamente con la responsabilità del produttore o del distributore, ma di tutti i soggetti potenzialmente coinvolti, dalla nascita fino all'uso e alla

[26] Si veda, Forti, G. *Colpa ed evento nel diritto penale*. Giuffrè, 1990, 149 e ss.

[27] Così, Piergallini, *Danno da prodotto e responsabilità penale: profili dommatici e politico-criminali*, cit., 303 e ss..

morte del prodotto. Il danno, può infatti essere il risultato dell'apporto delle condotte di soggetti diversi dal produttore, quali per esempio organi certificatori, soggetti incaricati dell'installazione e utilizzatori. Qualora sia accertato il contributo causale di uno o più soggetti alla realizzazione dell'evento, l'attribuzione della responsabilità penale dovrà necessariamente ruotare intorno all'individuazione dell'elemento soggettivo colposo, che può innestarsi anche involontariamente sulla condotta altrui.

4. La struttura e i contenuti del giudizio di colpevolezza.

   Infine, sotto il profilo della colpevolezza si pone in discussione la morfologia della riconoscibilità stessa del rischio, precondizione irrinunciabile del successivo rimprovero per dolo o colpa[28]. Il concetto di rischio non sviluppa un vero e proprio statuto giuridico, con la conseguenza che esso è adoperato, in definitiva, come sinonimo di "pericolo". In tal modo, il momento cognitivo del dolo finisce per dissolversi in una mera intuizione o sospetto del rischio, e la componente volitiva degrada a mera "attesa" dell'evento[29]. Quanto alla colpa, la difficoltà di formulare giudizi di prevedibilità e di evitabilità mina alla radice la capacità di individuare la diligenza dovuta, incrinando la possibilità di individuare la misura oggettiva e soggettiva della colpa, che costituiscono lo scheletro del relativo giudizio di rimprovero[30]. Il rischio è quello di mascherare, mediante il ricorso alla colpa, forme di imputazione per responsabilità oggettiva.

Una volta richiamati i problemi evidenziati dalla dottrina penalistica in tema di responsabilità per danno da prodotto difettoso, cerchiamo di comprendere più nel dettaglio cosa accade quando tale fenomeno entra in contatto con la disciplina giuridica dei dispositivi medici e che tipo di ripercussioni può avere sulla posizione dei medici che utilizzano tali tecnologie.

3.3.1 Responsabilità penale per danni cagionati da dispositivi medici difettosi

L'attività medica è una delle aree in cui più rilevano i profili di controllo, gestione e prevenzione del rischio. Nel primo capitolo abbiamo dedicato ampio spazio ai sistemi di *risk management* e *incident reporting* in ambito clinico, evidenziando i rapporti che intercorrono tra rischio, relazioni intersoggettive, divisione

[28] Stortoni, «Angoscia tecnologica ed esorcismo penale», cit., 87 e ss., il quale approfondisce dettagliatamente il tema della riconoscibilità del rischio

[29] Piergallini, *Danno da prodotto e responsabilità penale: profili dommatici e politico-criminali*, cit., 101 e ss.

[30] Forti, *Colpa ed evento nel diritto penale*, cit., 149 e ss.

dei compiti, organizzazione, tecnologie e criteri di ripartizione della responsabilità. L'interprete è chiamato a bilanciare costantemente l'esercizio di un'attività rischiosa, quale quella medica, e beni giuridici di rango costituzionale come la salute, l'integrità fisica e la vita stessa.

Vi è poi un area in cui la responsabilità medica, le tecnologie ad elevata automazione e i pericoli ad essa intrinsechi, e il fenomeno del danno da prodotto difettoso si intersecano creando una sorta di campo gravitazionale capace di attirare tutti i problemi finora analizzati in materia di responsabilità plurisoggettiva e colpa, nelle ipotesi di danni cagionati dal malfunzionamento di dispositivi medici. In particolare, diviene fondamentale domandarsi se l'esistenza di una certificazione di conformità dei *medical device* alla normativa CE, ad opera di un organo notificato che ne dovrebbe garantire la sicurezza, sia sufficiente a richiamare l'applicazione del principio di affidamento, quale limite all'attribuzione della responsabilità colposa del medico, che abbia utilizzato un prodotto poi rilevatosi difettoso, o se debba in ogni caso essere chiamato a rispondere, a titolo di colpa, per i danni cagionati dal dispositivo medico utilizzato. A tal fine, è necessario richiamare e analizzare la disciplina in materia di dispositivi medici dettata dalla Direttiva 93/42 CEE e successive modifiche[31], recepita in Italia dal d.lgs. 24 febbraio 1997, n. 46.

3.3.2 La Direttiva 93/42/CEE e il d.lgs di attuazione 24 febbraio 1997, n. 46: Il sistema completo di assicurazione di qualità

Esaminando la definizione di dispositivo medico e analizzando la molteplicità dei sistemi robotici e d'intelligenza artificiale, in uso in una vasta gamma di aree in ambito clinico, riabilitativo e assistenziale, è emersa l'enorme diffusività di tali sistemi nell'area oggetto di indagine, e la pluralità dei soggetti potenzialmente autori di fattispecie criminose, intimamente legate alla produzione, distribuzione e uso di tali tecnologie.

In un quadro sì fatto, uno degli obiettivi principali perseguiti dalla Direttiva 93/42/CEE, e fatto proprio dal legislatore nazionale, riveste un ruolo di centrale importanza. Tale obiettivo è identificato dal V Considerando e risiede nell'esigenza di garantire un elevato livello di protezione e sicurezza, contro i rischi per la salute di tutti coloro che vengono in contatto con dispositivi medici. Un rischio che, come più volte evidenziato, assume contorni sfumati e spesso di difficile prevedibilità, ripercuotendosi sulla difficoltà di attribuire in modo certo le responsabilità, e che pare difficilmente conciliabile con l'esistenza in commercio di un prodotto "sicuro". A ciò si aggiunga che qui la sicurezza attiene ad un

[31] La Direttiva 2007/47 CE, recepita nel nostro ordinamento dal D.Lgs. n. 37 del 25 gennaio 2010, ha modificato la Direttiva 93/42 CEE cui facciamo riferimento.

concetto di natura prettamente tecnica, che accompagna il prodotto lungo tutto il suo ciclo di vita. In caso di danni, la responsabilità sarà potenzialmente in grado di coinvolgere tutti i soggetti che hanno apportato modifiche o sono entrati in contatto con il dispositivo, dal momento della sua progettazione fino al suo concreto utilizzo, originando un fenomeno di concatenazione funzionale e cronologica di posizioni e di eventuali soggetti attivi del reato. Peraltro, all'interno di un'area caratterizzata da attività pericolose e tuttavia necessarie, il requisito della sicurezza prescritto dalla direttiva CE deve necessariamente confrontarsi con il parametro del cd. rischio consentito e con il potenziale affidamento dei soggetti coinvolti sull'altrui diligenza. È evidente che il rischio è consentito anche laddove per l'attività da esercitare sia richiesta l'esistenza di autorizzazioni amministrative e certificazioni di sicurezza o qualità. Queste, infatti, in via eccezionale, rendono esplicitamente lecito lo svolgimento di determinate attività, subordinandone l'esercizio al rispetto di precise norme cautelari.

La Direttiva 93/42 CEE individua a tal fine un sistema di tutela e controlli *ex ante*, complesso e articolato, affidato ad autorità amministrative o privati, che intervengono a decorrere dalla fase immediatamente successiva alla produzione del dispositivo. Anche il legislatore nazionale, nel recepire la direttiva, ha disposto (a) un sistema di controllo *ex ante* affidato al produttore e, ove previsto, a organismi notificati o privati, e (b) una fase *ex post*, successiva alla messa in commercio del prodotto, che vede coinvolti soggetti privati e autorità amministrative.

La procedura di conformità ai requisiti essenziali di efficacia e sicurezza da applicare, ex art. 11 del d.lgs. n. 46/97, varia in base alla classe di rischio cui appartiene il dispositivo, secondo le regole e i parametri dettati nell'Allegato IX. Occorre precisare che l'assoluta assenza di rischio, pur idealmente auspicabile, non è raggiungibile per quanto riguarda i dispositivi medici; peraltro, in considerazione del beneficio clinico per il paziente derivante dall'utilizzo del dispositivo, si può accettare un certo livello di rischio residuo, maggiore quanto più grande sarà il beneficio apportato[32].

I dispositivi medici sono classificati, ex d.lgs.46/97, secondo quattro criteri:

[32] Anche per la gestione dei rischi, e specificatamente di quelli connessi ai dispositivi medici, esiste una specifica norma tecnica, la EN 14971, che consente al fabbricante, che decida di seguirla, di compiere tutte le fasi previste per tale gestione, a partire dall'analisi dei rischi. Questa consiste nell'identificazione dei pericoli connessi con l'utilizzo di un dato dispositivo e nella quantificazione del rischio che il danno si verifichi. Tale quantificazione è correlata direttamente alla probabilità dell'evento sfavorevole e alla severità degli effetti negativi possibili, ed inversamente alla facilità di evidenziarlo. Per la gestione del rischio esistono numerose metodologie; le principali tra esse sono anche richiamate nella norma tecnica EN 14971; nella norma è precisato che la gestione del rischio deve essere affidata a personale adeguatamente formato e chiaramente identificato e che il fabbricante deve dedicare risorse umane ed economiche adeguate a tale aspetto.

1. Invasività del dispositivo: dispositivi non invasivi, invasivi negli orifizi del corpo, invasivi chirurgici, impiantabili;
2. Durata del contatto con il corpo: temporanea, breve termine, lungo termine;
3. Sede anatomica su cui incide il dispositivo medico: in particolare sistema circolatorio centrale e sistema nervoso centrale; e infine
4. Dipendenza da una fonte di energia: dispositivo non attivo, dispositivo attivo terapeutico, dispositivo attivo diagnostico.

Sulla base dei parametri elencati, si distinguono quattro classi di rischio:

- Classe I: appartengono a questa classe i dispositivi medici meno critici, quali la gran parte di quelli non attivi e non invasivi;
- Classe IIa: comprende i dispositivi medici a medio rischio, quali alcuni dispositivi non attivi (invasivi e non) e dispositivi attivi che interagiscono con il corpo in maniera non pericolosa;
- Classe IIb: comprende dispositivi medici a rischio medio/alto, quali alcuni dispositivi non attivi (specie invasivi) e i dispositivi medici attivi che interagiscono con il corpo in maniera pericolosa;
- Classe III: comprende i dispositivi medici ad alto rischio, quali gran parte dei dispositivi impiantabili, quelli contenenti farmaci o derivati animali ed alcuni dispositivi che interagiscono sulle funzioni di organi vitali, come per esempio i dispositivi invasivi ad uso chirurgico di tipo temporaneo, destinati a diagnosticare, sorvegliare o correggere difetti del cuore o del sistema circolatorio centrale, e quelli destinati ad essere utilizzati in contatto diretto con il sistema nervoso centrale. Appartengono a tale classe anche le protesi d'anca, spalla e ginocchio.

La sezione 1 dell'allegato IX riporta tutte le definizioni dei termini utilizzati

per la classificazione[33].

Maggiore è la classe di rischio, maggiore è la complessità della procedura di valutazione e attestazione di conformità da seguire. Per i *medical device* che appartengono alla classe più rischiosa (Classe III), si dovrà ricorrere ad un sistema cd. qualificato di garanzia della qualità, volto a verificare il possesso dei requisiti essenziali previsti dalla disciplina comunitaria, e necessaria al fine di apporre la marcatura CE di cui all'art. 16 del D.Lgs. n. 46/97, indispensabile per la libera circolazione del dispositivo all'interno della Comunità Europea. In particolare, per i dispositivi di classe superiore alla I la conformità viene valutata, con diverse modalità, anche da un soggetto di terza parte, denominato Organismo Notificato, che la attesta mediante una certificazione rilasciata al fabbricante.

Tale sistema non può quindi che riverberarsi sul valore da attribuire alle garanzie di sicurezza del dispositivo che si riveli difettoso e fonte di danno.

In particolare, in fase di accertamento di responsabilità per danni derivanti dal malfunzionamento del prodotto, dovrà tenersi necessariamente conto del ruolo dell'organo notificato e verificare l'eventuale sussistenza, in capo al medesimo, di eventuali profili di responsabilità penale, e valutare come l'apporto dell'Organismo Notificato e l'apposizione della certificazione CE, incidano su profili di responsabilità altrui.

L'Organismo notificato non si limita a garantire la semplice qualità del prodotto, ma altresì, e in conformità a quanto disposto ex art. 11, comma 1, lett. a), la qualità del sistema di produzione e di fabbricazione, dando luogo al cd. *"sistema completo di assicurazione di qualità"*, o, più in generale, *premarket control*. Oltre

[33]In particolare, (a) i dispositivi non invasivi sono quelli che non penetrano in alcuna parte del corpo, né attraverso un orifizio né attraverso la cute; (b) i dispositivi invasivi negli orifizi del corpo sono quelli che penetrano negli orifizi, intendendo per orifizio qualsiasi apertura naturale del corpo, compresa la superficie esterna del globo oculare, oppure qualsiasi apertura artificiale e permanente, quale uno stoma; (c) i dispositivi di tipo chirurgico sono quelli che penetrano attraverso la superficie del corpo sia nel contesto di un intervento chirurgico sia al di fuori di tale contesto, quindi per esempio anche l'ago di una siringa; (d) i dispositivi impiantabili sono quelli destinati ad essere impiantati totalmente nel corpo (oppure a sostituire una superficie epiteliale o la superficie oculare) mediante intervento chirurgico e a rimanere in tale sede dopo l'intervento, oppure quelli destinati ad essere introdotti parzialmente nel corpo mediante intervento chirurgico e a rimanere in tale sede per almeno trenta giorni; e infine (e) i dispositivi attivi sono quelli che ricorrono a qualche forma di energia - elettrica o di altro tipo - per funzionare (il software indipendente o stand-alone è considerato un dispositivo medico attivo). I dispositivi impiantabili attivi, sono legati per il loro funzionamento ad una fonte di energia elettrica o a qualsiasi altra fonte di energia diversa da quella prodotta direttamente dal corpo umano e dalla gravità. Sono destinati ad essere impiantati interamente o parzialmente, mediante intervento chirurgico o medico, nel corpo umano o mediante intervento medico in un orifizio naturale e destinati a restarvi dopo l'intervento. Pur non essendo distinti in classi, di fatto sono equiparabili ai dispositivi medici di Classe III ai fini delle procedure di marcatura.

alla certificazione della conformità del prodotto alle disposizioni della Direttiva comunitaria, tale sistema prevede:

- la revisione del sistema di qualità da parte dell'Organismo certificatore, al fine di stabilire se esso risponda a quanto documentato dal fabbricante, mediante visita presso la sede, per controllare i procedimenti di fabbricazione;
- sorveglianza dell'Organismo certificatore sul fabbricante, per garantire che costui soddisfi correttamente gli obblighi derivanti dal sistema di qualità approvato; e infine
- periodiche ispezioni e valutazioni dell'Organismo, per accertarsi che il fabbricante applichi il sistema di qualità approvato[34].

Una volta esaurita la procedura di controllo e accertata la rispondenza del dispositivo ai requisiti essenziali di cui all'art. 3 del D.Lgs. 46/97 e alla Direttiva, l'attività dell'organismo notificato culmina, necessariamente, con l'apposizione della marcatura di conformità CE. Essa attesta la rispondenza di prodotti industriali, tra cui rientrano anche i dispositivi medici, ai requisiti essenziali di sicurezza, sanità pubblica, tutela del consumatore e rispetto dell'ambiente[35]. A questo punto, è d'obbligo chiedersi quali siano i profili di rilevanza penale del contrassegno CE. Il valore attribuito alla certificazione comunitaria e al marchio di conformità CE del dispositivo muta a seconda che alla stessa sia attribuita valenza meramente formale oppure propriamente sostanziale. Qualora il controllo svolto dall'organismo notificato si riducesse a un esame della documentazione cartacea in possesso del produttore, che attesta il rispetto della normativa comunitaria in sede di produzione e l'esistenza dei requisiti essenziali del prodotto, il marchio di conformità avrebbe rilevanza puramente formale. Al contrario, la certificazione di conformità assumerebbe rilievo sostanziale qualora fosse rilasciata a seguito di un vero e proprio controllo "materiale" della produzione e del dispositivo, per esempio, a seguito di ispezioni e verifiche sul posto. Inoltre, dopo una prima analisi della disciplina e dei criteri di classificazione, ex d.lgs. 46/97, sembra possibile rilevare un primo punto di criticità. Il legislatore, non sembra aver tenuto conto di un parametro, vale a dire il livello di automazione, che a chi scrive pare essenziale per stabilire il livello di rischio dei *medical device*.

[34] In alternativa il produttore potrà scegliere la procedura di cui Allegato III (certificazione CE) e Allegato IV (verifica CE) o V (garanzia di qualità della produzione) che implica,in ogni caso, l'intervento dell'organismo *de quo*.

[35] Si veda, Cipolla, P. «Profili penali del contrassegno CE». in: *Giurisprudenza di merito* 10, 2133 (2012), 2147 e ss., il quale peraltro afferma che in tali casi potrebbe assegnarsi al contrassegno CE il valore di autocertificazione della rispondenza del prodotto agli standard di sicurezza con cui il produttore pertanto assume la responsabilità di eventuali difetti e/o danni.

Abbiamo avuto modo di vedere come una visione socio-tecnica del sistema sanità porti al superamento di regole univoche di organizzazione, in favore di una visione di coerenza tra tutte le variabili del sistema. In particolare le variabili tecniche, che scaturiscono dalle modalità operative e applicative di cui si avvale il lavoro umano, e di cui fanno parte le tecnologie, non possono non essere prese in considerazione tra i criteri utilizzati per la classificazione dei dispositivi in termini di rischio. Il rischio da ignoto tecnologico, è un rischio generalizzato che coinvolge tutte le componenti del sistema sanità che non può non rilevare in questa sede. Una visione socio-tecnica del rischio non può prescindere da un analisi che tenga conto dell'interazione tra uomo e tecnologie e dunque di una tassonomia dei livelli di automazione, potenzialmente capace di incidere, a sua volta, sulla disciplina di certificazione, sul danno e sui profili di responsabilità.

3.3.3 Ripercussioni in tema di responsabilità penale

Dall'esame della direttiva 93/42 CEE e del D.lgs. di attuazione, è emerso come per l'attività di produzione, distribuzione, commercio e utilizzo di dispositivi medici, i soggetti necessariamente coinvolti dalla nascita alla morte del prodotto siano almeno tre, e più in particolare, il produttore, l'organismo certificatore, e infine l'utilizzatore[36]. Ciò conferma il potenziale apporto plurisoggettivo allo svolgimento di attività rischiose, che richiedono livelli di conoscenza molto elevati da parte degli operatori, e in cui l'evoluzione scientifica costante, può assumere un valore determinante in sede di giudizio. È indubbio, che tutti i soggetti sopra elencati, attraverso le proprie condotte attive, o omissive, siano potenzialmente in grado di contribuire alla realizzazione del danno cagionato da un *medical device* difettoso, con evidenti ripercussioni sulla colpevolezza e la ripartizione della responsabilità.

Possiamo ora identificare, almeno due profili di collegamento con i temi trattati, di cui si sono esposte le criticità nel corso della trattazione. Da un lato, sembra evidente il facile riproporsi delle difficoltà relative al giudizio di causalità della colpa, in particolare sotto i profili di prevedibilità *ex ante* e di evitabilità dell'evento, in relazione alla responsabilità penale per danni cagionati da dispositivi medici difettosi.

In particolare, nei casi in cui sia rilevabile un difetto del *medical device* è proprio l'elemento della riconoscibilità della regola cautelare prima, e del difetto successivamente, ad assumere un ruolo discriminante nell'attribuzione della responsabilità penale a titolo di colpa. Sarà fondamentale individuare il momento esatto in cui il pericolo e il difetto, insito nel prodotto, può e deve essere

[36] In questa sede, per ciò che sembra rilevante ai fini della trattazione, si è omesso di analizzare il ruolo dell'importatore, così come disciplinato dalla Direttiva esaminata.

riconoscibile.

A ciò si aggiunga che la riconoscibilità del difetto dovrà essere declinata, almeno nel *quantum*, in relazione al soggetto di volta in volta considerato e alle competenze e capacità che tali soggetti hanno o dovrebbero avere in base alla relativa figura di agente modello. Si pensi ad esempio, alle particolari conoscenze che il progettista modello deve avere, così come l'organismo notificato che attesta la conformità del *medical device*, diverse tra loro, e a loro volta distinte da quelle del medico utilizzatore.

Sotto il profilo dell'imputazione soggettiva dell'evento, la colpevolezza, e più in particolare l'esistenza del dovere di prevedere, riconoscere e prevenire le imprudenze altrui, dovrà essere vagliata in relazione alla condotta del produttore, dell'organismo certificatore e del medico utilizzatore, così da individuare i corretti criteri di ripartizione della responsabilità penale, tra i diversi soggetti coinvolti. Il ruolo di tali soggetti deve essere valutato anche alla luce delle condotte altrui, a causa della tecnicità della materia e della continua evoluzione scientifica capaci di influenzare in modo determinante il comportamento di tali soggetti, sia come singoli, sia in rapporto al loro inserimento all'interno dell'intero sistema socio-tecnico, di cui fanno parte.

Abbiamo evidenziato, ancora una volta, la frammentazione, sotto il profilo del soggetto attivo del reato, in relazione alla responsabilità penale per danni cagionati da dispositivi medici difettosi.

Nella sezione successiva, si procederà all'analisi della posizione del medico che utilizzi un dispositivo rivelatosi difettoso, con particolare attenzione alla relazione tra medico e organismo notificato e al ruolo da attribuire alla certificazione di conformità CE, in sede di attribuzione di eventuali profili di responsabilità penale colposa.

3.4 La Responsabilità medica: percezione del rischio e imprevedibilità. Il principio di affidamento e la posizione del medico per danni cagionati da dispositivi medici difettosi.

Il tema della responsabilità professionale del medico, e in particolare della responsabilità penale, è di particolare difficoltà, sia sotto i profili che riguardano una sua definizione teorica, sia dal punto di vista delle soluzioni pratiche. Essa riguarda una funzione sociale di tutela di beni giuridici di valore primario, quali la vita e l'integrità fisica della persona umana, costituzionalmente riconosciuti, e che postulano una tutela penale contro le condotte colpose, lesive di quel bene.

La responsabilità medica, in una visione socio-tecnica, appare legata ad un sistema composito, costituito dal complesso dei rapporti che, oltre a quello di tipo personale, si istituiscono nel momento in cui un soggetto è destinatario di prestazioni mediche di ogni tipo, diagnostiche, preventive, ospedaliere, terapeutiche, chirurgiche o assistenziali. L'attività medico-sanitaria non coinvolge solo i medici, ma anche personale con diversificate qualificazioni, come per esempio, infermieri e assistenti sanitari, ostetriche, tecnici di radiologia medica e tecnici di riabilitazione[37]. A questo, si deve aggiungere il particolare contesto in cui medici e operatori sanitari, si trovano ad operare. Uno dei temi di maggiore attualità per gli ordinamenti giuridici contemporanei, riguarda la tutela degli interesse fondamentali, della vita e della salute, contro i rischi da cosiddetto *ignoto tecnologico*. Analizzando le aree mediche e i sistemi di automazione, abbiamo visto come robot estremamente sofisticati e sistemi d'intelligenza artificiale, coadiuvino attivamente e talvolta sostituiscano l'opera di medici e operatori sanitari. Ancora, in relazione alla responsabilità per danni cagionati da dispositivi medici difettosi, abbiamo evidenziato la frammentazione sotto il profilo del soggetto attivo potenzialmente responsabile.

In un contesto così eterogeneo, il tema della responsabilità dovrà necessariamente tenere conto della relazione tra tecnologie, personale medico, organismo notificato e organizzazione nel suo complesso.

Tra le fenomenologie criminose la cui complessità, sia oggettiva, relativa cioè al tipo di attività, sia soggettiva, relativa alle entità che realizzano tale attività, si scontra con i tradizionali modelli di ascrizione della responsabilità, la dottrina penalistica italiana, segnala quale esempio paradigmatico, la responsabilità medica, per danni, lesioni e morte dei pazienti.

La prima questione centrale, che la giurisprudenza ha dovuto affrontare riguarda la qualificazione causale della condotta del medico rispetto al danno penalmente rilevante, su cui negli ultimi anni dottrina e giurisprudenza si sono lungamente soffermate, spesso con risultati mutevoli[38]. Vi sono situazioni in cui la condotta del medico può essere facilmente giudicata causale, rispetto all'evento infausto, come per esempio nel caso in cui un chirurgo, eseguendo una colecistectomia per via laparoscopica, laceri un vaso importante, provocando un'emorragia e la morte del paziente. Vi sono casi in cui la condotta- sicuramente errata- del medico, seguita dalla morte del paziente, non può tuttavia

[37] In questo sensoAlpa, G. «La responsabilità medica». In: *Resp. civ. prev* (1999), pp. 315–336,p. 315.

[38] Per approfondimenti, si vedano Aleo, S., Centonze, A. e Lanza, E. *La responsabilità penale del medico*. Giuffrè Editore, 2007; eBilancetti Mauro e Bilancetti, F. *La responsabilità civile e penale del medico*. Cedam, Padova, 2010

essere giudicata causale rispetto a questa, perché la condotta corretta non avrebbe affatto garantito con sicurezza un risultato diverso.[39]. E vi sono ipotesi di incidente, occorso a seguito di malfunzionamento del dispositivo, dovuto, non a usura e mancata manutenzione, nel cui caso sarebbero responsabili i soggetti tenuti alla manutenzione della stessa, ma per esempio a causa di un difetto del sistema, per cui il medico si troverà costretto a intervenire manualmente, e stabilire se la causa di eventuali danni al paziente sia riconducibile al malfunzionamento del dispositivo o per esempio all'imperizia del medico è questione di non poco conto. Ancora, nel caso di lesioni o di morte del paziente incorsi, per esempio, a seguito di errore di programmazione di un robot chirurgico il soggetto imputabile sarà il produttore della macchina, o ancora più specificamente, nel caso in cui si riesca ad individuarlo con certezza, il soggetto che ha elaborato il programma di quello specifico strumento, o il medico, nel caso in cui sia riscontrabile un suo comportamento colposo. In tutte queste ipotesi, la principale difficoltà nella ricostruzione del nesso causale si rinviene nella presenza di un deficit conoscitivo sui meccanismi che hanno prodotto l'evento, che mal si prestano ad una spiegazione di tipo deterministico, incentrata sulla causa necessaria e sufficiente e rimandano, ancora una volta, a modelli caratterizzati da una rete causale, incentrata sul concetto di causazione multipla[40]. Tali difficoltà sono implementate dalla manifestazione diacronica del danno[41].

Come anticipato, in questa sede si è scelto di approfondire la posizione del medico in relazione all'organismo notificato e al ruolo da attribuire alla certificazione di conformità CE, in sede di attribuzione di eventuali profili di responsabilità penale colposa.

In base al principio dell'equivalenza delle cause vigente nel nostro ordinamento[42], un determinato fatto lesivo è ascrivibile a tutti coloro che pongono in essere una condotta che rappresenti, nella progressione causale, un antecedente

[39] Aleo, Centonze e Lanza, *La responsabilità penale del medico*, cit., 2 e ss.

[40] Cfr. Vineis, *Modelli di rischio: epidemiologia e causalità*, cit.

[41] Facciamo riferimento a casi come quello Roland Mracek v. Bryn Mawr Hospital e v. Intuitive Surgical Inc. (D.C. Civil No. 08-cv-00296 Pennsylvania) del 2005. Mracek subì un intervento alla prostata a seguito di una diagnosi di tumore. Il chirurgo utilizzò per l'intervento il robot chirurgico Da Vinci, prodotto e venduto dalla Intuitive Surgical, Inc.. Durante l'intervento il Da Vinci andò in stallo e comparve un messaggio di errore sul display. Nessuno del personale in sala riuscì a farlo ripartire. Il chirurgo fu costretto a terminare l'intervento in laparoscopia. Ad una settimana dall'operazione il paziente rilevò ematuria, disfunzioni erettili e forti dolori e decise di fare causa all'ospedale e alla casa produttrice del "Da Vinci". Il caso è noto soprattutto per i profili di responsabilità civile, ma astraendo dalla fattispecie concreta, che si è risolta per i parametri della legislazione americana con il rigetto della domanda (Mraceck non ha fornito, infatti, la prova dell'esclusione dell'esistenza di cause secondarie di malfunzionamento e che lo stesso abbia causato il danno) lo stesso offre spunti di riflessione in tema di responsabilità medica.

[42] In accordo a quanto affermato anche dalla Corte di Cassazione nella sentenza n. 40897/2011

necessario al verificarsi del fatto medesimo. In particolare, in relazione al fenomeno dei danni cagionati da dispositivi medici difettosi, il momento di inizio della progressione causale, che può eventualmente condurre alla realizzazione di eventi di morte o lesioni, va individuato, secondo la giurisprudenza[43], nella commercializzazione dei prodotti, coinvolgendo tutti i soggetti cui sopra si è fatto riferimento. Inoltre, come più volte sottolineato, il fenomeno del danno da dispositivi medici difettosi è caratterizzato da una concatenazione funzionale e cronologica di posizioni che dovrà necessariamente riflettersi in sede di accertamento della responsabilità. In conformità ai principi generali del diritto penale, tale accertamento dovrà necessariamente essere legato all'elemento soggettivo del reato, così da stabilire se nella condotta dei soggetti coivolti, siano ravvisabili elementi di colpa, sotto i profili della negligenza, imprudenza e imperizia o della violazione di leggi, regolamenti, ordini o discipline, e come ciò si rapporti con l'ingerenza di eventuali imprudenze altrui.

In caso di danno derivante dall'uso di un dispositivo medico difettoso marcato CE, e con riferimento alla posizione del medico utilizzatore, la scelta tra pronuncia di condanna a titolo di concorso, cooperazione colposa, o assoluzione in virtù del principio di affidamento, quale limite all'imputazione dell'evento a titolo di colpa, dipenderà dal valore attribuito al parametro di riconoscibilità (a) del difetto, e/o (b) del pericolo e del dovere di prevedere e riconoscere inosservanze altrui. È utile ricordare che secondo la teoria del reato colposo, la riconoscibilità è legata alle conoscenze nomologiche e ontologiche del soggetto agente nel caso concreto[44].

Tuttavia, è stato più volte evidenziato come, nell'ambito del fenomeno del danno da prodotto, la riconoscibilità del difetto ceda il passo alla mera percepibilità del pericolo o del vizio insito nel prodotto e alla consapevolezza di operare in una situazione di rischio da cui deriva un dovere di intervento molto sfumato e che talvolta appare illimitato. Tale svalutazione del parametro di riconoscibilità si ripercuote inevitabilmente sui fenomeni di interazione tra più soggetti, dove la prevedibilità e la riconoscibilità dell'evento, coinvolge le condotte di tutti i soggeti menzionati, rischiando di sconfinare in un dovere di diligenza relazionale assoluto.

Fino a pochi anni fa, parte della dottrina e della giurisprudenza di legittimità, con riferimento alla teoria del reato colposo, avevano ritenuto l'esistenza in

[43] Così anche Cass. pen., sez. IV 13.4.2011, n. 15002 in DeJure.

[44] Così, Forti, *Colpa ed evento nel diritto penale*, cit., 201 e ss., il quale, in riferimento al rilievo assunto dal parametro della riconoscibilità in rapporto alla tipicità colposa, ritiene si debba parlare di dovere di riconoscere la realizzazione del fatto, atta a determinare il contenuto della regola cautelare.

astratto di un obbligo di prevedere le imprudenze altrui[45], escludendo o limitando l'applicabilità del principio di affidamento. Se si applicase una tale visione alle ipotesi di danno da dispositivo medico marcato CE, sorgerebbe in capo al medico un dovere di prevedere e impedire le negligenze di produttori, distributori e, addirittura, organismi notificati, a causa dello svolgimento di un'attività pericolosa e intimamente connessa a quella del produttore. Solo così, potrebbe escludersi il valore di attestato di sicurezza e affidabilità da attribuire alla marcatura CE. In pieno accordo, con quanto più volte sostenuto da Mantovani[46], si ritiene che, ai fini di una corretta applicazione del giudizio di responsabilità ex art. 43 c.p. sia necessario riferirsi alla situazione personale dell'agente, nel momento in cui ha agito – o si è astenuto dall'agire -[47]. Solo la riconoscibilità del rischio, in virtù delle circostanze concrete, attiva il dovere di evitare l'evento[48] e giustifica un intervento anche nelle ipotesi di inosservanze riconducibili a terzi.

In conclusione, si ritiene che, nelle fattispecie di omicidio o lesioni che derivino da un difetto di un dispositivo medico certificato CE, debba potersi applicare il principio di affidamento, salvi i casi in cui il difetto sia palese e riconoscibile[49]. L'orientamento secondo cui la certificazione rilasciata da un Organismo notificato autorizzato avrebbe valenza puramente formale[50], escludendo così l'efficacia scusante della marcatura CE, nei confronti di distributori e medici che abbiano messo in commercio e utilizzato un dispositivo difettoso, è da escludere anche a fronte di alcune recenti sentenze.

Un caso di particolare interesse è quello noto alle cronache come il caso delle "valvole Killer". Un chirurgo aveva installato valvole cardiache di ultimissima generazione, munite di certificazione CE, su una pluralità di pazienti. Il successivo malfunzionamento delle valvole, dovuto a un difetto congenito della

[45]Con particolare riferimento al settore della circolazione stradale, tale dottrina era arrivata a precludere, *tout court*,l'applicazione del principio di affidamento. Cfr. Duni, G. «L'obbligo di prevedere le condotte altrui». In: *Rivista giuridica della circolazione e dei trasporti* (1964).

[46]Mantovani, M. *Il principio di affidamento nella teoria del reato colposo*. A. Giuffrè, 1997, 218 e ss. e in contrasto rispetto a quanto sostenuto da Duni, «L'obbligo di prevedere le condotte altrui», cit., il relazione al settore specifico della circolazione stradale.

[47]Pulitanò, D. *Diritto penale*. Giappichelli, 2009, 350.

[48]Mantovani, *Il principio di affidamento nella teoria del reato colposo*, cit., 220

[49]In questi termini si è pronunciata la Suprema Corte di Cassazione, sez. IV, con sentenza n. 18140/2012, in DeJure, stabilendo che: "Il principio di affidamento, che è coerente applicazione del principio di personalità della responsabilità penale, in forza del quale ciascuno risponde delle conseguenze della propria condotta, commissiva od omissiva, nell'ambito delle proprie conoscenze e specializzazioni mentre non risponde dell'eventuale violazione delle regole cautelari da parte di terzi, non è utilmente richiamabile quando l'affidante risulti già in colpa, per avere non osservato le regole cautelari rientranti nella propria sfera di competenza".

[50]Si veda Trib. Torino, sez. III penale, 19 febbraio 2009, cui ha fatto seguito Corte d'appello Torino 2010 e Cass. sez. IV 13 aprile 2011, n. 15002, in DeJure. In questo caso il giudice aveva attribuito alla certificazione CE, valore meramente cartolare.

protesi, aveva determinato in alcuni casi la morte dei pazienti e in altri aveva reso necessari interventi d'urgenza per la sostituzione dei dispositivi, prolungando lo stato di malattia dei pazienti, e derminando l'insorgenza di psicopatie, depressione, disturbi post-traumatici da stress e altre conseguenze penalmente rilevanti. Le protesi valvolari cardiache, ai sensi della Direttiva 93/42/CEE[51], sono annoverate tra i dispositivi di classe di rischio III.

Il Tribunale di Padova, con sentenza 9 giugno 2008, aveva dichiarato la colpevolezza per i fatti sopra descritti: a) dei soci e dirigenti della società produttrice delle valvole cardiache difettose, in quanto responsabili della progettazione, sperimentazione, produzione, controllo di qualità, immissione in commercio e messa in servizio dei suddetti dispositivi medici; b) dei titolari della ditta importatrice, responsabili per aver importato e fornito il menzionato dispositivo difettoso e dannoso; e infine c) del cardiochirurgo e direttore del Centro di cardiochirurgia universitario per aver impiantato le valvole in questione su più pazienti, alcuni dei quali erano poi deceduti per edema polmonare e insufficienza cardiaca, mentre altri erano stati sottoposti ad intervento operatorio di urgenza, per la sostituzione con altre protesi afferenti a una categoria più affidabile. In particolare, il chirurgo era stato ritenuto responsabile a titolo di colpa per "avere utilizzato valvole di nuova generazione, senza una sufficiente letteratura ed esperienza sull'argomento, senza essersi informato presso altri centri di cardiochirurgia sull'utilizzo di tali valvole, senza avere informato adeguatamente i pazienti, senza uno scrupoloso monitoraggio dopo l'intervento, e violando l'art. 12 del codice deontologico circa l'utilizzo di terapie e tecniche nuove". Il giudice di primo grado non aveva, quindi, tenuto in considerazione la certificazione con marcatura CE.

Il giudice di secondo grado, aveva confermato la colpevolezza a titolo di concorso, dei tre soggetti apicali della società produttrice e dei distributori, mentre aveva escluso la responsabilità del cardiochirurgo, a qualsiasi titolo, sia in relazione ai decessi che alle lesioni contestatigli per aver effettuato gli interventi di cui sopra. Secondo il giudice di appello, il medico aveva correttamente fatto affidamento su quanto illustratogli dagli esperti del ramo, e valutato positivamente, dal punto di vista medico, quanto appreso da quella fonte. Il chirurgo aveva riposto la propria fiducia nell'esito positivo degli impieghi all'estero di valvole già testate in vitro e in vivo, munite della certificazione CE e quindi "presuntivamente affidabili".

La Suprema Corte di Cassazione[52] ha confermato la responsabilità dei produttori per non avere vigilato adeguatamente sulla produzione delle protesi; ha

[51] Direttiva 93/42/CEE, punto 15 preambolo, art. 9, all. 9, parte 3, sez. 2.4, regola 8

[52] sentenza della sez. IV della Suprema Corte di Cassazione n. 40897/2011, in DeJure.

prosciolto gli apici della ditta distributrice, poiché costoro "potevano legittimamente immettere sul mercato, a norma dell'articolo 5 del d.lgs. 24 febbraio 1997, n.46, il dispositivo medico di cui si discute, perché regolarmente munito della marcatura di conformità CE di cui all'art. 16 dello stesso decreto, requisito questo idoneo a determinare sicuro affidamento, circa i requisiti di sicurezza del prodotto, nei commercianti". La Corte ha aggiunto che i distributori erano privi delle necessarie cognizioni ingegneristiche o mediche e per tale motivo su di essi "non gravava alcun obbligo di esprimere una qualsiasi autonoma valutazione ed eventuali riserve sull'affidabilità e sulla sicurezza delle valvole distribuite". Infine, ha confermato l'assoluzione del medico cardiochirurgo, sottolineando il valore della certificazione di qualità CE, di cui i dispositivi incriminati erano dotati e quindi, implicitamente, la relazione intersoggettiva, di fiducia e aspettativa, che naturalmente si crea tra medico-utilizzatore e organismo notificato.

In conclusione, e in linea con quanto stabilito dalla Suprema Corte di Cassazione nella sentenza n. 40897/2011 e in conformità al principio di affidamento, si ritiene di poter attribuire efficacia sostanziale alla certificazione CE. Questa, per altro, sembra essere l'unica strada percorribile per poter garantire il rispetto del principio di responsabilità penale personale ex art. 27, comma 1 Cost., escludendo la sussistenza del dovere di riconoscere la negligenza altrui anche nei casi in cui ciò non sia in alcun modo prevedibile, per la sola interconnessione soggettiva che si crea in certi settori. In questo modo, si garantisce che il giudizio sul rischio e la prevedibilità ed evitabilità dell'evento, intimamente legato al principio di affidamento, come limite all'attribuzione di responsabilità a titolo di colpa, sia interpretato correttamente e in linea con il principio di colpevolezza. La responsabilità penale può aversi solo qualora siano posti in essere atti ed eventi, la cui realizzazione rientri nella sfera di controllo del soggetto agente[53].

3.5 Responsabilità penale e sistemi d'intelligenza artificiale

Il mondo tecnologico cambia rapidamente e robot e avanzati sistemi d'intelligenza artificiale sostituiscono sempre più spesso gli esseri umani nello svolgimento di attività complesse.

Secondo la definizione di Marvin Minsky "Artificial intelligence is the science of making machines do things that would require intelligence if done by men" Minsky, M. «Steps toward artificial intelligence». In: *Computers and thought* 406 (1963), p. 450. Con il termine intelligenza artificiale si intende, quindi, l'abilità di un computer di svolgere funzioni e ragionamenti tipici della mente umana.

[53] Pulitanò, *Diritto penale*, cit., 290.

Nel suo aspetto puramente informatico, essa comprende la teoria e le tecniche per lo sviluppo di algoritmi che consentano alle macchine, e tipicamente ai calcolatori, di sviluppare un'abilità e/o un' attività intelligente, in domini specifici. Finché l'uomo ha utilizzato i computer come meri strumenti, non vi era alcuna reale differenza tra un computer e un cacciavite o un telefono. Il problema è nato quando i computer si sono evoluti da macchine programmate per eseguire processi computazionali definiti, in macchine pensanti, dotate d'intelligenza artificiale e di capacità computazionale autonoma Padhy, N. *Artificial intelligence and intelligent systems*. Vol. 337. Oxford University Press Oxford, 2005.

Gli agenti intelligenti, anche detti agenti software (AS), sono sistemi informatici in grado di agire autonomamente all'interno di un determinato ambiente, senza il controllo diretto del loro utilizzatore. Ogni AS è in grado di percepire l'ambiente, attraverso i suoi organi sensori, e spesso di modificarlo, mediante i suoi organi effettori. L'attività dell'agente software è autonoma: essa è determinata da processi cognitivi compiuti dallo stesso AS. Gli agenti elettronici possono operare sia nel mondo reale (come ad esempio i robot) sia in quello virtuale.

Da alcuni anni, gli scienziati si scontrano sulla vera essenza degli agenti dotati d'intelligenza artificiale[54]. I futurologi hanno proclamato la nascita di una nuova specie, machina sapiens: creature intelligenti che divideranno con l'uomo lo spazio sulla terra Winograd, T. «Thinking machines: Can there be? Are we». In: *The boundaries of humanity: Humans, animals, machines* (1991), pp. 198–223. I critici hanno sostenuto che il concetto stesso di "machina sapiens" risulta ontologicamente un ossimoro. Le macchine, compresi i robot, con la loro logica fredda, non saranno mai in grado di essere penetranti e creative come lo sono gli esseri umani ibid. Tale polemica solleva questioni sull'essenza stessa dell'uomo (gli esseri umani funzionano come macchine pensanti?) e dell'intelligenza artificiale (possono esistere macchine pensanti?).

Secondo alcuni [55] per poter definire un'entità dotata d'intelligenza dovrebbe essere possibile ascrivergli cinque attributi: a) capacità di comunicazione: con un'entità intelligente è possibile comunicare e più è semplice, più tendiamo ad ascrivere intelligenza al soggetto che abbiamo di fronte. È possibile comunicare con un cane, ma non della teoria della relatività di Einstein, mentre è possibile parlarne con un bambino, anche se la discussione deve essere affrontata in

[54]Si vedano Winograd, T. «Thinking machines: Can there be? Are we». In: *The boundaries of humanity: Humans, animals, machines* (1991), pp. 198–223; Dreyfus, H. L. *What computers can't do: The limits of artificial intelligence*. Vol. 1972. Harper & Row New York, 1979; Boden, M. A. *The Philosophy of Artificial Intelligence*. 1990

[55]Schank, R. C. «What is AI anyway?» In: *The foundation of artificial intelligence—a sourcebook*. Cambridge University Press. 1990, pp. 3–13; Russel, S. e Norvig, P. «Artificial Intelligence: A Modern Approach, 2003». In: *EUA: Prentice Hall* ()

termini che l'interlocutore è in grado di comprendere; b) conoscenza interna: un'entità intelligente dovrebbe avere una certa conoscenza di sé; c) conoscenza esterna: a un'entità intelligente è richiesta la percezione del mondo esterno, per poter imparare e utilizzare le informazioni; d) il quarto attributo è ciò che la letteratura scientifica chiama goal-driven behavior, ovvero la capacità di agire per raggiungere i propri obiettivi; e) infine, la creatività: un entità intelligente dovrebbe avere un certo grado di creatività, intesa come capacità di agire in modo alternativo quando l'azione inizialmente posta in essere ha dato esito negativo. Una mosca che tenta di uscire da una stanza e sbatte contro il vetro di una finestra continua a ripetere infinite volte la medesima azione. Quando un robot intelligente urta contro una finestra, dovrebbe essere progettato perché tenti di uscire dalla porta. La maggior parte degli agenti dotati d'intelligenza artificiale dovrebbe quindi possedere uno o più di questi attributi [56]. Alcuni tipi di agenti intelligenti possiedono caratteristiche ancor più sofisticate.

La letteratura ha spesso cercato di inquadrare il rapporto tra agenti autonomi intelligenti, sviluppatori e utenti, specialmente in ambito civilistico e, in particolare, secondo lo schema del contratto di agenzia e mandato per le ipotesi di contratti conclusi da agenti software, interrogandosi anche sull'opportunità di riconoscere personalità giuridica e patrimonio autonomo agli agenti intelligenti.

Sul fronte del diritto penale, la responsabilità dei sistemi d'intelligenza artificiale è un territorio ancora molto inesplorato. Esistono tuttavia alcune eccezioni. Un primo modello elaborato dalla dottrina propone un'analogia tra agenti intelligenti (AI), dotati di capacità cognitive e stati mentali, e animali. Più di recente Gabriel Hallevy, professore di Diritto Penale presso la Facoltà di Giurisprudenza dell'Ono Accademic College in Israele, ha proposto tre modelli di responsabilità penale degli AI: (1) Perpetration-via-Another, (2) Natural-Probable- Consequence, e (3) Direct Liability[57]. Alla base dei modelli elaborati dalla dottrina, vi è la percezione che gli agenti autonomi intelligenti, non possano essere considerati alla stregua di semplici prodotti, a fronte delle loro autonomia e delle loro capacità cognitive e di apprendimento automatico. Tali modelli saranno presentati nelle sezioni successive.

[56] Schank, «What is AI anyway?», cit.; Hallevy, G. «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control». In: *Akron Intell. Prop. J.* 4 (2010), p. 171 e Karnow, C. E. «Liability for distributed artificial intelligences». In: *Berkeley Technology Law Journal* (1996), pp. 147–204

[57] Hallevy, G. «I, Robot—I, Criminal"—When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses». In: *Syracuse Sci. & Tech. L. Rep.* 22 (2010), pp. 1–9; idem, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit.

3.5.1 Il modello zoologico

La maggior parte degli ordinamenti giuridici disciplina il rapporto tra esseri umani e animali, da un lato, considerandoli come possibile oggetto del diritto di proprietà e, dall'altro, come oggetto di tutela contro abusi e crudeltà.

Il primo aspetto riguarda il diritto di proprietà e possesso sugli animali, generalmente disciplinato dal Diritto Civile[58], secondo cui se un animale causa un danno, il proprietario è giuridicamente responsabile di tale danno[59]. Allo stesso modo, se un cane attacca e morde un essere umano, in caso di lesioni, è il proprietario ad essere responsabile per il danno arrecato. Gli ordinamenti disciplinano queste ipotesi attraverso il Diritto Civile o la Tort Law e in una minoranza di casi il Diritto Penale ma, in ogni caso, il soggetto giuridicamente responsabile rimane il proprietario.

Nessun ordinamento giuridico considera gli animali soggetti di diritto.

Rispetto a ipotesi di abusi e crudeltà, essi sono tutelati dai vari ordinamenti attraverso il Diritto Penale. Per esempio, il codice penale italiano disciplina il reato di maltrattamento di animali ex art. 544-ter c.p.. L'oggetto giuridico del reato tutelato dall'ordinamento non è l'animale ma il comune sentimento di pietà umana e la sensibilità degli esseri umani verso gli animali[60]. Per tali fattispecie, la protezione dei diversi ordinamenti non ha nulla a che vedere con il diritto di proprietà, chiunque può essere incriminato per il reato di maltrattamento di animali a prescindere da eventuali diritti di proprietà o possesso.

Su queste premesse nasce il modello zoologico.

Esso propone un'analogia tra AI, dotati di capacità cognitive e stati mentali, e animali. Si tratta di un modello discusso in dottrina da molto tempo ed esaminato anche in relazione al controllo di aerei senza pilota e ad alcuni robot di nuova generazione[61].

[58] Secondo il nostro Codice Civile possono costituire oggetto del diritto di proprietà, acquistabile a titolo originario e a titolo derivativo. A titolo originario possono essere acquistati tramite occupazione, ex articolo 923 c.c.; per inerzia del proprietario nel casco dello sciame di api; per una speciale forma di usucapione, abbreviata a venti giorni, per gli animali domestici ex articolo 925 c.c..

[59] Il Codice Civile italiano, per esempio, disciplina la responsabilità del proprietario o utilizzatore per i danni cagionati da un animale ex art. 2052 c.c..

[60] L'attuale formulazione dell'articolo 544-ter c.p. , introdotta dalla legge 20 luglio 2004, n. 189 e successivamente aggiornata dalla Legge 4 novembre 2010, n. 201, tutela la sensibilità degli esseri umani verso gli animali, diversamente dall'art. 727 c.p. che, prima dell'entrata in vigore della legge 20 luglio 2004, n. 189, puniva il reato di maltrattamento di animali tutelando il comune sentimento di pietà umana.

[61] Si vedano per esempio, Gabbai, J. M. «Complexity and the aerospace industry: Understanding emergence by relating structure to performance using multi-agent systems». Tesi di dott. Citeseer, 2005; Reynolds, C. W. «Flocks, herds and schools: A distributed behavioral model». In: *ACM SIGGRAPH computer graphics*. Vol. 21. 4. ACM. 1987, pp. 25–34; Newman, W. S. «Automatic

Il parallelo tra AI e animali ha così trovato ampio spazio in letteratura. Per esempio, in *Guilty robots, happy dogs: the question of alien minds*, David McFarland[62] sostiene che i rapporti giuridici degli esseri umani con i robot debbano essere inquadrati nello stesso modo in cui lo sono quelli con gli animali, per le ipotesi in cui provochino un danno e sia riscontrabile un comportamento colposo o doloso del proprietario.

Un resoconto interessante su questo modello si trova nel report *Legal Issues of Software Agents* del consorzio LEGAL-IST[63]. Gli autori costruiscono un modello di responsabilità, chiamato *Dog Model*, attraverso un'analogia tra agenti software (AS), dotati di capacità cognitive e stati mentali, e cani addestrati. Nonostante abbiano natura profondamente diversa, e prescindendo da considerazioni di tipo etico, entrambi sono addestrati e programmati per perseguire autonomamente i compiti e gli obiettivi assegnati, ed entrambi sono considerati come oggetti dall'ordinamento giuridico. Le conseguenze giuridiche dei loro comportamenti ricadrebbero sul proprietario dell'AS nello stesso modo in cui ricadono sul proprietario di un cane. In particolare, l'esempio è quello di un cane addestrato come cane da pastore, o a condurre persone non vedenti o a prestare soccorso agli esseri umani in caso di annegamento. Obiettivi e compiti sono assegnati sia al cane che all'AS, entrambi sono addestrati o programmati per svolgere autonomamente azioni finalizzate al raggiungimento di obiettivi predeterminati. L'addestratore di animali è quindi paragonabile allo sviluppatore del software. Successivamente, il cane e l'AS vengono trasferiti ai relativi proprietari e utenti, che ne acquistano la proprietà o il possesso, e che li utilizzeranno secondo le funzioni stabilite. Il comportamento del cane, dotato di propri stati mentali e cognitivi, non può essere del tutto prevedibile *ex ante*. Esso impara dall'esperienza, agisce autonomamente ed è dotato di intelligenza. Adotterà un certo comportamento sulla base dei propri compiti e obiettivi e della propria interazione con l'ambiente. Situazioni inaspettate e imprevedibili potrebbero portare a risultati indesiderati. Tali considerazioni restano valide anche nel caso in cui si tratti di agenti software dotati di capacità cognitive e di sviluppare propri stati mentali[64].

Gli autori costruiscono due scenari differenti a seconda che il cane e l'AS siano addestrati a svolgere compiti e raggiungere obiettivi leciti o illeciti.

---

obstacle avoidance at high speeds via reflex control». In: *Robotics and Automation, 1989. Proceedings., 1989 IEEE International Conference on*. IEEE. 1989, pp. 1104–1109; Teubner, G. «Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law». In: *Journal of Law and Society* 33.4 (2006), pp. 497–521.

[62] McFarland, D. *Guilty robots, happy dogs: the question of alien minds*. OUP Oxford, 2008.

[63] LEGAL-IST. «Report on Legal Issues of Software Agents». In: *Doc.* D14.Rev. 2 (2006), pp. 1–140, 76 e ss..

[64] ibid.

Nella prima ipotesi, qualora si verifichi un danno, l'addestratore, il proprietario e l'utente potranno essere penalmente responsabili solo nel caso in cui abbiano agito con negligenza, imprudenza o imperizia, mettendo in atto un comportamento colposo. Inoltre, sia il cane che l'AS potrebbero aver sviluppato propri sotto-obiettivi e stati mentali in contrasto con quelli per cui erano originariamente stati addestrati e programmati. In mancanza di dolo o colpa dell'addestratore, del programmatore e del proprietario, nessuna forma di responsabilità penale potrà essergli addebitata.

Diverso è il caso in cui sia riscontrabile una forma di dolo dell'addestratore, del programmatore o del proprietario, o di entrambi. La responsabilità penale dipenderà dal rapporto di causa-effetto tra il danno e chi lo ha causato[65].

In entrambi gli scenari, trascurando la possibilità di addebitare il fatto al cane o all'AS, ritenendoli direttamente responsabili delle proprie azioni e del danno, gli unici soggetti penalmente responsabili saranno l'addestratore, il programmatore, il proprietario o l'utente, qualora soddisfino i requisiti della *mens rea*.

In *The Laws of Robots*, sulla base di alcuni criteri sviluppati da Colin Allen, Gary Varner e Jason Zinser, e sviluppati ulteriormente da Luciano Floridi e Jeff Sanders, Pagallo[66] evidenzia tre caratteristiche principali degli AI, che spiegano il paragone con gli animali piuttosto che con le cose o i prodotti. In primo luogo robot e AI sono interattivi, capaci di percepire l'ambiente che li circonda e rispondere agli stimoli, modificando i valori delle loro proprietà e stati interni. La seconda caratteristica riguarda l'autonomia. Robot e AI modificano i propri stati interni esercitando un controllo sulle proprie azioni senza la necessità di un intervento diretto da parte degli esseri umani. Infine, essi hanno capacità adattive, sono in grado di perfezionare le regole attraverso cui le loro proprietà e stati interni si modificano[67].

3.5.2 Alcune considerazioni, critiche e obiezioni

Secondo una parte della dottrina, che pure riconosce il valore del modello zoologico e la sua capacità di fornire risposte in certi contesti, come quello aeronautico[69], alcuni problemi di fondo non posso essere risolti utilizzando questo

[65] ibid.

[66] Pagallo, U. *The Laws of Robots*. Vol. 200. Springer, 2013.

[67] Si vedano Allen, C., Varner, G. e Zinser, J. «Prolegomena to any future artificial moral agent». In: *Journal of Experimental & Theoretical Artificial Intelligence* 12.3 (2000), pp. 251–261; [68].

[69] Gabbai, «Complexity and the aerospace industry: Understanding emergence by relating structure to performance using multi-agent systems», cit.; Newman, «Automatic obstacle avoidance at high speeds via reflex control», cit.

approccio[70]. Quando un AI è in grado di comprendere e risolvere autonomamente una certa situazione, mediante il proprio software, qualcosa nel puzzle della responsabilità si perde.

Secondo Hallevy[71] la comunicazione di idee e concetti particolarmente sofisticati è molto più semplice con un AI che non con gli animali. Un AI è programmato per conformarsi al ragionamento logico formale degli esseri umani, esso è la base dei calcoli, del ragionamento e delle azioni dell'AI. Diversamente, la capacità di ragionamento degli animali non è basata necessariamente sulla logica formale, tipica del ragionamento umano.

Ancora, la sfera emotiva influenza in modo importante i comportamenti di esseri umani e animali. Ciò non accade con gli AI poiché privi di stati emotivi. Se l'unità di misura è la sfera emotiva, animali ed esseri umani sono molto più vicini tra loro di quanto non lo siano rispetto ad un AI. Se l'unità di misura è la capacità cognitiva e di ragionamento, allora, gli AI sono molto più vicini agli esseri umani che agli animali. Nonostante la sfera emotiva sia in grado di influenzare quella cognitiva, la legge mantiene una distinzione tra le due aree. In particolare, il diritto penale, nel valutare la *mens rea* e l'imputabilità, considera esclusivamente gli stati cognitivi e solo eccezionalmente quelli emotivi. Così, nell'ipotesi di furto gli stati emotivi del soggetto attivo del reato, quali per esempio invidia, frustrazione e odio, non sono giuridicamente rilevanti. Il codice penale italiano disciplina questa ipotesi ex art. 90 c.p., escludendo che gli stati emotivi e passionali possano incidere sull'imputabilità. Tuttavia, in certi casi hanno effetto di attenuante comune e sono elementi costitutivi dei delitti d'onore. La *ratio* alla base dell'art. 90 del nostro Codice Penale, così come delle norme di altri ordinamenti che ritengo irrilevanti gli stati emotivi ai fini dell'imputabilità, è quella di evitare che qualunque delitto impulsivo possa essere non punibile indipendentemente da uno stato di infermità. La legge ritiene infatti che salvo i casi di infermità, un normale individuo sia in grado di inibire i propri stati emotivi e passionali, frenando e controllando le proprie azioni. Sotto questo profilo, gli AI sono molto più vicini agli esseri umani che agli animali. Sulla base di tali considerazioni, il modello zoologico risulterebbe inadatto per valutare la responsabilità degli AI [72]. Anche se in linea teorica la legge potrebbe disciplinare il comportamento degli AI nello stesso modo in cui disciplina quello degli animali, è necessario prepararsi ad una nuova classe di azioni, giuridicamente rilevanti, che non sono del tutto riconducibili agli esseri umani e a

[70]Si veda per esempio Hallevy, G. «Unmanned vehicles–Subordination to criminal law under the modern concept of criminal liability». In: (2012).

[71]Hallevy, G. *When Robots Kill: Artificial Intelligence Under Criminal Law*. Northeastern University, 2013, 22 e ss.

[72]In questo senso, Lessig, L. *Code and other laws of cyberspace*. Vol. 3. Basic books New York, 1999e Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit.

malapena agli animali[73].

3.6 Personalità giuridica e responsabilità penale degli AI

Gabriel Hallevy, nell'ormai celebre articolo *The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control*[74] propone per la prima volta tre modelli di responsabilità penale applicabili agli AI, ampliati e approfonditi in alcuni lavori successivi.

Il presupposto di tali modelli risiede nel riconoscimento di personalità giuridica agli agenti autonomi intelligenti. Il tema del legame tra agenti e persone fisiche o giuridiche, che di essi si avvalgono, è al centro del dibattito dottrinale da molto tempo, e in particolare in ambito civilistico, in relazione agli effetti dei negozi giuridici compiuti dagli AI. Il primo ad ipotizzare la soggettività giuridica dei robot è stato il filosofo americano, Hilary Putnam, che aveva ipotizzato già negli anni 60 il riconoscimento dei diritti civili per i robot[75]. Come noto, numerosi ordinamenti giuridici riconoscono ad alcune entità, come per esempio associazioni e società, personalità giuridica, il cui tratto distintivo consiste nel possedere autonomia patrimoniale rispetto al patrimonio delle persone fisiche che le compongono o le amministrano. Autorevole dottrina ha sostenuto che un simile approccio potrebbe essere adottato anche con riferimento agli AI: essi potrebbero avere un proprio patrimonio e rispondere nei limiti di questo delle obbligazioni assunte[76]. Altra parte della dottrina ha sostenuto la possibilità di replicare la situazione che vigeva nel diritto romano per lo schiavo, poiché considerato una *res* che rispondeva nei limiti di un patrimonio separato detto *peculium*[77]. In questo caso però l'ipotesi di una soggettività fittizia avrebbe meno

[73] In questo senso Pagallo, *The Laws of Robots*, cit.

[74] Hallevy, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit.

[75] Putman, H. «I robot: macchine o vita creata artificialmente?» In: *Mente, Linguaggio e Realtà* (1987), pp. 416–438.

[76] Sartor, G. «Gli agenti software: nuovi soggetti del ciberdiritto». In: *Contratto e impresa* 2 (2002), pp. 57–91. Sartor, rifacendosi alle tesi di Dennett ritiene che i sistemi intelligenti operanti nella rete abbiano una intenzionalità nel senso che tendono con razionalità ad un obbiettivo. Avrebbero degli stati psicologici simili a quelli umani perché tendono apparentemente in modo ragionevole ad un obbiettivo. Sostiene che gli agenti sono molto autonomi, molto imprevedibili e dotati di una, sia pure estrinseca, intenzionalità. Sartor applica l'impostazione di Putnam e di Dennett secondo cui gli agenti obbediscono a leggi psicologiche. La soggettività intenzionale sarebbe una soggettività ridotta che porterebbe ad applicare agli agenti la disciplina delle persone e non delle cose a livello di contratti e di responsabilità.

[77] Cfr.Searle, J. R. «Minds, brains, and programs». In: *Behavioral and brain sciences* 3.03 (1980), pp. 417–424, Weitzenboeck, E. *Electronics agents: some other legal issues. ECLIP 2nd Summer School, Palme De Mallorca*. 2001.

forza. Nel diritto romano classico si riteneva, infatti, che il *peculium* continuasse ad appartenere al *dominus*,il quale era chiamato a rispondere dei debiti contratti dallo schiavo solo nei limiti del *peculium* stesso.

Sul fronte del diritto penale, uno studio recente di Human Rights Watch sui *"Robot Killer"*, automi da utilizzare in contesti di guerra con la capacità di selezionare gli obiettivi e ingaggiare un combattimento senza l'intervento umano, ha affrontato gli interrogativi più spinosi legati al loro uso, in relazione a eventuali condotte illecite e ad errori che il robot potrebbe compiere ove non sia in grado di distinguere tra obiettivi militari e obiettivi civili[78]. Lo studio ha rilevato che l'automa non potrebbe rientrare nella giurisdizione delle corti internazionali non essendo una *"natural person"* e che, in ogni caso, anche laddove le convenzioni internazionali fossero cambiate, un eventuale giudizio non raggiungerebbe alcun risultato utile per la società o per la vittima del crimine posto che il robot non subirebbe né un effetto deterrente dalla condanna, né tanto meno potrebbe percepirne il disvalore. Di parere contrario, Hallevy[79], il quale, non solo, è favorevole al riconoscimento della personalità giuridica degli AI, sulla base del modello adottato dagli ordinamenti giuridici per le società, ma anche ad una loro condanna penale. Nelle sezioni che seguono, sarà presentata una panoramica dei tre modelli, con particolare attenzione a quello della responsabilità diretta degli AI.

3.6.1 Primo modello: The perpetration through another

Il primo modello affonda le proprie radici lontano nel tempo, e più in particolare nel modello della responsabilità vicaria, riconosciuta dal diritto civile e penale e un tempo basata sul concetto di schiavitù e proprietà[80].

[78] Si veda, Mind the Gap. The Lack of Accountability for Killer Robots, lo studio è disponibile all'indirizzo https://ildirittodeirobot.files.wordpress.com/2015/04/arms0415_forupload_0.pdf. Si veda anche McFarland, T. e McCormack, T. «Mind the Gap: Can Developers of Autonomous Weapons Systems be Liable for War Crimes». In: *Int'l L. Stud. Ser. US Naval War Col.* 90 (2014), p. i

[79] Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit.

[80] L'origine storica della responsabilità vicaria è certamente inquadrabile nella tradizione giuridica romanistica, caratterizzandosi originariamente per la responsabilità del *pater familias* per i reati commessi dai membri della propria famiglia e dagli schiavi soggetti a patria potestà. Agli inizi del XIX secolo, l'evoluzione di questo modello ha portato a delineare tre ipotesi principali di responsabilità vicaria. La prima è riconducibile alla figura classica del concorso nel reato. Se più persone cooperavano nel commettere il medesimo reato si ricorreva alla figura del concorso, anche nel caso in cui fosse possibile riscontrare una relazione gerarchica tra i soggetti coinvolti. Tuttavia, nel caso in cui la relazione gerarchica fosse accompagnata da un gap informativo, o dall'uso di una posizione di forza per far sì che una delle parti coinvolte non fosse in grado di avere piena coscienza e volontà del fatto di reato, quest'ultima era considerata un agente innocente, e penalmente non responsabile per il reato commesso. Da questa ipotesi emergono le basi

Il modello della *perpetration through another*, già previsto da diversi ordinamenti nazionali, è stato formalmente accolto dal diritto penale internazionale, ex art. 25 comma 1, lett. a) dello Statuto di Roma della Corte Penale Internazionale là dove afferma che è personalmente responsabile anche chi ha commesso il fatto *through another person*. Il *perpetrator* di un crimine è colui che realizza l'ipotesi criminosa attraverso l'azione o l'omissione di un'altra persona, utilizzando quest'ultima come uno strumento per la commissione del crimine. Il reato, dunque, viene attribuito al soggetto che, dominando il fatto, utilizza l'autore materiale del crimine come uno strumento, senza incidere sul piano della formazione della volontà di quest'ultimo[81]. L'autore mediato, ovvero il *perpetrator*, è ritenuto responsabile a prescindere dalla responsabilità dell'autore materiale. La naturale applicazione di questa forma di responsabilità si ha quando il reato è materialmente commesso da un agente innocente, penalmente non perseguibile perché, per esempio, ha agito in stato di incapacità o intossicazione, o perché infermo di mente o perché minore non imputabile. Il soggetto che materialmente realizza la condotta è solo uno strumento nelle mani di chi lo utilizza per realizzare il crimine. Tale forma si distingue pertanto da una forma di istigazione in cui l'esecutore materiale agisce con un'intenzione formatasi in tutto o in parte sotto l'impulso dell'istigatore.

Il primo modello presentato da Hallevy considera l'AI privo di capacità cognitive e assimilabile all'incapace di intendere e volere, o al minore non imputabile. L'AI è quindi un agente innocente e le sue capacità cognitive non sono sufficienti a incontrare i requisiti della *mens rea*. Esso è quindi lo strumento di chi architetta e commette il reato, il *perpetrator*. Come tale, è quest'ultimo a essere penalmente responsabile per la condotta criminosa (*actus reus*) dell'AI. L'elemento soggettivo del reato (*mens rea*) è, infatti, determinato dallo stato mentale

del modello della*perpetration through another*. L'agente innocente era considerato uno strumento funzionale alla commissione del reato di cui era penalmente responsabile il cosiddetto *perpetrator*. Tra la fine del XIX e l'inizio del XX secolo, emerge un terzo modello di responsabilità vicaria. La figura dell'agente innocente si espande fino a ricomprendere ipotesi non riconducibili a relazioni gerarchiche tra i soggetti coinvolti, come per esempio casi di coercizione o minacce. Nel corso del XX secolo, questo modello finisce per inglobare l'ipotesi dell'agente semi-innocente: tipicamente un soggetto negligente, non pienamente consapevole della situazione di fatto. Oggi molti ordinamenti riconoscono la figura dell'agente semi-innocente come parte del modello della *perpetration through another*, riconoscendo il semi-innocente responsabile del reato a titolo di colpa e le altre parti a titolo di dolo. L'esempio classico è quello dell'infermiera che si trova in sala operatoria con il chirurgo, il quale si accinge ad operare una persona che tempo prima aveva aggredito l'infermiera. Quest'ultima decide di infettare gli strumenti operatori, ingannando il chirurgo che li utilizza successivamente durante l'intervento. In caso di morte del paziente, l'infermiera risponde a titolo di dolo e se il chirurgo aveva l'obbligo di controllare la sterilità degli strumenti, sarà considerato un agente semi-innocente e risponderà a titolo di colpa.

[81] Amati, E. *Introduzione al diritto penale internazionale*. Vol. 2. Giuffrè Editore, 2010, 116 e ss.

del *perpetrator*[82].

Esistono due candidati principali per il ruolo di *perpetrator through another*: il programmatore del software e l'utente. Il programmatore dell'AI potrebbe aver ideato e costruito un programma al fine di commettere un reato attraverso l'AI. Tra gli esempi portati da Hallevy c'è quello del programmatore che costruisce un robot operativo, intenzionalmente collocato all'interno di una fabbrica e il suo software è programmato per appiccare un incendio durante la notte, mentre i lavoratori non ci sono. Il robot appicca l'incendio ma sarà il programmatore, in questo caso *perpetrator through another*, ad essere penalmente responsabile per la condotta criminosa del robot, nel caso specifico incendio doloso. Il secondo candidato a svolgere il ruolo di autore mediato è l'utente, ovvero l'utilizzatore dell'AI. L'utente non ha progettato e programmato il software ma utilizza l'AI e il suo software per raggiungere scopi illeciti. L'esempio è quello dell'utente che acquista un robot domestico programmato per eseguire qualsiasi ordine datogli dal padrone. Il robot identifica il nostro utente come il padrone, che gli ordina di attaccare e assalire chiunque si introduca in casa. Il robot esegue l'ordine esattamente, attaccando l'intruso e provocandogli lesioni gravi. Questa ipotesi non è diversa dal caso in cui una persona ordini al proprio cane di attaccare qualsiasi intruso[83]. Il robot mette in atto la condotta criminosa ma sarà il programmatore a rispondere del reato commesso dall'AI.

In entrambi gli scenari l'*actus reus* è posto in essere dall'AI, il programmatore e l'utente non hanno compiuto nessuna azione conforme alla condotta criminosa del reato specifico. Il modello della *perpetration through another* considera l'azione commessa dall'AI come se fosse stata commessa dal programmatore o dall'utente. La *ratio* alla base della responsabilità di questi ultimi risiede nell'uso strumentale dell'AI come agente innocente. La *mens rea* richiesta per l'attribuzione della responsabilità è imputabile al programmatore e all'utente, in entrambi gli scenari. Tutte le volte in cui un programmatore o un utente utilizzano un AI in modo strumentale per commettere un reato ricadono nella figura del *perpetrator*. Questo modello di responsabilità assume che l'AI sia completamente dipendente dal programmatore o dall'utente e non possa autogovernarsi e autodeterminarsi. Esso è valido per due scenari possibili. Il primo riguarda il caso in cui l'AI sia strumentalmente utilizzato per commettere un reato senza che l'utilizzatore o il programmatore si avvalgano delle capacità avanzate dell'AI. Il secondo è il caso in cui si utilizzi un robot o un AI privo delle avanzate capacità cognitive dei moderni sistemi d'intelligenza artificiale. Tuttavia, l'AI è

[82] Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit.; idem, «I, Robot—I, Criminal"—When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses», cit.; idem, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit.

[83] idem, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., 73 e ss..

utilizzato per la sua capacità di eseguire un ordine e commettere un reato. Un semplice cacciavite non può eseguire un ordine, diversamente da un cane che tuttavia non è in grado di eseguire ordini sofisticati e complessi, mentre un AI sì[84]. Il modello della *perpetration through another* non può, quindi, essere applicato a ipotesi in cui l'AI decida di commettere un reato sulla base della propria esperienza o conoscenza. Non può essere applicato nel caso in cui commetta un reato pur non essendo stato programmato per commettere un reato specifico e nonostante questo l'AI abbia commesso un reato e, infine, nel caso in cui l'AI agisca come agente semi-innocente. [85].

3.6.2 Secondo modello: The Natural Probable Consequence

Il secondo modello assume un profondo coinvolgimento dell'utente o del programmatore nello svolgimento delle attività dell'AI. La responsabilità come *Natural Probable Consequence* si basa sulla capacità di programmatori e utenti di prevedere la potenziale commissione di reati da parte di robot e sistemi d'intelligenza artificiale. Essi saranno penalmente responsabili del reato commesso dall'AI, se tale reato è la conseguenza naturale o probabile di un loro comportamento colposo o doloso.

Hallevy distingue due ipotesi principali che portano a esiti differenti.

La prima si verifica quando programmatori o utenti non hanno intenzione di commettere alcun reato, ma a causa di un comportamento negligente, nella programmazione o nell'uso dell'AI, quest'ultimo commette un reato. Tra gli esempi proposti da Hallevy[86], troviamo quello di un AI progettato per identificare minacce provenienti da Internet e proteggere un dato sistema informatico. Pochi giorni dopo l'attivazione del software, l'AI comprende che il modo più efficace per rilevare tali minacce è accedere a siti internet e sistemi informatici, distruggendo qualsiasi software riconosciuto come una minaccia alla sicurezza del sistema protetto. L'AI commette il reato di accesso abusivo e danneggiamento di sistema informatico, nonostante il programmatore non abbia programmato l'AI, o l'utente non se ne sia servito, per commettere il reato. Questa ipotesi non permette di applicare il modello della *Perpetration Through Another*, che presume l'intenzione criminale del programmatore e/o dell'utente. Nel nostro esempio essi non sono a conoscenza del reato, e non lo hanno previsto, difettano del requisito dell'intenzione e non hanno partecipato alla condotta criminosa. Ciò

[84] ibid.

[85] ibid., 74 e ss.. Hallevy, descrive un agente semi-innocente come una parte negligente, non pienamente consapevole della situazione di fatto, mentre qualsiasi altra persona ragionevole, nelle medesime circostanze, avrebbe potuto essere a conoscenza della situazione.

[86] ibid., 75 e ss.

che rileva ai fini di una responsabilità criminale del programmatore o dell'utente, secondo il modello della *Natural Probable Consequence*, è l'aver adotatto un comportamento negligente, penalmente rilevante nell'ipotesi in cui, secondo un criterio di normale diligenza, avrebbero dovuto prevedere la possibilità che il reato fosse commesso, evitando quindi il verificarsi del danno[87].

La seconda ipotesi si occupa dei casi di responsabilità riconducibili alla figura del concorso nel reato, vale a dire quando programmatori o utenti hanno programmato o utilizzato un AI per commettere un reato, ma l'AI ha deviato il proprio comportamento rispetto al piano originario commettendo un'altro reato, in aggiunta o in sostituzione al reato programmato. In questo caso Hallevy sembra fare ricorso a una particolare ipotesi di *aberratio delicti*, e in particolare alla figura del concorso anomalo nel reato, prevista da numerosi ordinamenti e dal nosro Codice Penale ex art. 116[88]. Secondo questo istituto, qualora il reato commesso sia diverso da quello voluto da taluno dei concorrenti, anche questi ne risponde, se l'evento è conseguenza delle sua azione od omissione. Perché ricorra tale figura è necessaria la presenza di tre requisiti: a) l'adesione psichica dell'agente ad un reato concorsuale diverso; b) la commissione da parte di altro concorrente di un reato diverso; c) un nesso psicologico in termini di prevedibilità tra la condotta dell'agente compartecipe e l'evento diverso concretamente verificatosi. L'esempio di scuola è quello di un gruppo di correi che concorda inizialmente di commettere un furto: mentre uno dei partecipi si limita a fare da basista e palo, gli esecutori materiali commettono, in difformità dal piano criminoso originario, una rapina e un sequestro ai danni del soggetto rapinato[89].

Similmente, Hallevy costruisce l'esempio in cui un programmatore progetta un AI per commettere una rapina in banca, senza prevedere nel piano criminoso, e a causa della propria negligenza, l'uccisione di un essere umano. Durante l'esecuzione della rapina l'AI uccide una delle persone presenti all'interno della banca, che aveva tentato di resistere alla rapina.

È necessario analizzare separatamente la responsabilità per il reato pianificato e la responsabilità per in reato non pianificato.

Se il programmatore ha progettato l'AI per commettere il reato e l'AI ha agito come agente innocente, allora si applicherà il modello della *Perpetration Through Another*. Il programmatore o l'utente saranno gli unici ad essere penalmente responsabili per il reato pianificato, in quanto autori mediati. Nel caso in cui l'AI non sia un agente innocente, si avrà un ipotesi di concorso nel reato, applicando

[87] ibid.; idem, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit.

[88] cfr., Cass., I, 23 febbraio 1995, n. 3381.

[89] Caso tratto da Cass., 3 marzo 1978, in Giustizia Penale, 1979, II, 274.

il modello della responsabilità diretta, esaminato nella sezione successiva, in combinazione alla responsabilità del programmatore.

Per il reato non pianificato è necessario un approccio diverso. Hallevy distingue tra intelligeza artificiale forte e intelligenza artificiale debole.

Se l'AI appartiene alla prima categoria, ed è in grado di soddisfare i requisiti della *mens rea*, sarà penalmente responsabile per il reato commesso, secondo il modello della *Direct Liability*, e la responsabilità del programmatore, o dell'utente, sarà determinata secondo il modello della *Natural Probable Consequence*, appena descritta. Essi saranno responsabili per il reato pianificato e per il reato non pianificato.

Se l'AI è qualificabile come un sistema d'intelligenza artificiale debole, e non è in grado di soddisfare i requisiti della *mens rea*, l'AI non sarà responsabile del reato non pianificato. Se si verificano queste condizioni l'AI è considrato un agente innocente. La responsabilità per il reato non pianificato rimane in capo al solo programmatore o utente, a causa del comportamento negligente in fase di progettazione o nell'uso dell'AI, e si aggiunge alla responsabilità per il reato pianificato, secondo il modello della *Natural Probable Consequence*.

In conclusione, se l'AI ha agito come agente innocente, non sarà responsabile né del reato pianificato, né del reato non pianificato. In queste circostanze le azioni dell'AI ricadono nel modello della *Perpetration through another*. Tuttavia, se l'AI non ha agito come agente innocente, alla responsabilità del programmatore o dell'utente, secondo il modello della *Natural Probable Consequence*, si aggiunge la responsabilità dell'AI, secondo il modello della responsabilità diretta, come descritto di seguito.

3.6.3 Terzo modello: The Direct Liability

Il terzo e ultimo modello non assume alcuna dipendenza dell'AI dal programmatore o dall'utente[90], e mira a fornire un quadro teorico di equivalenza funzionale tra AI ed esseri umani, ai fini della responsabilità penale. Per questa ragione, tale modello merita maggiore attenzione nella nostra analisi. Hallevy parte dall'idea che la responsabilità penale richieda esclusivamente l'adempimento di due requisiti principali: l'elemento oggettivo, l'*actus reus*, e l'elemento

[90]Cfr. Frank, S. J. «Tort Adjudication and the Emergence of Artificial Intelligence Software». In: *Suffolk UL Rev.* 21 (1987), p. 623; Lehman-Wilzig, S. N. «Frankenstein unbound: towards a legal definition of artificial intelligence». In: *Futures* 13.6 (1981), pp. 442–457; Gerstner, M. E. «Liability issues with artificial intelligence software». In: *Santa Clara L. Rev.* 33 (1993), p. 239; Susskind, R. E. «Expert systems in law: A jurisprudential approach to artificial intelligence and legal reasoning». In: *The modern law review* 49.2 (1986), pp. 168–194.

soggettivo della *mens rea*, quali elementi costitutivi del reato [91]. Secondo l'autore, se un AI è in grado di soddisfare entrambi i requisiti, può essere penalmente responsabile per il reato commessso. A questo punto è necessario chiedersi se gli AI sono in grado di soddisfare i requisiti dell'*actus reus* e della *mens rea* o se differiscono dagli esseri umani in questo contesto.

Il primo dei due elementi costitutivi del rato, l'*actus reus*, può facilmente essere attribuito ad un AI, non solo quando l'AI compia un'azione attraverso un movimento fisico[92], controllando per esempio un braccio meccanico o idraulico, ma anche nel caso in cui si l'AI operi al di fuori dell'ambiente fisico. L'adempimento del requisito dell'*actus reus*, non può limitarsi a considerare i soli movimenti muscolari volontari, e in questo caso meccanici, perché si finirebbe per ignorare le fattispecie che non richiedono il compimento di un atto in senso tradizionale, come per esempio i reati informatici; salvo che per questo tipo di reati si voglia far coincidere l'atto fisico con una serie di impulsi elettrici, suggerendo in ogni caso che la definizione tradizionale di atto, inteso come movimento muscolare volontario, è inadeguata e non ammissibile[93]. Per queste ragioni, la visione tradizionale di atto, inteso come movimento muscolare volontario, è ormai sorpassata[94] Nell'ipotesi in cui si tratti di un reato omissivo, la mancata azione dell'AI, in presenza di un obbligo ad agire, sarà sufficiente a soddisfare l'elemento oggettivo della fattispecie[95].

La questione più delicata riguarda invece la possibilità di attribuire ad un AI l'elemento soggettivo del reato.

[91] Cfr. Dressler, J. *Cases and materials on criminal law*. West Group Publishing, 2007, 123 e ss.; Marinucci, G. e Dolcini, E. *Manuale di diritto penale: parte generale*. Giuffrè editore, 2012; Enzo, F. G.-.-.-M. *Diritto penale, parte generale*. 2009, 154 e ss..

[92] Hallevy fa l'esempio di un robot che controlli e attivi il proprio braccio elettrico o idraulico, e attraverso il suo movimento colpisca un essere umano, soddisfando il requisito dell'*actus reus* della fattispecie di aggressione. Hallevy, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit., 187 e ss.; idem, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., 38 e ss.

[93] In questo senso Freitas, P. M., Andrade, F. e Novais, P. «Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible». In: *AI Approaches to the Complexity of Legal Systems*. Springer, 2014, pp. 145–156, 152 e ss.

[94] Si vedano in questo sensoHerring, J. *Criminal law: text, cases, and materials*. Oxford University Press, USA, 2014; e Ormerod, D. *Smith and Hogan: Criminal Law*. 2008, che sottolineano come sia fuorviante, per esempio, affermare che nel reato di diffamazione, l'atto corrisponde al movimento della propria lingua, bocca e corde vocali.

[95] Hallevy, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit., 187 e ss.; idem, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., 39 e ss.

Proviamo a riassumere il ragionamento seguito da Hallevy[96]. Molte delle avanzate capacità cognitive dei moderni sistemi d'intelligenza artificiale non sono rilevanti ai fini della responsabilità penale, così come alcune caratteristiche tipiche degli esseri umani, per esempio la creatività, posseduta, seppur in modo differente, anche dagli animali. Secondo le teorie generali del diritto penale, affinché il requisito della *mens rea* sia soddisfatto, sono necessari elementi quali conoscenza e volontà, declinabile secondo livelli differenti come per esempio, l'intenzione o la colpa. La coscienza è definita come capacità di ricevere e comprendere informazioni sulla realtà di fatto, e la maggior parte dei sistemi d'intelligenza artificiale sono in grado, attraverso i propri recettori di acquisire informazioni, spesso in modo anche più dettagliato rispetto alle capacità di un essere umano. Tali informazioni vengono poi trasferite alle unità centrali di elaborazione che analizzano i dati raccolti. Secondo Hallevy, il processo di analisi e in generale i processi cognitivi dei sistemi d'intelligenza artificiale sono equivalenti ai corrispondenti processi umani[97]. Gli esseri umani attraverso vista, olfatto, tatto e udito ricevono informazioni e, attraverso la loro analisi, ne comprendono il significato. Nei moderni sistemi d'intelligenza artificiale, lo stesso processo di analisi e comprensione, è svolto mediante avanzati algoritmi.

Più in particolare, egli parte dall'analisi e dalla definizione dei due elementi strutturali del delitto intenzionale: coscienza e volontà.

La prima si riferisce alla rappresentazione da parte dell'agente di tutti gli elementi strutturali del fatto tipico, ovvero la condotta e la possibilità che l'evento si verifichi. Per soddisfare i requisiti della *mens rea*, l'agente dovrà rappresentarsi la possibilità che l'evento si verifichi come risultato della condotta. I sistemi d'intelligenza artificiale hanno la capacità, attraverso i propri *device* di percepire tutti gli elementi fattuali percepiti da un essere umano mediante i sensi, e di rappresentarsi una visione complessiva di tali dati e della realtà circostante[98]. Tale rappresentazione avviene attraverso processori capaci di analizzare i dati, integrarli con altri dati e informazioni, trasferirli e agire sulla base di tali risultati. Hallevy considera il caso di un *Security Robot* basato su un sistema d'intelligenza artificiale. Il task assegnato al robot è quello di identificare eventuali intrusi, fermarli e allertare le forse di polizia[99]. Attraverso l'uso dei propri sensori, il robot è in grado di percepire gli elementi fattuali, processarli, identificare gli intrusi

[96] ibid.; idem, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit.

[97] Cfr. Dennett, D. C. «Evolution, error, and intentionality». In: *Contemporary Materialism* (1986), p. 254

[98] Partridge, D. e Wilks, Y. *The foundations of artificial intelligence: A sourcebook*. Cambridge University Press, 1990, 112-118

[99] Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., 53 e ss.

e rappresentarsi correttamente l'intrusione. L'identificazione si basa sugli elementi e i dati percepiti: per esempio l'aspetto fisico o il colore dei vestiti. Il robot fa un calcolo di tipo probabilistico e se tale risultato non è sufficiente a produrre un identificazione accurata, il robot inizia un processo d'identificazione vocale, o ancora potrebbe richiedere al potenziale intruso un codice di identificazione o una password. Una volta acquisite le nuove informazioni, processerà tali dati e li integrerà con le informazioni già presenti in memoria, così da essere certo di non confondere gli intrusi con i membri delle forze di polizia o del personale di sicurezza, ad esempio[100]. Una volta raccolti e processati tutti gli elementi fattuali, il robot avrà le informazioni necessarie per prendere una decisione e agire, sulla base di tale processo.

Il secondo elemento strutturale, la volontà, ha come oggetto l'evento, come risultato della condotta[101]. L'elemento volitivo è soddisfatto qualora l'agente si prefiguri e abbia coscienza della possibilità che l'evento si verifichi come conseguenza naturale della propria condotta e la volontà che questo si verifichi. La maggior parte dei sistemi basati sull'intelligenza artificiale forte sono in grado di apprendere e soppesare correttamente e accuratamente determinati fattori: è il caso, per esempio, dei cosiddetti sistemi di *machine learning*. Inoltre, in alcuni casi il processo di *decision making* dell'AI può essere monitorato. La coscienza dell'AI circa la possibilità che l'evento si verifichi viene monitorata insieme a tutti i fattori presi in considerazione nel processo decisionale. Circa la volontà che l'evento si verifichi, l'autore costruisce un parallelo tra un giocatore di scacchi umano e un giocatore di scacchi basato su un sistema d'intelligenza artificiale[102]. In particolare, se si tratta di un essere umano, siamo soliti affermare che il giocatore ha intenzione di vincere la partita. La volontà è uno stato interno all'agente e non è possibile essere certi della sua esistenza, tuttavia questa può essere dedotta esaminando il corso della sua condotta. Per la prova dell'intenzione, Hallevy fa ricorso alla *Foreseebility rule presumption*, utilizzata da alcuni ordinamenti per provare l'esistenza del dolo. Si presume che l'agente abbia voluto il verificarsi dell'evento se, mentre metteva coscientemente in atto la propria condotta, aveva previsto il verificarsi di tale evento, con una probabilità molto elevata. Il riferimento è rivolto alla figura del *dolus indirectus*. La *ratio* alla base di questa presunzione è che l'agente, pur prevedendo il verificarsi dell'evento con un elevato grado di probabilità, come risultato della propria condotta, decide di

[100] ibid., p. 53.

[101] L'autore rileva anche che i sentimenti, come odio, amore, invidia e gelosia, pur appartenendo esclusivamente agli esseri umani, sono generalmente irrilevanti ai fini della colpevolezza, salvo rare eccezioni come per esempio i crimini razziali o i crimini d'odio, per cui è richiesta una particolare forma di intenzione. Gli AI non potranno essere soggetti a responsabilità penale nei rari casi in cui la fattispecie di reato richieda questa forma di intenzione.

[102] Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit., p. 59.

non modificarla. Da ciò è possibile presumere che l'agente voglia il verificarsi dell'evento[103], come risultato della propria condotta. Così, secondo il parallelo costruito da Hallevy, se possiamo affermare, analizzando la condotta del giocatore umano, che egli ha intenzione di vincere la partita, allora possiamo dire lo stesso del nostro AI. L'AI esamina diverse possibili mosse di gioco, per ognuna calcola la probabilità che si verifichi un certo evento, e, nel caso specifico prevede le possibili mosse dell'avversario. Infine sulla base dei risultati, decide e adotta un comportamento.

Una volta stabilito che i moderni sistemi d'intelligenza artificiale sono in grado di soddisfare i requisiti della *mens rea* dei reati intenzionali, Hallevy costruisce il medesimo schema, seppur adattandolo alle diverse forme di *mens rea*, per i reati colposi e per le ipotesi di *Strict liability*, giungendo alla conclusione che i moderni sistemi d'intelligenza artificiale sono in grado di soddisfare i requisiti dell'elemento oggettivo e dell'elemento soggettivo del reato. Appurato questo, secondo l'autore, non vi sarebbe alcun motivo valido per non sottoporre gli AI alla responsabilità penale, nel caso in cui commettano un reato[104].

Dal momento che un AI è in grado di soddisfare entrambi gli elementi costitutivi del reato, e in questo specifico contesto, non vi è alcuna differenza sostanziale tra esseri umani e AI, anche la maggior parte delle cause di giustificazione e di esclusione dell'imputabilità, previste dagli ordinamenti penali, potranno essere applicata a robot e sistemi d'intelligenza artificiale, qualora se ne verifichino i presupposti[105].

Nel sistema costruito da Hallevy, la responsabilità penale di un AI non esclude e non sostituisce la responsabilità di utenti e programmatori. Essa si combina e si aggiunge a quella degli esseri umani e, tuttavia, la responsabilità di un AI è indipendente rispetto a quella del programmatore o dell'utente. Inoltre, i tre modelli descritti, non sono alternativi ma possono essere applicati in combinazione tra loro, così da costruire un quadro completo della responsabilità penale nei casi in cui siano coinvolti AI. In conclusione, la responsabilità penale di un AI secondo il modello della responsabilità diretta non è diversa dalla responsabilità penale di un essere umano, a parità di condizioni; essa si basa sugli stessi elementi e può essere analizzata nello stesso modo.

[103] Hallevy, G. *Liability for Crimes Involving Artificial Intelligence Systems*. Springer, 2014, 96 e ss.

[104] idem, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit., 191 e ss.; idem, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit.

[105] Per un approfondimento sull'applicabilità delle cause di giustificazione si vedaibid., 120-155. L'autore fa l'esempio di un AI che opera nelle forze di polizia locale e a cui viene dato l'ordine di arrestare una persona illegalmente. Se l'ordine non è manifestamente contrario alla legge, l'esecutore dell'ordine non sarà penalmente perseguibile.

3.6.4 La punibilità degli AI

Una volta elaborati i modelli di responsabilità applicabili agli AI, e appurata la possibilità che siano ritenuti penalmente responsabili, Hallevy costruisce un sistema di pene applicabili, sulla base dei medesimi fondamenti teorici che il diritto penale prevede per gli esseri umani. Secondo l'autore, le pene previste dagli ordinamenti penali possono essere applicate, con alcuni adattamenti agli AI. Il vantaggio più significativo di tale adattamento è che il significato della pena rimane identico sia che si tratti di esseri umani, sia che si tratti di AI.

Così, Hallevy costruisce un sistema di equivalenze, per esempio tra pena capitale, che priva il condannato della propria vita, rendendolo ovviamente incapace di commettere ulteriori reati in futuro, e la cancellazione del programma d'intelligenza artificiale che controlla l'AI. Una volta terminato il programma, l'AI sarà a sua volta incapace di commettere ulteriori reati. La cancellazione del programma, termina l'esistenza autonoma e indipendente dell'agente, nello stesso modo in cui la pena di morte pone fine alla vita degli esseri umani. Ancora, nei sistemi giuridici occidentali, il significato della carcerazione per gli esseri umani risiede nella forte limitazione o privazione della libertà personale. La libertà di un AI, include la libertà di agire nella propria area di pertinenza, così una pena equivalente consiste nel mettere fuori uso l'agente per un periodo di tempo determinato, limitando la sua libertà di azione. La condanna ai servizi sociali ha una funzione risarcitoria rispetto al danno arrecato alla collettività e allo stesso modo un AI può essere impiegato per svolgere lavori di pubblica utilità, nelle aree più disparate. In conclusione, secondo Hallevy, le pene più comuni previste dagli ordinamenti giuridici nazionali sono applicabili agli AI, con gli opportuni adattamenti, senza che tali sanzioni perdano la propria funzione e natura. L'autore sostiene ancora che gli AI non hanno un'anima, e alcuni di essi non hanno né anima né corpo, e tuttavia non vi sarebbe alcuna differenza giuridica sostanziale tra la *ratio* alla base della responsabilità penale imposta alle società e la *ratio* alla base di una responsabilità penale imposta agli AI.

Alcune prime considerazioni

Dopo aver analizzato i modelli elaborati dalla dottrina, per regolare le ipotesi in cui un AI ponga in essere una condotta astrattamente riconducibile a una fattispecie criminosa, secondo i criteri del diritto penale, possiamo fare alcune prime considerazioni. La responsabilità penale delle persone giuridiche ha certamente costituito una svolta innovativa, nell'area del diritto penale, e i modelli utilizzati forniscono preziosi indizi, per costruire e approfondire un quadro dogmatico plausibile per la responsabilità penale degli agenti artificiali. Il diritto penale,

contrariamente a quanto possa immaginarsi a un primo sguardo, dimostra un certo grado di flessibilità qualora, per ragioni di politica criminale, tale flessibilità risulti necessaria e a condizione che certe premesse dogmatiche siano soddisfatte. Nel prossimo capitolo, proveremo a verificare più nel dettaglio se un AI sia potenzialmente in grado di soddisfare i requisiti dell'*actus reus* e della *mens rea*, analizzando il funzionamento e le tecniche utilizzate per costruire alcuni sistemi d'intelligenza artificiale. Procederemo all'analisi di un caso concreto che, seppur non attinente alla sfera dell'e-health, a chi scrive, appare significativo del fenomeno, costruendo scenari e verificando possibili responsabilità di agenti autonomi intelligenti, produttori e utilizzatori del sistema, secondo i modelli di responsabilità qui esposti.

Inoltre, qualora riuscissimo a dimostrare la capacità di un AI di soddisfare i requisiti dell'elemento oggettivo e soggettivo del reato, dovremmo chiederci non tanto se sia possibile applicare il diritto penale agli AI, quanto piuttosto se dovremmo farlo, a quale scopo, e in che modo. Quale utilità avrebbe applicare sanzioni agli AI, nel modo in cui sono descritte da Hallevy? Esistono altri modi per regolare le ipotesi in cui un AI commetta dei reati? Ci sembra possibile fin da ora anticipare che, prima di arrivare ad applicare sanzioni penali agli AI, seppur con alcuni adattamenti, allo stato attuale, vi siano modelli di regolazione informatico-giuridica, utili a prevenire e a controllare l'emergenza di fenomeni criminali riconducibili agli AI.

Casi di studio, scenari e modelli di regolazione

4.1 Una breve introduzione

Sulla base dei modelli presentati nella sezione 3.6, le prime sezioni di questo capitolo sono dedicate a verificare se un agente autonomo intelligente sia concretamente in grado di soddisfare i requisiti dell'*actus reus* e della *mens rea*, quali condizioni necessarie all'attribuzione di responsabilità secondo il diritto penale. In particolare, si è scelto di analizzare la capacità di un AI di soddisfare i requisiti del dolo, poiché esso rappresenta la più alta forma di intenzione. Qualora un AI sia capace di sviluppare forme di intenzione e volontà, si cercherà di verificare quali soluzioni siano disponibili per ridurre le ipotesi che un agente autonomo intelligente ponga in essere condotte criminose.

Nella sezione 4.5 sarà esaminato un caso concreto sulla base dei modelli discussi nel capitolo precedente. Il caso, seppur non attinente alla sfera dell'e-health, appare a chi scrive esemplificativo del fenomeno. Non sempre, infatti, è possibile attribuire la responsabilità per il reato compiuto da un AI a programmatori o utilizzatori, con il rischio che si verifichi un vuoto di responsabilità davanti a tali fenomeni e che questi diventino incontrollabili.

A tal fine saranno presi in considerazione alcuni dei modelli informatico-giuridici, utili a regolare i casi in cui un AI violi una norma, ponendo in essere una condotta criminosa che, se compiuta da un essere umano, sarebbe punibile secondo le leggi del diritto penale.

Infine, torneremo all'area dell'e-health, provando a costruire uno scenario che prenda in considerazione un sistema d'intelligenza artificiale utilizzato come supporto alla diagnostica e ad indagare i rapporti tra uomo e macchina, eventuali aspetti problematici e a delineare infine profili di responsabilità, nel caso in cui, da una diagnosi errata del sistema derivino danni al paziente. Cercheremo anche di verificare l'applicabilità del principio di affidamento, di cui si è trattato nella prima parte del capitolo precedente, quale limite alla responsabilità penale del medico.

4.2 Requisiti per l'attribuzione della responsabilità penale

Tradizionalmente si definisce reato "ogni fatto cui la legge ricollega una sanzione penale". Tale definizione di natura formale indica i fatti che costituiscono reato per un determinato ordinamento positivo, in funzione delle conseguenze giuridiche che il legislatore riconnette ai fatti in questione. Nella struttura del reato si distinguono due specie di elementi: essenziali, e dunque indispensabili per l'esistenza del reato, e accidentali, la cui presenza non influisce sull'esistenza del reato ma solo sull'entità della pena. Questi ultimi si distinguono a loro volta in circostanze attenuanti e aggravanti. Gli elementi essenziali per l'imputazione della responsabilità penale sono l'elemento oggettivo del reato, l'*Actus Reus*, e l'elemento soggettivo del reato, la *mens rea*. La struttura di base dell'elemento materiale è la medesima per tutti i tipi di violazione, sia essa intenzionale, colposa o preterintenzionale. Essa consiste di tre componenti principali: "la condotta", eventuali presupposti e circostanze che la caratterizzano, e l'evento lesivo legato alla condotta da un rapporto causale. I presupposti e le circostanze hanno carattere solo eventuale. La condotta, a sua volta, può essere espressa attraverso azioni o omissioni. L'elemento soggettivo, richiamato anche dall'articolo 27 della Costituzione, è costituito dall'atteggiamento psicologico del soggetto agente, che può essere declinato nelle forme del dolo, della colpa o della preterintenzione. Una volta riassunti brevemente gli elementi necessari all'attribuzione della responsabilità penale, per condotte astrattamente riconducibili a fattispecie criminose, nelle sezioni che seguono cercheremo di verificare se un agente autonomo intelligente è in grado di soddisfare tali requisiti.

4.3 *Actus Reus* e sistemi d'intelligenza artificiale

Sul terreno del reato commissivo, la condotta criminosa assume la forma di un'azione in grado di modificare la realtà mentre, il reato omissivo identifica la mancanza di una determinata azione che per legge si aveva l'obbligo di compiere.

Tradizionalmente, la dottrina penalistica ha inteso il concetto di azione in senso stretto come "il movimento del corpo idoneo ad offendere l'interesse protetto dalla norma o l'interesse statuale perseguito dal legislatore attraverso l'incriminazione"[1]. Secondo questa definizione un agente intelligente che opera nel

[1]Così,Mantovani, F. «Diritto penale». In: *Parte generale, Padova: CEDAM* (2001). Cfr. anche Fiandaca Giovanni e Musco, E. «Diritto penale. Parte generale». In: *Bologna, Zanichelli* (2004).

mondo fisico, come un robot, è certamente in grado di compiere atti che soddisfino i requisiti della condotta. Esso è infatti capace di governare le proprie parti in movimento, come per esempio un braccio meccanico o idraulico. Tuttavia, esistono agenti che non operano nel mondo fisico, ma in quello virtuale, come per esempio gli agenti software (AS). Le loro azioni risulterebbero giuridicamente irrilevanti sotto il profilo materiale.

È già stato evidenziato come l'adempimento del requisito dell'*actus reus*, non possa limitarsi a considerare i soli movimenti muscolari, e in questo caso meccanici, perché si finirebbe per ignorare quelle fattispecie che non richiedono il compimento di un atto in senso tradizionale, come appunto i reati informatici; salvo che per questo tipo di reati si voglia far coincidere l'atto fisico con una serie di impulsi elettrici, suggerendo in ogni caso che la definizione tradizionale di atto, inteso come movimento muscolare, è inadeguata e non ammissibile. Alcuni autori hanno sottolineato come tale definizione risulti inadeguata anche rispetto a reati come la diffamazione, ove è fuorviante ritenere che, in tale fattispecie, l'atto giuridicamente rilevante coincida con il movimento della propria lingua, bocca e corde vocali[2]. Per questi motivi, la visione tradizionale secondo cui l'azione equivale ad un movimento corporeo è stata ritenuta, negli ultimi anni, obsoleta[3].

Accettando una visione più ampia del concetto di azione, anche un agente che operi unicamente nel mondo virtuale come un AS, qualora per esempio la sua condotta integri la fattispecie di accesso abusivo e danneggiamento di un sistema informatico, soddisferebbe il requisito materiale della condotta.

L'azione punibile deve essere accompagnata dal requisito della coscienza e volontà, anche se poi tale requisito assume significato diverso in funzione della natura dolosa o colposa del reato commesso. Tuttavia, nell'analisi dell'*actus reus* non rileva se l'azione sia voluta o meno dal soggetto agente, la volontà è infatti parte dell'elemento soggettivo del reato. Gli atti involontari e i riflessi condizionati possono quindi considerarsi azioni[4]. Secondo questa definizione, un agente intelligente è in grado di compiere atti che soddisfino il requisito della condotta secondo le norme penali. Questo è vero non solo quando l'azione è il risultato di calcoli interni condotti dall'agente intelligente, ma anche quando questo esegue le istruzioni di un operatore umano remoto. Per le ipotesi di condotte omissive, affinché i requisiti della condotta siano soddisfatti, sarà sufficiente un'inazione

[2] Tra gli altri Ormerod, *Smith and Hogan: Criminal Law*, cit.

[3] In questo senso, Freitas, Andrade e Novais, «Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible», cit., 152 e ss.; Herring, *Criminal law: text, cases, and materials*, cit.

[4] Nel caso in cui venga a mancare il potere di signoria dell'agente sulla propria azione, come nell'ipotesi di costringimento fisico, i requisiti materiali della condotta saranno in ogni caso soddisfatti, e tuttavia l'agente non sarà punibile. Il costringimento fisico è una specificazione della forza maggiore e ha quindi valore di esimente.

da parte dell'AI a fronte di un obbligo giuridico di agire. In conclusione, gli agenti intelligenti sono in grado di porre in essere una condotta che soddisfi i requisiti dell'*actus reus*. Proviamo ora a verificare se un AI sia astrattamente in grado di soddisfare la componente soggettiva del reato.

4.4 *Mens Rea* e sistemi d'intelligenza artificiale

La società contemporanea non ha interesse a punire eventi accidentali, o casuali e la sola condotta non è sufficiente per l'imputazione della responsabilità penale. Assumono valore solo gli eventi che sono il risultato di un comportamento attivo o omissivo colpevole. La *mens rea* riassume le condizioni psicologiche che consentono l'imputazione del fatto di reato al suo autore. Nel giudizio di colpevolezza rientrano la valutazione del legame psicologico o, comunque, del rapporto di appartenenza tra fatto e autore e la valutazione delle circostanze, di natura personale e non, che incidono sulle capacità di autodeterminazione del soggetto. La Mens rea copre la relazione interna e soggettiva tra il soggetto che pone in essere la condotta e l'elemento materiale del reato, che include, come già ricordato, la condotta, le circostanze e l'evento[5]. Il rimprovero di colpevolezza presuppone la possibilità di agire diversamente da parte del soggetto cui viene attribuito il fatto e, nella maggior parte degli ordinamenti, può essere graduato secondo i parametri del dolo e della colpa.

La nozione di dolo è incentrata su tre elementi: coscienza, volontà ed evento dannoso o pericoloso. I primi due elementi sono di natura strutturale, poiché indicano le componenti che caratterizzano il dolo come fenomeno psicologico. Il terzo elemento attiene invece all'oggetto che deve riflettersi nella rappresentazione e nella volizione.

La colpa è un criterio di imputazione sussidiaria rispetto al dolo, perché la condotta antigiuridica che dà luogo al delitto colposo è punibile nei soli casi espressamente previsti dalla legge. Per la configurazione del delitto colposo, è necessario che la condotta sia cosciente e volontaria, che l'evento, salvo determinati casi, non sia voluto e che il fatto sia imputabile all'agente per negligenza, imprudenza o imperizia.

Per soddisfare i criteri soggettivi di attribuzione della responsabilità penale non è necessario che all'agente siano ascrivibili tutte le capacità umane. Nel diritto penale, i requisiti della *mens rea* sono soddisfatti da abilità di gran lunga inferiori rispetto a quelle richieste affinché si possa paragonare una macchina ad un essere umano; questo è lo standard che si applica a tutti gli agenti, umani e non umani, come per esempio le società, purché dotati di personalità giuridica

[5] Fiandaca, «Diritto penale. Parte generale», cit., 275 e ss.

[6]. La questione essenziale è se solo gli esseri umani siano in grado di soddisfare i requisiti della *mens rea* secondo i parametri del diritto penale, o se tali requisiti possano essere soddisfatti anche da alcuni sistemi d'intelligenza artificiale particolarmente avanzati. Come anticipato nella parte introduttiva di questo capitolo, si è scelto di prendere in considerazione il dolo, poiché, per il diritto penale, esso rappresenta il grado più alto della volontà.

4.4.1 Mens rea, AI e reati intenzionali

Il dolo è strutturalmente caratterizzato da due componenti psicologiche: rappresentazione (o coscienza o conoscenza o previsione) e volontà. La tesi che assegna al dolo una duplice dimensione, cognitiva e volitiva, è accettata nella maggior parte degli ordinamenti giuridici.

La componente conoscitiva del dolo si atteggia diversamente a seconda che abbia come punto di riferimento elementi descrittivi della fattispecie di reato o elementi normativi. Nel primo caso l'elemento cognitivo si riferisce alla capacità di rappresentarsi correttamente la realtà e gli elementi del mondo esterno, così come appaiono nella loro dimensione naturalistica (ad esempio, uomo, morte, etc.). Qualora la componente conoscitiva, si riferisca, invece, ad elementi normativi (ad esempio altruità della cosa, documento etc.), per l'esistenza del dolo non sarà sufficiente che l'agente sia a conoscenza di meri dati di fatto. Il dolo non è infatti semplice rappresentazione ma anche volontà consapevole.

L'elemento cognitivo nei sistemi d'intelligenza artificiale

L'elemento cognitivo del dolo, consiste nella capacità di percepire e comprendere gli elementi del mondo esterno. Affinché un essere umano sia cosciente della realtà esterna devono essere soddisfatte due condizioni: (1) egli deve essere in grado di acquisire dati e informazioni (attraverso i sensi) e (2) di crearsi una rappresentazione generale e rilevante della realtà, sulla base delle informazioni acquisite. Proviamo ora a valutare se, almeno in linea teorica, un sistema d'intelligenza artificiale sia in grado di soddisfare queste due condizioni.

Endsley definisce il concetto di stato di coscienza, cd. *Situation Awareness* (SA), come la "percezione di elementi presenti nell'ambiente, in uno specifico volume di spazio e tempo, la comprensione del loro significato e la stima e la previsione dei loro stati futuri"[7]. In particolare, tale definizione di SA è strutturata

[6]Nello stesso senso Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit.; Hallevy, G. *The Criminal Liability of Artificial Intelligence Entities*. Akron Intellectual Property, 2010, il quale distingue tra due tipologie di agenti autonomi intelligenti, *Machina Sapiens* e *Machina Sapiens Criminalis*.

[7]Endsley Mica R e Garald, D. «Theoretical underpinnings of situation awareness: A critical review». In: *Situation awareness analysis and measurement* (2000), pp. 3–32.

su tre livelli:

- Livello 1: percezione degli elementi che caratterizzano l'ambiente;
- Livello 2: comprensione di tali elementi e della realtà in uno stato presente;
- Livello 3: previsione degli stati futuri.

Il primo livello consiste nella percezione dello stato, degli attributi e delle dinamiche degli elementi rilevanti dell'ambiente. I requisiti della *Situation Awareness* possono variare a seconda del singolo dominio di applicazione. Ad esempio, in ambito militare un ufficiale di comando deve essere in grado di individuare e distinguere l'obiettivo militare dai civili e dagli alleati. Così, un controllore di volo e l'autista di un autobus, in base al dominio in cui operano, hanno bisogno di informazioni differenti affinché tale condizione sia soddisfatta. L'acquisizione dei dati può avvenire attraverso un numero più o meno elevato di canali fisici, virtuali, o attraverso una combinazione di entrambi.

Il secondo livello consiste nella comprensione dei dati e degli elementi acquisiti e nella capacità di relazionarli agli obiettivi rilevanti. La comprensione è basata su una sintesi di tutti i dati e di tutti gli elementi percepiti, nel primo livello, e sul confronto di tali elementi con gli obiettivi finali, che potremmo indicare come *desiderata*. La nozione di comprensione è quindi basata su una continua estrazione di informazioni dall'ambiente circostante, e dall'integrazione di frammenti di dati con la conoscenza acquisita in precedenza, utile a formare una visione coerente della realtà nello stato presente.

Questo tipo di conoscenza, che potremmo definire aggregata, può essere utilizzata in un momento successivo, per dirigere la percezione su determinati elementi e anticipare stati futuri ed eventi. Tale previsione, costituisce il terzo livello della *Situation Awareness*. Se un essere umano conosce gli elementi che caratterizzano l'ambiente, il loro significato in relazione all'obiettivo che intende raggiungere, ed è in grado di stimare come tali elementi evolveranno, almeno nel breve periodo, avrà soddisfatto i requisiti necessari a determinare lo stato di coscienza[8].

- **Livello 1: percezione e sistemi d'intelligenza artificiale**

  Proviamo a valutare la possibilità che un sistema d'intelligenza artificiale soddisfi il primo livello dello stato di coscienza. Il primo livello del processo di coscienza consiste nell'abilità di osservare lo spazio circostante, percepire l'ambiente e formalizzare tale percezione, affinché il nostro AI sia in grado di comprendere le informazioni acquisite e fare previsioni sui

[8] ibid.

possibili stati futuri dell'ambiente in cui opera. Come più volte rilevato, un AI può esistere e agire nell'ambiente fisico - e in questo caso si tratterà di un robot- oppure in un ambiente virtuale.

Nel caso dei robot, la percezione può essere implementata nell'hardware, ad esempio, attraverso una videocamera o un sensore laser per acquisire informazioni sulla geometria dell'ambiente circostante. In questo caso, il ruolo principale è giocato dai sistemi utilizzati per catturare ed elaborare i dati ambientali, che operano in modo analogo al funzionamento dei cinque sensi negli esseri umani. Esattamente come gli occhi percepiscono la luce e trasmettono dati alla corteccia visiva, le telecamere del nostro agente software assorbono la luce per trasferire i dati agli apparati di elaborazione, cd. processori. Inoltre, le tecnologie più recenti sono in grado di acquisire informazioni e dati in modo molto più avanzato e preciso di quanto siano in grado di fare gli esseri umani attraverso i propri sensi. Nel nostro esempio, le telecamere possono assorbire la luce anche qualora la gamma di frequenze non sia visibile -ad esempio la radiazione infrarossa- dall'occhio umano. In quest'ottica, un sistema intelligente è in grado di soddisfare il requisito del primo livello della *situation awareness* in modo anche più efficiente di quanto non sia possibile per un essere umano.

Per gli AI che agiscono e operano in un ambiente virtuale, la percezione avviene mediante la tracciatura di attività, messaggi, e informazioni. Questa funzione può essere implementata nel software. Nel 2004 Weens, Steegmans e Holovoet[9] hanno proposto un modello generale di percezione per gli AI che operano in ambienti virtuali. Il modello è indipendente da qualsiasi dominio applicativo o dalle specifiche topologie dell'ambiente in cui gli agenti si trovano ad operare. La figura 4.1 ne mostra una rappresentazione sintetica.

Il modello si compone di tre moduli funzionli: il primo è costituito da sensori per la percezione dei dati (*sensing*), il secondo si occupa dell'interpretazione dei dati (*interpreting*) e il terzo di filtrarli (*filtering*).

Nel primo modulo, i sensori creano una corrispondenza biunivoca tra lo stato dell'ambiente e la sua rappresentazione. Tale corrispondenza tra stato e rappresentazione dipende dall'insieme dei *foci* e delle leggi percettive scelte. La selezione del *focus* consente ad un agente software di dirigere la propria capacità percettiva, e di acquisire dati sull'ambiente selezionando le informazioni rilevanti. La possibilità di dirigere gli stati percettivi, e di creare una rappresentazione ridotta della realtà, permette di tarare un

[9] Weyns, D. et al. «Environments for multi-agent systems». In: *First International Workshop (E4MAS 2004), New York, NY, July 19, 2004, Revised Selected Papers*. Vol. 3374. 2004, 23 e ss.

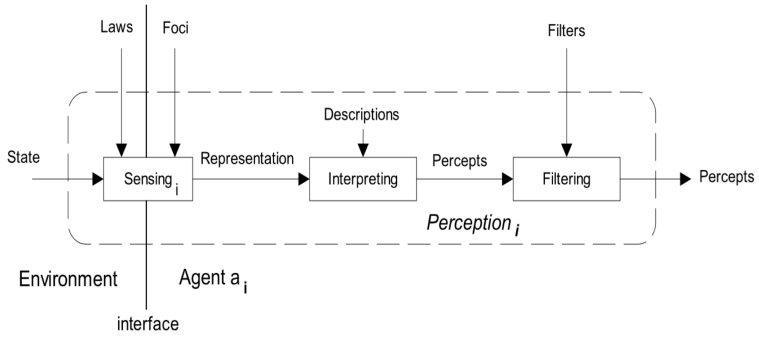


FIGURA 4.1: Model for active perception, Weens, Steegmans e Holovoet

agente software in base al dominio di applicazione. Le leggi percettive sono quindi uno strumento nelle mani del designer per creare e modellare vincoli alla percezione. Una legge percettiva potrebbe, per esempio, specificare che l'area al di là di un certo ostacolo sia non utile allo scopo corrente dell'agente. Ancora, per ragioni di efficienza, si potrebbero introdurre limiti nominali alla percezione, in modo da vincolare e ridurre l'insieme di informazioni che l'agente software dovrà poi processare. Questo primo modulo supporta anche la percezione simultanea da parte di più sensori e gli agenti possono selezionare ed attivare più *foci* contemporaneamente.

Il secondo modulo riguarda la fase di interpretazione dei dati acquisiti. A tal fine, gli agenti software usano alcune descrizioni, rappresentazioni sintetiche di concetti, in grado di mappare le percezioni e le rappresentazioni in una forma comprensibile dall'agente.

Infine, nel terzo modulo, utilizzando un insieme coerente di filtri, un AS è in grado di selezionare un sottoinsieme di dati, sulla base di specifici criteri di scelta. Ognuno di questi filtri impone condizioni sugli elementi oggetto della percezione, così da determinare se tali elementi siano di interesse o meno.

Un osservazione importante riguarda la dinamicità degli stati percettivi degli agenti. In un sistema aperto, è infatti necessario che i componenti di un sistema percettivo siano in grado di mutare dinamicamente, per adattarsi alle indicazioni dei progettisti o alle necessità dell'agente stesso. Secondo quanto indicato dagli autori, all'interno del modello, l'agente può modificare l'insieme dei *foci* selezionati, e i relativi filtri, in modo dinamico e sulla base del compito che l'agente deve portare a termine.

D'altro canto, le leggi percettive sono definite a priori dal progettista e, per questo motivo, difficilmente riescono a gestire correttamente cambiamenti non previsti nell'ambiente circostante. In questi casi si è in presenza di un cosiddetto ambiente debolmente strutturato. Tuttavia, per poter gestire cambiamenti imprevisti, il modello può essere esteso con opportune infrastrutture. Tali infrastrutture consentono sia di adattare il set di leggi percettive ai cambiamenti che si verificano nell'ambiente, sia di modificare le leggi dinamicamente[10].

- **Livello 2: comprensione e sistemi d'intelligenza artificiale**

  Riguardo al secondo step del processo di coscienza, la maggior parte delle tecnologie basate sull'intelligenza artificiale, così come i robot e i computer, sono dotati di un "cervello artificiale" incorporato nell'hardware (processori, memoria, e così via). Esattamente come gli esseri umani creano immagini rilevanti della realtà esterna, attraverso l'analisi dei dati che caratterizzano l'ambiente - mediante la raccolta, il trasferimento e l'integrazione delle informazioni - gli AI sono in grado di creare un'immagine generale e rilevante del mondo che li circonda, attraverso l'assorbimento dei dati, l'integrazione di tutte le informazioni raccolte, il loro trasferimento alle unità di elaborazione e processazione, e la loro comparazione con schemi ricorrenti presenti in memoria, cd. *pattern*.

- **Livello 3: previsione degli stati futuri e sistemi d'intelligenza artificiale**

  Guardiamo infine, al terzo livello del processo di *Situation Awareness*: la capacità di compiere previsioni sugli stati futuri. Un AI è in grado di calcolare le probabilità che certi esiti determinati si verifichino come conseguenza di azioni fra loro alternative, di utilizzare queste informazioni come base di possibili scelte, e infine adottare una scelta conforme e ragionevole rispetto agli obiettivi preposti.

Una prospettiva computazionale dello stato di *Situation Awareness*

Da un punto di vista computazionale, lo stato di coscienza consiste (a) in un processo guidato di aggregazione di percezioni e dati e (b) nella costruzione di relazioni concettuali e credenze, che riflettono lo stato dell'ambiente e della realtà in cui l'agente opera. Per esempio, i computer che giocano a scacchi - si pensi al famoso caso di Deep Blue dell'IBM, che alla fine degli anni novanta batté a scacchi l'allora campione del mondo Garry Kasparov - sono in grado di

[10] Weyns, D., Steegmans, E. e Holvoet, T. «Towards active perception in situated multi-agent systems». In: *Applied Artificial Intelligence* 18.9-10 (2004), pp. 867–883, passim.

analizzare lo stato corrente del gioco a seconda della posizione dei pezzi sulla scacchiera. Il sistema compie un'analisi delle possibili mosse di gioco e, per ciascuna opzione, analizza possibili contromisure in risposta all'avversario. Per ogni contromossa compie la medesima analisi, utilizzando vari possibili metodi di restringimento delle strategie di ricerca. Questo tipo di processo identifica il merito di una specifica sequenza di mosse, individuando il cosiddetto albero di gioco -ovvero l'insieme di mosse e contromosse per un certo frangente di gioco- più promettente[11].

Negli ultimi anni, uno degli approcci e delle tecniche più utilizzate nella progettazione di sistemi d'intelligenza artificiale è stato il modello probabilistico. Esso gioca un ruolo centrale nell'analisi scientifica dei dati, nel *machine learning*, nella robotica, nelle scienze cognitive e più in generale in tutti i campi dell'intelligenza artificiale. Il modello probabilistico fornisce una struttura per la comprensione dei meccanismi alla base dell'apprendimento. Per tale ragione, è diventato uno dei principali approcci teorici e pratici per la progettazione di sistemi in grado di apprendere dall'esperienza.

Un sistema di rappresentazione probabilistico descrive in che modo rappresentare e manipolare gli stati di incertezza relativi a modelli di realtà, e fare previsioni sulla base di questi modelli. L'idea chiave è che una macchina possa utilizzare tali modelli per fare predizioni su dati e stati futuri, e adottare decisioni che siano razionali alla luce di tali predizioni. Ovviamente, l'incertezza gioca un ruolo fondamentale in questo processo. È incerto quale sia il modello appropriato, poiché i dati osservati possono essere coerenti con modelli diversi e, allo stesso modo, possono essere incerte le predizioni relative ai risultati futuri di certe azioni, dato un certo modello[12].

Alcuni modelli e metodi per gestire l'incertezza non rappresentano esplicitamente le probabilità associate ai fatti e alle predizioni. Ad esempio, le reti neurali, utilizzate in molte applicazioni di cd. *pattern recognition* e altri sistemi, caratterizzati dalla disponibilità di grandi quantità di dati, non rappresentano esplicitamente l'incertezza nella struttura o nei parametri delle reti neurali, come per esempio, i sistemi di riconoscimento vocale[13] e delle immagini[14], o di

[11] Hsu, F.-H. *Behind Deep Blue: Building the computer that defeated the world chess champion*. Princeton University Press, 2002.

[12] Per una panoramica delle principali aree di ricerca relative al modello probabilistico, si vedaGhahramani, Z. «Probabilistic machine learning and artificial intelligence». In: *Nature* 521.7553 (2015), pp. 452–459.

[13] Hinton, G. et al. «Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups». In: *Signal Processing Magazine, IEEE* 29.6 (2012), pp. 82–97.

[14] Sermanet, P. et al. «Overfeat: Integrated recognition, localization and detection using convolutional networks». In: *Proc. International Conference on Learning Representations* (2014).

predizione di parole all'interno di un testo[15]. Tali sistemi non producono una diretta rappresentazione delle probabilità coinvolte nella processazione dei dati.

In altri sistemi, l'incertezza è un ingrediente chiave. Ad esempio, la decisione può dipendere dal grado di incertezza dei dati o delle predizioni. Esempi tipici riguardano la rilevazione di pedoni nelle immagini riprese da veicoli autonomi o la classificazione in sottotipi di *pattern* genici, di pazienti affetti da leucemia, sulla base dei risultati clinici. Tuttavia, gli scopi e le applicazioni di *machine learning* sono molto più ampi rispetto alla semplice classificazione di *pattern* e alla mappatura di *task*. Essi possono includere attività di ottimizzazione e *decision making*, sintetizzazione di dati ed estrazione automatica dai dati di modelli interpretabili.

Negli esseri umani, la coscienza prevede anche la capacità di inferire elementi dello stato corrente, non direttamente percepiti. Tale inferenza avviene tramite processi mentali come, per esempio, l'abduzione. Questa capacità di inferenza, può essere presente nei sistemi d'intelligenza artificiale che utilizzano tecniche di *machine-learning*. Tali tecniche consentono, infatti, di inferire dati mancanti o latenti, rispetto ai dati osservati. Tutte le attività di *machine learning* consistono nel fare inferenze circa i dati mancanti o latenti dai dati osservati. Nell'esempio citato di classificare le persone con leucemia in uno dei quattro sottotipi principali di questa malattia, sulla base di modelli di espressione genica misurati di ogni persona, i dati osservati sono coppie di modelli di espressione genica e sottotipi etichettati, e non osservato o dati mancanti da dedurre sono i sottotipi di nuovi pazienti. Per fare inferenze circa i dati non osservabili dai dati osservati, il sistema di apprendimento ha bisogno di fare alcune ipotesi; nel loro insieme, queste ipotesi costituiscono un modello, e un modello possono essere considerati ben definito se può fare previsioni o le previsioni sui dati non osservabili essere stati addestrati su dati osservati. Nell'esempio citato, relativo alla classificazione dei pazienti affetti da leucemia, in una delle quattro sottoclassi principali di questa malattia, i dati osservati sono le coppie di genotipo-fenotipo -l'espressione del gene in un pattern classificato e riconoscibile della malattia- appartenenti ai pazienti a cui è già stata fatta una diagnosi, mentre i dati mancanti sono le sotto-classi dei nuovi pazienti da diagnosticare.

L'elaborazione inferenziale dei dati non osservati attraverso i dati osservati è possibile mediante ipotesi e assunzioni elaborate dal sistema; esse costituiscono collettivamente un modello della relazione input-output. Ogni modello può essere considerato valido nella misura in cui è in grado di elaborare predizioni corrette sui dati non osservati[16].

[15] Bengio, Y. et al. «Neural probabilistic language models». In: *Innovations in Machine Learning*. Springer, 2006, pp. 137–186.

[16] Ghahramani, «Probabilistic machine learning and artificial intelligence», cit.

Alcune conclusioni preliminari

Dopo aver analizzato e compreso il significato di cognizione secondo il diritto penale ed esposto in modo dettagliato il modello di *Situation Awareness* elaborato da Endsley e i livelli in cui esso è strutturato, abbiamo cercato di comprendere come alcune tecniche d'intelligenza artificiale siano in grado di dotare un AI di stati cognitivi, in grado di soddisfare il concetto di coscienza e rappresentazione, all'interno di sistemi complessi.

Sulla base di quanto visto fin'ora, è stato dimostrato che i sistemi d'intelligenza artificiale sono in grado di soddisfare i requisiti dell'elemento cognitivo, e in particolare della rappresentazione o coscienza, attraverso l'assorbimento, la raccolta e l'inferenza di dati e informazioni che caratterizzano l'ambiente in cui l'agente opera, l'elaborazione di tali informazioni e la loro integrazione con informazioni acquisite in precedenza o presenti *ab origine* all'interno del sistema, e la creazione di un'immagine generale e complessiva della realtà. In conclusione, un AI è in grado di raggiungere lo stato di coscienza acquisendo dati e informazioni e rappresentandosi un'immagine generale della realtà, passando attraverso le tre fasi di percezione, comprensione, e predizione.

Nella maggior parte dei casi, i processi che portano allo stato di coscienza, possono essere accuratamente monitorati e registrati dal sistema d'intelligenza artificiale stesso. È quindi possibile dimostrare che un agente è a conoscenza di informazioni e stati di fatto particolari. La figura 4.2 offre una panoramica dei requisiti dell'elemento cognitivo in relazione alle capacità di un AI.

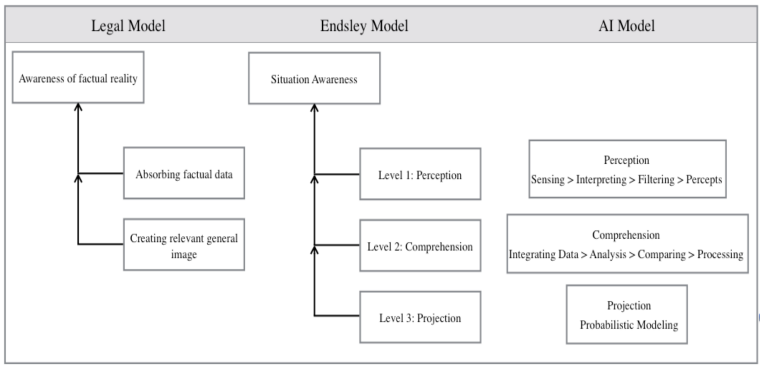


FIGURA 4.2: L'elemento cognitivo nei sistemi d'intelligenza artificiale

L'elemento volitivo nei sistemi d'intelligenza artificiale

L'elemento volitivo della *mens rea* può essere declinato secondo vari livelli d'intensità. Come già anticipato, in questa sede si è scelto di valutare la possibilità che un agente autonomo intelligente sia in grado di soddisfare i requisiti del più alto grado di volontà, secondo il diritto penale: l'intenzione.

Il processo deliberativo che conduce ad un'azione dolosa o intenzionale è un processo interno alla mente umana che ha come oggetto il verificarsi di uno stato futuro. Come già evidenziato nella prima parte della sezione 4.4.1, la volontà investe l'azione e il fatto tipico colto nella sua unità significativa. Essa abbraccia tutti gli elementi costitutivi del fatto tipico: condotta, circostanze ed evento. È bene precisare che l'evento non va inteso in senso strettamente naturalistico o in senso giuridico – come lesione o messa in pericolo del bene protetto - poiché nel primo caso rimarrebbero esclusi i cd. reati di mera condotta, e nel secondo, perché nei reati di creazione legislativa[17], la consapevolezza del carattere lesivo del fatto non può prescindere dalla conoscenza effettiva del divieto penale, contravvenendo al principio secondo cui *ingnorantia legis non excusat*. Oggetto del dolo è quindi il fatto tipico[18]. Se manca la volontà di realizzare il fatto di reato, non bastano a integrare il dolo desideri, speranze, proponimenti, tendenze, inclinazioni e simili. L'imputazione a titolo di dolo, in omaggio al principio *cogitationis poenam nemo patitur*, presuppone che la volontà si traduca in realizzazione, almeno allo stadio del tentativo punibile.

Il dolo consiste in una volontà consapevole, e presuppone un processo cognitivo cosciente, differenziandosi così da una volontà non cosciente, riscontrabile in un'urgenza, un istinto o un impulso interno e non controllabile. Tale volontà cosciente ha come oggetto la condotta, le circostanze e l'evento. Per esempio nell'omicidio intenzionale, oggetto del dolo e della volontà cosciente è la morte del soggetto passivo del reato, e affinché la volontà sia rilevante ai fini della responsabilità penale, deve essere presente simultaneamente alla condotta.

Esistono numerose teorie della volontà e la dottrina continua a scontrarsi sul suo significato e il rapporto tra intenzione, desiderio e la capacità di prevedere le conseguenze dell'azione; tra avere intenzione rispetto a un risultato, e prevedere o rendersi conto che una certa azione realizzerà o potrebbe realizzare il risultato voluto[19].

[17] Si pensi ad esempio al reato di inosservanza del provvedimento del giudice che reprime la condotta antisindacale ex art. 28, comma 4, dello Statuto dei Lavoratori

[18] Tra i molti,Wittgenstein, L. et al. *Philosophische Untersuchungen: Philosophical Investigations*. Wiley-Blackwell, 2009.

[19] Per una panoramica del concetto giuridico di intenzione, si veda Duff, R. A. «Intention, agency and criminal liability: Philosophy of action and the criminal law». In: *Intention, Agency and Criminal Liability: Philosophy of Action and the Criminal Law, Blackwell* (1990). I principali esiti del

Il filosofo Michael Bratman ha sviluppato un'importante teoria sul rapporto tra intenzione e azione, su cui si basa il modello *Belief Desire Intention* (BDI), che descrive l'azione razionale umana[20]. Secondo Bratman, l'intenzione è un elemento chiave nella pianificazione degli stati futuri. Essa presuppone il desiderio, ma il desiderio da solo è insufficiente, poiché l'intenzione consiste in un impegno ad agire.

Il paradigma BDI fornisce un concetto di azione basato su tre attitudini o atteggiamenti mentali: (1) Beliefs, ovvero le credenze, (2) Desires, i desideri e (3) Intentions, le intenzioni.

1. Beliefs. Le credenze rappresentano lo stato informativo dell'agente. Esse corrispondono a ciò che egli percepisce, dati e informazioni, ma comprendono anche tutte le informazioni ottenute mediante il ragionamento e i processi inferenziali. Le credenze sono fatti che rappresentano il modo il cui l'agente vede la realtà e il mondo. Il concetto di credenza deve essere distinto da quello di conoscenza. Quest'ultima infatti presuppone la verità dei dati e delle informazioni, mentre ciò che un agente crede non deve necessariamente essere vero.

2. Desires.

   I desideri incarnano lo stato motivazionale, rappresentano gli obiettivi o gli stati futuri che l'agente vorrebbe e/o potrebbe realizzare. Un agente può avere contemporaneamente più desideri, che possono anche essere o entrare in conflitto tra loro.

3. Intentions.

   Le intenzioni rappresentano il risultato delle deliberazioni dell'agente, cioè, ciò che l'agente ha scelto di fare. Le intenzioni sono desideri che egli si è

dibattito dottrinale sul dolo sono rappresentati dalle teorie dell'intenzione, della rappresentazione e della volontà. Secondo una prima teoria, il dolo è la volontà di cagionare l'evento come fine ultimo della condotta, mentre altri lo ravvisano nella volontà della condotta e nella previsione dell'evento. Si tratta di visioni criticate, la prima infatti sarebbe troppo restrittiva poiché escluderebbe le ipotesi di dolo eventuale, mentre la seconda finirebbe per ricomprendere all'interno del del dolo anche la colpa cosciente. La teoria della volontà oggi dominante ritiene il dolo coscienza e volontà sia del fatto costitutivo che dell'evento tipico, ricomprendendo così il dolo eventuale, ma non la colpa cosciente. In ogni caso vi è concordanza nell'identificare nell'elemento soggettivo in esame due momenti: quello rappresentativo ed quello volitivo. Quindi il soggetto attivo del reato deve rappresentarsi gli elementi precedenti o concomitanti alla sua condotta. Alla rappresentazione deve accompagnarsi la volizione da parte dell'agente sia della condotta che dell'evento. Alcuni ritengono che si debba poi aggiungere un terzo momento, ovvero la coscienza dell'antigiuridicità del fatto. Tuttavia la dottrina non trova accordo sul punto e a nostro avviso, l'impossibilità di accogliere tale posizione, è già stata espressa trattando il tema dei reati di creazione puramente legislativa.

[20] Bratman, M. «Intention, plans, and practical reason». In: (1987).

in qualche misura impegnato a realizzare. Esse non possono entrare in conflitto e devono necessariamente essere coerenti tra loro.

Nel campo dell'intelligenza artificiale esistono diversi approcci per la progettazione di un agente intenzionale[21].

Le attitudini mentali del modello di ragionamento elaborato da Bratman possono essere replicate all'interno di un agente, secondo l'architettura dei sistemi BDI. In tale architettura, credenze, desideri e intenzioni sono strutturate secondo le seguenti componenti[22]:

1. *Belief store*: esso contiene le informazioni che l'agente ha sulla realtà e sul mondo. Attraverso un processo percettivo, l'agente osserva l'ambiente e cerca di interpretare i dati e le informazioni raccolte attraverso sensori, fornendo *input* al Belief store in termini di nuove credenze.

2. *Goal store*: esso contiene gli obiettivi o i desideri che un agente ha adottato e gli eventi cui può reagire. Un singolo obiettivo può avere diversi piani di realizzazione a seconda delle diverse situazioni. Per raggiungere questi obiettivi, l'agente costruisce piani e li memorizza nella *plan library*.

3. *Plan library*: raccoglie l'insieme dei piani che un agente può utilizzare, per raggiungere un determinato obiettivo. Un piano, a sua volta, è costituito da tre componenti principali:

    - il corpo: esso definisce una serie concreta di azioni, necessarie per la realizzazione di un piano;
    - le condizioni di attivazione: definiscono le circostanze in cui l'agente deve considerare e attivare il piano. Quando le condizioni di attivazione si verificano il piano diventa "attivo" e l'agente lo considera eseguibile;
    - le condizioni di termine: definiscono le condizioni in cui l'agente può riconsiderare la sua intenzione attuale, come per esempio nel caso in cui conduca all'impossibilità di realizzare il piano.

Quando un agente forma nuove credenze, valuterà quali piani possano essere attuati sulla base delle condizioni di attivazione che corrispondono alle nuove credenze, o costruirà nuovi piani, o adatterà quelli esistenti, per raggiungere i propri obiettivi, sulla base delle nuove condizioni. L'insieme emergente di piani corrisponde alle intenzioni dell'agente e ogni piano definisce una possibile

[21] Vedi, Wooldridge, M. J. *Reasoning about rational agents*. MIT press, 2000.

[22] Kinny, D., Georgeff, M. e Rao, A. «A methodology and modelling technique for systems of BDI agents». In: *Agents breaking away*. Springer, 1996, pp. 56–71.

linea di condotta. Pertanto, le intenzioni si riferiscono I desideri scelti dall'agente corrispondono alle sue intenzioni. Pertanto, le intenzioni si riferiscono sia all'impegno dell'agente verso i propri desideri (gli obiettivi da raggiungere attraverso i piani selezionati), sia al suo impegno verso i piani selezionati per raggiungere questi obiettivi.

Gli Agenti BDI sono sensibili ai cambiamenti ambientali e sono attualmente utilizzati in numerose applicazioni di successo, in ambienti complessi e dinamici, come ad esempio i veicoli senza pilota, il trading online, la logistica e il controllo del traffico aereo.

In conclusione, sembra che un agente che adotti un piano attraverso l'applicazione del modello BDI sia in grado di avere un'intenzione giuridicamente rilevante per il diritto penale. Proveremo ora a considerare come si possa dimostrare la presenza di intenzione all'interno di un agente.

Provare l'intenzione

Come più volte evidenziato, l'intenzione è uno "stato mentale" interno all'agente. Essa ha come oggetto il verificarsi di uno stato futuro e, per tale motivo, è molto più difficile dimostrare l'intenzione rispetto alla coscienza, che, invece, ha come oggetto stati attuali e presenti. La dottrina è unanime nel ritenere la sostanziale impossibilità di cogliere con precisione fatti che non appartengono ad una dimensione esteriore, bensì psichica, e quindi interiore, risultando gravoso lo sforzo di dare veste giuridica ad una realtà che è essa stessa incerta, in quanto non osservabile. Per affrontare la complessità relativa alla prova dell'intenzione, in particolare per alcune forme di dolo, parte della dottrina e della giurisprudenza ha adottato un criterio di accertamento di tipo ipotetico, in base a cui si presume che l'agente abbia voluto e inteso le conseguenze "naturali e probabili" delle proprie azioni.

Più precisamente, si presume che il soggetto agente abbia inteso i risultati della propria azione se (1) questa è stata compiuta con piena coscienza e (2) le conseguenze di tale azione erano altamente probabili e potevano essere previste dall'agente[23].

Secondo il criterio della prevedibilità devono essere soddisfatte due condizioni:

1. l'azione deve essere stata commessa con piena coscienza, e
2. i risultati dell'azione devono essere stati previsti dall'agente con un elevato grado di probabilità.

[23]Si veda per esempio Shute, S. «Knowledge and Belief in the Criminal Law». In: *Criminal law theory – doctrines of the general part* (2002)

Consideriamo se tale criterio di prevedibilità possa essere applicato ad un AI. Come osservato nell'esempio dei computer che giocano a scacchi, un AI è in grado di valutare le probabilità che si verifichino certi stati futuri, e agire di conseguenza: l'AI esamina opzioni alternative, costruisce piani di azione considerando i rispettivi risultati futuri, prende decisioni consapevoli per attuare un piano, e agisce di conseguenza.

In molti contesti, gli AI hanno la capacità di valutare la probabilità che una determinata azione conduca ad un certo risultato, e ne sono capaci così come lo sono gli esseri umani, e talvolta anche in modo molto più preciso, sulla base di quanto esposto fin'ora. In tali contesti, è quindi possibile affermare che un AI abbia avuto l'intenzione non solo di agire, ma anche di ottenere un certo risultato come conseguenza delle proprie azioni.

Abbiamo visto come la concezione giuridica e filosofica di intenzione possa essere d'aiuto per progettare e caratterizzare un sistema d'intelligenza artificiale capace di intenzione: un AI è in grado di sviluppare intenzioni, e tali intenzioni posso essere verificate. Esaminando l'elemento cognitivo, abbiamo evidenziato che gli AI sono capaci di consolidare una coscienza degli stati presenti, e i processi che conducono a tale consapevolezza possono essere monitorati con precisione, registrati e provati. Riguardo alla prova dell'intenzione sulla base del criterio di prevedibilità, le attività di un AI, che conducono al consolidamento dell'intenzione, possono essere anch'esse monitorate e registrate, così da avere una prova diretta dello stato intenzionale dell'agente.

Sulla base di quanto è stato considerato fino a questo momento, gli AI sono capaci di sviluppare intenzione, secondo i principi del diritto penale[24]. Ovviamente la sussistenza della *mens rea* all'interno di un AI dovrà essere valutata caso per caso, prendendo in considerazione le specifiche caratteristiche del singolo agente, nella situazione concreta. Proviamo ora ad esaminare un caso concreto, costruendo possibili scenari di responsabilità penale, sulla base dei modelli elaborati dalla dottrina e delle considerazioni svolte fino a questo momento.

4.5 Il caso del *Random Darknet Shopper Bot*

A novembre del 2014, il Random Darknet Shopper, un sistema automatico di transazioni telematiche, e più precisamente un bot[25], è stato utilizzato da un collettivo artistico di nome !Mediengruppe Bitnik, per fare acquisti casuali in uno

[24]Similmente, e giungendo alle stesse conclusioni Hallevy, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit.

[25]In generale, il termine bot, abbreviazione di robot, identifica un agente software che accede alla rete attraverso lo stesso tipo di canali utilizzati dagli utenti umani (per esempio, accede alle pagine Web, invia messaggi in una chat, si muove all'interno di videogiochi, e così via). Programmi di questo tipo sono diffusi in relazione a numerosi servizi in rete, con scopi vari, ma in

specifico *marketplace* sul Dark Web. Il bot è stato dotato di un budget di 100$ in BitCoin a settimana, per scandagliare i negozi online non indicizzati del deep web e acquistare oggetti, per poi spedirli nella sede del collettivo. L'acquisto era completamente randomizzato: il collettivo non sapeva cosa avesse comprato il bot fino al momento della consegna. Il bot ha acquistato gli oggetti più svariati, tra cui: alcune páia di jeans, delle sneakers, due cappelli muniti di telecamera e un mazzo di chiavi passe-partout dei vigili del fuoco, una borsa falsa di Louis Vuitton, una lattina di Coca-Cola piena di soldi, 200 pacchetti di sigarette Chesterfield e un DVD contenente una partita di ecstasy. Tutti gli acquisti sono stati fatti, e successivamente esposti, nell'ambito di una performance artistica intitolata "The Darknet: From Memes to Onionland", in mostra presso la Kunst Halle Sankt Gallen, a Zurigo e terminata l'11 gennaio 2015, con il sequestro del materiale illecito[26]. Da quanto riportato, sembra che il bot abbia posto in essere una condotta astrattamente riconducibile alla fattispecie criminosa di acquisto di sostanze stupefacenti.

Questo caso ci spinge a sollevare alcune domande: è possibile identificare un soggetto attivo del reato? Se sì, chi è penalmente responsabile? Esiste un vuoto di responsabilità? E in caso, come possiamo regolare questo tipo di fenomeni?

La maggior parte degli ordinamenti europei regola, facendo ricorso al diritto penale, numerose condotte legate alle cosiddette droghe pesanti, e in particolare: la coltivazione illegale, la produzione, la fabbricazione, l'estrazione, la preparazione, l'acquisto e la detenzione, la vendita, la distribuzione, la consegna a qualsiasi titolo, la mediazione, la spedizione, la spedizione in transito, il trasporto, l'importazione e l'esportazione di sostanze stupefacenti illegali[27].

Nel caso del *Random Darknet Shopper bot*, la fattispecie rilevante, ai fini della nostra indagine, è l'acquisto di sostanze stupefacenti.

Sotto il profilo materiale, nella maggior parte degli ordinamenti, per acquisto si intende la condotta di colui che ponga in essere con la controparte- che sia

genere legati all'automazione di compiti spesso troppo gravosi o complessi per gli utenti umani. Il tipo più diffuso di bot programmato per navigare sul Web è il cosiddetto web crawler o anche spider. Questi AS, navigano le pagine Web e seguendo i link ipertestuali presenti nel testo, passano dall'una all'altra, e raccolgono informazioni sui contenuti delle pagine, generalmente allo scopo di indicizzarle opportunamente nel database principale del motore di ricerca. Su progetti di tipo wiki, i bot svolgono soprattutto, ma non solo, compiti di riordino automatico delle pagine, compilazione dei collegamento, correzione di reindirizzamenti, e creazione di pagine di sintesi.

[26] La notizia è stata riportata da numerose testate giornalistiche tra cui *The Guardian*. http://www.theguardian.com/technology/2014/dec/05/software-bot-darknet-shopping-spree-random-shopper

[27] Nell'ordinamento italiano, il testo di riferimento è il Testo Unico delle leggi in materia di sostanze stupefacenti e psicotrope, di cui al decreto del Presidente della Repubblica n. 309 del 1990, e il DL 272/05 convertito dalla legge 49/06. In particolare, l'acquisto di sostanze stupefacenti è disciplinato ex art. 73 del DPR 309/1990.

in possesso dello stupefacente o che possa ragionevolmente procurarselo per cederlo al primo- il relativo negozio avente ad oggetto la sostanza stupefacente[28]. La condotta di acquisto non richiede necessariamente che l'acquirente entri in possesso dello stupefacente[29]. Essa si ritiene integrata in presenza del solo accordo tra le parti, in ossequio al principio consensualistico, senza che sia necessaria la materiale consegna della sostanza o il pagamento del corrispettivo[30].

Così, tornando al nostro caso, e tenendo in considerazione quanto detto trattando il tema dell'*actus reus*, possiamo asserire che la condotta di acquisto di sostanze stupefacenti è stata posta in essere dal bot. Possiamo chiederci se tuttavia, anche l'utente, e nel caso specifico i membri del collettivo artistico, possa essere considerato soggetto attivo del reato, avendo acquistato le sostanze stupefacenti attraverso il bot, o se sia possibile ravvisare un ipotesi di concorso o altre condotte penalmente rilevanti da parte del collettivo. Riguardo all'elemento soggettivo del reato di acquisto di sostanze stupefacenti, si tratta di un reato punibile a titolo di dolo. Gli ordinamenti disciplinano poi in modo diverso il grado di intenzione necessario ad integrare i requisiti della *mens rea*.

Le notizie riportate sui giornali, relative al caso del *Random Darknet Shopper bot*, non contengono tutte le informazioni necessarie per allocare correttamente eventuali responsabilità penali. Abbiamo quindi costruito cinque possibili scenari, prendendo in considerazione i modelli di responsabilità esaminati nel capitolo precedente .

1. **Primo scenario: il bot è stato progettato o impiegato con l'intenzione di commettere il reato, e il bot non è in grado di soddisfare i requisiti della *mens rea*.**.

   Il bot è utilizzato come mero strumento per commettere il reato ed esegue esattamente ciò che gli è stato ordinato dall'utilizzatore o dal programmatore. Il programmatore e/o l'utente non eseguono alcuna azione necessaria a integrare l'elemento materiale del reato.

   In questa ipotesi possiamo avere tre differenti sotto-scenari:

   (a) il bot non ha capacità cognitive e volitive;

   (b) le capacità cognitive e volitive del bot sono assimilabili a quelle di un incapace, come un minore non imputabile o un infermo di mente;

[28] Diversamente, per ricezione si intende la condotta di colui che materialmente riceva la sostanza, a prescindere dal titolo in forza di cui l'abbia ricevuta (acquisto, donazione, amicizia, deposito, mandato, etc.).

[29] Con riferimento all'ordinamento penale italiano, si veda in questo senso, Cass., sezione IV, 11 ottobre 2011, n. 3950, in C.E.D., Cass. n. 251736.

[30] Sempre con riferimento all'ordinamento italiano, si veda, Cass. sezione IV, 23 gennaio 2014, n. 6781, in C.E.D. Cass., n. 259284.

(c) le capacità cognitive e volitive del bot sono assimilabili a quelle degli animali;

Nei casi (a) e (b) il bot deve essere considerato come un agente innocente. Si tratta di ipotesi riconducibili alla figura dell'autore mediato. Essa si riferisce a colui che strumentalizza un altro essere umano, non colpevole o non punibile, come esecutore materiale del reato[31]. L'agente è considerato innocente poiché non imputabile o in ogni caso non in grado di soddisfare i requisiti della *mens rea*. Il bot ha posto in essere la condotta criminosa, mentre colui che ha orchestrato il reato è il soggetto penalmente responsabile di quello specifico reato. In questo caso la responsabilità penale dell'autore del reato (perpetrator) è determinata sulla base della condotta dell'agente innocente e della *mens rea* dell'autore mediato[32]. Esistono due principali candidati per il ruolo di autore mediato: il primo è il programmatore del bot, il secondo è l'utilizzatore. Sarà considerato penalmente responsabile quello tra i due che soddisfa i requisiti della *mens rea* e ha orchestrato il reato. Questo modello considera la condotta posta in essere dal bot come se fosse stata eseguita dal programmatore o dall'utente. La *ratio* alla base della responsabilità dell'autore mediato è l'uso strumentale del bot che difetta dei requisiti della *mens rea*.

Consideriamo ora il terzo sotto-scenario (c) e il modello di responsabilità applicabile. Molti ordinamenti considerano gli animali come oggetto di un diritto di proprietà e possesso da parte degli esseri umani. Se un animale provoca un danno, colui che ha il diritto di proprietà o possesso sull'animale è giuridicamente responsabile del danno. Per esempio, se un cane attacca un passante, il proprietario è legalmente responsabile per eventuali danni o lesioni. Mentre nella maggior parte degli ordinamenti si tratta di ipotesi di responsabilità civile, in altri sono le norme di diritto penale ad essere applicate. In entrambi i casi è il proprietario o colui che ne ha il possesso, e non l'animale, ad essere civilmente o penalmente responsabile. Nessun ordinamento giuridico considera gli animali come soggetti giuridici imputabili. Se consideriamo il caso in cui un cane attacca un essere umano su ordine, ad esempio, del proprietario, l'animale è utilizzato come

[31] Il concetto di autore mediato nasce nel sistema penale tedesco, in quanto tale ordinamento - ancorato, in tema di concorso di persone, alla teoria dell'accessorietà estrema- necessitava di tale concetto per giustificare la punibilità di comportamenti che, seppur ritenuti meritevoli di punizione, non avrebbero potuto essere sanzionati attraverso il ricorso alle norme sul concorso di persone. Al contrario, la giurisprudenza italiana ha generalmente accolto e applicato alle ipotesi di concorso la teoria dell'accessorietà minima. Tuttavia, di recente, tale figura sembra essere stata accettata anche nel nostro ordinamento. In particolare, si veda, Cass. pen., sezione I, 23 novembre 2011, n. 47667.

[32] Si veda anche, Gillies, P. *The Law of criminal complicity*. Law Book Company, 1980.

mero strumento per la commissione dello specifico reato e, per le eventuali lesioni subite dalla vittima, sarà penalmente responsabile il proprietario. Nel terzo sotto-scenario, quindi, il bot agisce su ordine dell'utilizzatore o del programmatore, acquistando le sostanze stupefacenti e saranno questi ultimi ad essere penalmente responsabili per la condotta posta in essere dal bot.

2. **Secondo scenario: il bot è stato progettato o impiegato con l'intenzione di commettere il reato, e il bot è in grado di soddisfare i requisiti della *mens rea*.**

   In questo secondo scenario, l'*actus reus* è attribuibile al bot, e sia il programmatore e/o l'utente che il bot soddisfano i requisiti dell'elemento soggettivo del reato.

   Proviamo a questo punto a fare un esperimento mentale: se sostituissimo il bot con un essere umano, si verificherebbe un ipotesi di concorso nel reato. Uno dei principali problemi nell'applicare al caso concreto l'ipotesi di concorso nel reato, e più in generale le norme di diritto penale, risiede nella mancanza di personalità giuridica degli AI. Essi non sono soggetti alle norme di diritto penale. Tuttavia, come già evidenziato, la maggior parte dei moderni ordinamenti giuridici riconosce forme di responsabilità penale a carico delle persone giuridiche. Esse possono esercitare diritti di proprietà e possesso, concludere contratti, essere civilmente responsabili, e in alcuni casi anche penalmente responsabili, per esempio per frode e danni ambientali. Un aspetto cruciale nel considerare le società come persone giuridiche risiede nella possibilità di assoggettarle a sanzioni sia civili che penali, anche se spesso non è così semplice come nel caso in cui il soggetto attivo del reato sia un essere umano. Tale condizione non è attualmente soddisfatta per le ipotesi in cui un AI commetta un reato. Per esempio, mentre le pene pecuniarie (multa o ammenda) generalmente riescono a esplicare le proprie funzioni nei confronti delle società, la cui sopravvivenza e il cui obiettivo principale e risiede nel patrimonio sociale, lo stesso non può certo essere dato per scontato nel caso degli agenti autonomi intelligenti. In primo luogo essi non possiedono un proprio patrimonio, e in ogni caso non è chiaro per quale motivo dovrebbero preoccuparsi di eventuali perdite finanziarie. Torneremo più nel dettaglio su tale questione nella sezione successiva.

   Nello scenario presentato, e secondo le leggi penali dei moderni ordinamenti giuridici, solo il programmatore e/o l'utente saranno penalmente responsabili per il reato commesso.

3. **Terzo scenario: il bot non è stato progettato o impiegato con l'intenzione di commettere il reato, ma il programmatore e/o l'utente hanno irragionevolmente accettato una serie di rischi che hanno portato al verificarsi della condotta criminosa.**

   In questo scenario, il bot soddisfa i requisiti dell'*actus reus*, ma né il programmatore, né l'utilizzatore avevano intenzione di commettere il reato. Essi non hanno pianificato il reato e non avevano intenzione di commetterlo attraverso l'uso strumentale del bot. Questo scenario si basa sulla capacità del programmatore e/o dell'utente di prevedere la potenziale commissione del reato.

   Possiamo distinguere quattro diversi sotto-scenari:

   (a) il bot non ha alcuna capacità cognitiva e volitiva;

   (b) le capacità cognitive e volitive del bot sono assimilabili a quelle di un incapace, come un minore non imputabile o un infermo di mente;

   (c) le capacità cognitive del bot sono assimilabili a quelle degli animali;

   (d) il bot soddisfa i requisiti della *mens rea*.

   Nelle ipotesi (a) e (b), il bot non soddisfa i requisiti della *mens rea*. Il programmatore e/o l'utente non erano a conoscenza del reato commesso e non avevano intenzione di commettere il reato, tuttavia pur rappresentandosi la possibilità che il bot commettesse un reato, hanno accettato tale possibilità. Essi, pongono in essere uno o più comportamenti pur sapendo che da essi origina un rischio sostanziale e ingiustificabile. Qui il programmatore e/o l'utente sono consapevoli delle potenziali conseguenze negative del proprio comportamento ma decidono di portare avanti la propria condotta costi quel che costi. Nel caso del *Random Darknet Shopper bot*, sembra che il programmatore e/o l'utente non abbiano posto vincoli e restrizioni al bot, riguardo al tipo di merce acquistabile e ai siti web da cui fare acquisti, permettendogli anzi di navigare il dark-web. Essi hanno lasciato che il bot operasse in un ambiente in cui potevano prevedere con un elevato grado di probabilità che il bot acquistasse merce illegale. Così se il reato di acquisto di sostanze stupefacenti, in base all'ordinamento giuridico in cui il reato viene commesso, è punibile anche a titolo di dolo eventuale, l'utilizzatore e/o il programmatore saranno ritenuti penalmente responsabili per il reato commesso dal bot.

   Nell'ipotesi (c), in accordo con quanto già visto nel primo scenario, potremmo applicare il modello zoologico. Per esempio, la maggior parte degli ordinamenti prevede una serie di prescrizioni nel caso in cui si sia

proprietari di animali pericolosi. In particolare possiamo prendere il caso in cui il proprietario di un cane, ufficialmente dichiarato pericoloso, abbia l'obbligo di mettere la museruola al cane quando lo porta a passeggio. Qualora il proprietario non ottemperi e il cane ferisca o uccida un essere umano, il proprietario sarà responsabile a titolo di dolo eventuale o per colpa, a seconda di quanto previsto dall'ordinamento giuridico concretamente preso in esame. Nel caso qui esaminato, se il reato di acquisto di sostanze stupefacenti fosse un reato colposo o punibile a titolo di dolo eventuale, l'utente e/o il programmatore saranno penalmente responsabili per la condotta del bot.

Nell'ipotesi (d), se sostituissimo il bot con un essere umano avremmo un caso di concorso nel reato. Tuttavia, come osservato in precedenza, gli AI non sono dotati di personalità giuridica e dunque non possono essere penalmente responsabili. Per la condotta del bot, gli unici responsabili saranno l'utilizzatore e/o il programmatore.

4. **Quarto scenario: il bot è stato progettato o impiegato con l'intenzione di commettere un reato, ma il bot, nella commissione del reato, eccede quantitativamente o qualitativamente rispetto al piano originale.**

   In questo scenario, il programmatore e/o l'utilizzatore progettano e utilizzano intenzionalmente il bot per commettere un reato, ma il bot si discosta dal piano originario e commette un reato diverso o più reati dello stesso tipo di quello pianificato. Questo scenario è simile all'ipotesi del concorso anomalo nel reato, come illustrato nel capitolo precedente. Supponiamo che un gruppo di persone si accordi per commettere una rapina in banca e che il piano prevede di utilizzare delle armi, per esempio pistole, solo a fini intimidatori. Tuttavia, durante la rapina una delle guardie rimane uccisa per mano di uno dei rapinatori. L'omicidio non era parte del piano e i complici non hanno posto in essere la condotta prevista dalla fattispecie di omicidio, non si sono accordati per commettere l'omicidio e tuttavia, nelle circostanze specifiche, un ipotetico agente ragionevole avrebbe potuto prevedere la possibilità di quanto accaduto.

   Negli ordinamenti giuridici dell'Europa continentale, così come negli ordinamenti di Common Law inglese, la responsabilità penale per il reato non pianificato è attribuita a tutti i complici. Sul modello del concorso anomalo nel reato[33], tutti i complici saranno penalmente responsabili sia per le fattispecie di rapina e omicidio.

[33] Il concorso anomalo nel reato è riconducibile al modello di responsabilità penale della *Natural Probable Consequence* costruito da Hallevy, come analizzato nel capitolo precedente.

In questo scenario, il bot ha qualitativamente (ha commesso più reati di tipo diverso) o quantitativamente (ha commesso più reati dello stesso tipo) ecceduto rispetto al reato pianificato. È necessario analizzare separatamente la responsabilità per il reato pianificato e la responsabilità per il reato non pianificato. Possiamo distinguere quattro diversi sotto-scenari:

(a) il bot non ha capacità cognitive e volitive;

(b) le capacità del bot sono assimilabili a quelle di un incapace, come per esempio un minore non imputabile o un infermo di mente;

(c) le capacità del bot sono assimilabili a quelle degli animali;

(d) il bot soddisfa i requisiti della *mens rea*.

Nelle prime tre ipotesi, secondo il modello dell'autore mediato, come analizzato nel primo scenario, la responsabilità cadrà sul programmatore e/o utente che ha programmato o utilizzato il bot per commettere il reato. Il programmatore e/o l'utente hanno programmato e/o utilizzato il bot come strumento per commettere il reato pianificato. Il bot non soddisfa i requisiti della *mens rea*, né per il reato pianificato, né per il reato non pianificato. Secondo il modello del concorso anomalo, il programmatore e/o l'utente saranno penalmente responsabili per il reato non pianificato, se questo era prevedibile in relazione al reato pianificato.

Nella quarta ipotesi, il programmatore e/o l'utente saranno penalmente responsabili per il reato pianificato. Inoltre, nel sotto-scenario considerato, il bot soddisfa i requisiti della *mens rea*. Se al posto del bot avesse agito un essere umano si sarebbe verificato un caso di concorso nel reato. Se il reato non pianificato è la conseguenza probabile del reato pianificato, il programmatore e/o l'utente saranno penalmente responsabili sia per il reato pianificato che per quello non pianificato. Il bot soddisfa i requisiti della *mens rea* e se in sua vece avesse agito un essere umano, si sarebbe verificato un caso di concorso nel reato. Al contrario, se il reato non pianificato, non era la conseguenza probabile di quello pianificato, solo il bot sarebbe dovuto essere penalmente responsabile per il reato non pianificato. Anche in questo caso, poiché gli AI non hanno personalità giuridica, il bot non potrà essere ritenuto responsabile di alcun reato.

5. **Quinto scenario: il bot soddisfa i requisiti dell'*actus reus*, ma nessuna forma di intenzione può essere ascritta al programmatore e/o all'utente.**

   Possiamo distinguere due sotto-scenari:

   (a) il bot non soddisfa i requisiti della *mens rea*;

(b) il bot soddisfa i requisiti della *mens rea*.

Come più volte ricordato, ai fini dell'attribuzione della responsabilità penale devono essere soddisfatti due requisiti: l'elemento oggettivo e l'elemento soggettivo del reato.

Nella prima ipotesi, nessuna forma di intenzione o *mens rea* può essere ascritta al programmatore e/o all'utente, e al bot. Questo caso non sarebbe quindi soggetto ad azione penale.

Nella seconda ipotesi, l'unico a poter essere penalmente responsabile sarebbe il bot, ma poiché non dotato di personalità giuridica, la responsabilità per il reato commesso non potrà essere attribuita ad aluco.

Esiste un modo per regolare casi di questo tipo, così da ridurre l'ipotesi di reati commessi da AI? Nella sezione successiva saranno presentate e discusse alcune ipotesi di regolazione del fenomeno.

4.6 Possibili soluzioni: proposta di un modello di regolazione informatico-giuridico del fenomeno

Alcuni dei casi considerati nella sezione precedente mostrano un vuoto di responsabilità per le ipotesi in cui un AI ponga in essere una condotta che, se commessa da un essere umano, integrerebbe una fattispecie di reato, e tuttavia, la mancanza di *mens rea* del programmatore e/o dell'utente, non permette di allocare la responsabilità penale, per il fatto criminoso, ad alcuno dei soggetti coinvolti.

Proviamo ora a considerare come monitorare l'eventualità che tale fenomeno divenga sempre più emergente, e a valutare possibili opzioni per limitare la commissione di reati da parte di agenti autonomi intelligenti.

Una prima possibilità è quella di limitare il tipo di *task* assegnati agli AI e limitare la loro autonomia nell'esecuzione dei compiti ad essi affidati, considerando il contesto specifico in cui gli AI si troveranno ad operare e i rischi ad esso legati. In una area altamente sensibile, come il settore militare ad esempio, una buona scelta potrebbe essere quella di limitare l'autonomia degli AI e prevedere un controllo da parte degli esseri umani prima che l'AI possa procedere e compiere azioni particolarmente rischiose come sganciare una bomba.

Alcuni autori hanno proposto soluzioni volte a prevenire la commissione di reati da parte degli AI, senza tuttavia limitarne le capacità cognitive e volitive e

la loro autonomia. Una soluzione estrema, sostenuta da una parte della dottrina[34], consiste nell'attribuire agli AI personalità giuridica, patrimonio, fino a riconoscere la possibilità che siano ritenuti penalmente responsabili delle proprie azioni e assoggettabili alle sanzioni penali, secondo un sistema di equivalenze e opportuni adattamenti[35]. Questo approccio è stato sostenuto in particolare da Hallevy, come illustrato nel capitolo precedente. Una parte della dottrina afferma l'applicabilità agli agenti autonomi intelligenti del modello zoologico, secondo cui gli effetti delle condotte di un AI ricadrebbero sempre nella sfera giuridica del programmatore e dell'utilizzatore. Altri ancora, senza arrivare a ipotizzare l'applicabilità del diritto penale agli AI, hanno constatato come alcuni agenti autonomi intelligenti siano in grado di essere sensibili e reattivi rispetto alle norme, e rieducati e corretti nel loro comportamento normativo, nelle ipotesi in cui violino una norma. Negli ultimi anni, numerose discipline, che spaziano dalle scienze sociali all'ingegneria computazionale, hanno mostrato un crescente interesse nella necessità di assicurare un controllo sociale, in particolare nel dominio dei sistemi multi-agente[36]. Uno degli approcci più utilizzati e significativi, utilizza le norme per governare e vincolare il comportamento degli agenti autonomi intelligenti, e modellare i loro processi computazionali e cognitivi[37]. Tale approccio presuppone la progettazione di agenti normativi, vale a dire, agenti responsabili verso le norme e capaci di agire applicando tali norme e di rispondere a eventuali sanzioni. Gli agenti normativi hanno la capacità di (i) rappresentarsi le norme, essere motivati a seguire tali norme e a modificarle durante il loro ciclo di vita (rappresentazione della conoscenza); (ii) riconoscere e inferire le norme seguite da altri agenti (teoria dell'apprendimento); comunicare norme ad altri agenti (comunicazione e teoria delle reti); e (iv) imporre sanzioni ad altri agenti nel caso in cui tali norme non siano da essi rispettate (morale e diritto)[38]. Ad oggi, gli agenti normativi sono per lo più basati su architetture BDI, sia in relazione alla scelta degli obiettivi da perseguire, che all'elaborazione di piani, così da fornire loro un target di ragionamento[39]. Possiamo distinguere

[34]Così Hallevy, «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control», cit.; idem, *When Robots Kill: Artificial Intelligence Under Criminal Law*, cit.

[35]idem, *Liability for Crimes Involving Artificial Intelligence Systems*, cit.

[36]In particolare, si vedaHollander, C. D. e Wu, A. S. «The current state of normative agent-based systems». In: *Journal of Artificial Societies and Social Simulation* 14.2 (2011), p. 6.

[37]Si veda, per esempio Boella, G., Van Der Torre, L. e Verhagen, H. «Introduction to Normative Multi-agent Systems». In: *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2007

[38]Si vedano, Neumann, M. «Norm internalisation in human and artificial intelligence». In: *Journal of Artificial Societies and Social Simulation* 13.1 (2010), p. 12; Hollander e Wu, «The current state of normative agent-based systems», cit.

[39]Per tutti, Castelfranchi, C. et al. «Deliberative normative agents: Principles and architecture». In: *Intelligent agents VI. Agent theories, architectures, and languages*. Springer, 1999, pp. 364–378.

due approcci differenti.

Secondo un primo approccio, le norme sono strutturate in modo statico all'interno dei protocolli e del codice dell'agente[40], come vincoli interni all'architettura[41], in modo che l'agente non possa decidere di violarli nel perseguimento di obiettivi che esso ritenga più importanti o vantaggiosi. Questo tipo di agenti cosiddetti *norm follower*, essi non possono adattare e mutare il proprio comportamento nel tempo, sulla base dell'esperienza maturata.

Tuttavia, per quanto la società possa beneficiare di questo tipo di agenti, è necessario considerare anche la possibilità di progettare agenti capaci di "violazioni intelligenti" delle norme. Si consideri, per esempio, il caso in cui un'auto senza conducente debba evitare di travolgere un gruppo di pedoni, che abbia improvvisamente attraversato la strada, senza rendersi conto dell'arrivo dell'auto e del rischio di essere travolti. Supponiamo che sia troppo tardi per frenare, e sia necessario sterzare invadendo la corsia opposta di marcia, superando la doppia linea continua che divide le due carreggiate. Nonostante la manovra sia vietata dalle norme del codice della strada, può essere la scelta più ragionevole, o "intelligente", qualora l'alternativa consista nel travolgere i pedoni. Abbiamo quindi bisogno di un agente autonomo capace di:

1. avere consapevolezza dell'esistenza di una norma;
2. prendere in considerazione la norma nel corso proprio processo decisionale e nella valutazione di possibili azioni;
3. decidere se seguire o meno la norma nel caso concreto; e anche
4. verificare se una determinata azione (sia la propria o quella di altri agenti che operano nel medesimo sistema) è o meno conforme alla norma.

In questo caso le norme non possono essere costruite nell'architettura di un agente come vincoli alla selezione degli obiettivi e al processo decisionale. È importante notare che, considerare una norma non significa necessariamente seguirla, ma solo che gli obiettivi che l'agente seleziona, e i piani che decide di attuare in base agli obiettivi selezionati, saranno sviluppati sulla base della convinzione e della consapevolezza dell'agente che la norma esiste. È necessaria un'architettura che permetta all'agente di ragionare sull'opportunità di applicare le norme rilevanti, nel fissare gli obiettivi e selezionare i piani che porteranno

[40] Jennings, N. R. «Commitments and conventions: The foundation of coordination in multi-agent systems». In: *The knowledge engineering review* 8.03 (1993), pp. 223–250.

[41] Per tutti, Shoham, Y. e Tennenholtz, M. «On the synthesis of useful social laws for artificial agent societies (preliminary report)». In: *AAAI*. 1992, pp. 276–281; Shoham, Y. e Tennenholtz, M. «Emergent conventions in multi-agent systems: Initial experimental results and observations». In: *KR-92* (1992), pp. 225–231.

all'adozione di un certo comportamento da parte dell'agente[42]. Questo tipo di agente normativo non è un semplice *norm-follower*. Esso può conformare il proprio comportamento a norme sociali e giuridiche, ma può anche violare una norma per necessità o convenienza, sulla base delle circostanze del caso concreto. Mentre nel primo caso l'agente normativo rispetterà sempre e in ogni caso le norme (quando possibile), nel secondo deciderà se conformare o meno le sue azioni ad esse, sulla base del proprio ragionamento.

In entrambi gli approcci, le norme hanno un ruolo nella standardizzazione dei comportamenti di un AI, ma il secondo, (come illustrato nell'esempio dell'auto senza conducente) sembra essere più adatto per gli agenti che operano in un ambiente fortemente dinamico, permettendo loro di agire in maniera più flessibile, alla luce dei cambiamenti e delle circostanze ambientali, e anche nel caso in cui abbiano informazioni incomplete sull'ambiente in cui operano.

Alcune considerazioni conclusive

Il rapido sviluppo di sofisticati sistemi d'intelligenza artificiale pone la necessità urgente di adottare misure capaci di proteggere e garantire la sicurezza della società, da possibili danni agli interessi protetti dal diritto penale, siano essi individuali o collettivi. Abbiamo esaminato diversi possibili approcci. A chi scrive sembra che in alcuni contesti specifici, ci riferiamo in particolare a quello delle armi autonome intelligenti, l'autonomia degli AI debba necessariamente essere limitata, a fronte dei rischi che essa comporta e delle potenziali ripercussioni disastrose in termini di vite umane. Il modello zoologico, valido per alcune tipologie di sistemi d'intelligenza artificiale, appare talvolta inadeguato poiché, a fronte di livelli di autonomia elevati e capacità cognitive e volitive sviluppate, non propone soluzioni efficaci per arginare la possibile commissione di reati da parte di agenti autonomi intelligenti. Tale modello si limita a regolare i rapporti tra AI e persone fisiche o giuridiche e a far ricadere gli effetti di eventuali condotte criminose all'interno della sfera giuridica di programmatori e proprietari e, come già evidenziato, anche se in linea teorica la legge potrebbe disciplinare il comportamento degli AI allo stesso modo in cui disciplina quello degli animali, sembra ormai evidente la necessità di prepararsi ad una nuova classe di azioni giuridicamente rilevanti e difficilmente riconducibili agli animali.

Ancora, la responsabilità penale delle persone giuridiche è senza dubbio un enorme passo in avanti nell'area del diritto penale e i modelli utilizzati per sostenere tale responsabilità rivelano preziosi indizi per costruire un quadro dogmatico plausibile per un eventuale responsabilità penale degli AI. Inoltre, come

[42] Per un esempio di architettura per agenti normativi deliberativi e la approfondimenti sulla relazione tra norme, obiettivi, piani, e attuazione di un comportamento, si veda, Castelfranchi et al., «Deliberative normative agents: Principles and architecture», cit.

abbiamo evidenziato cercando di stabilire la possibilità che certi agenti autonomi siano capaci di soddisfare i requisiti dell'*actus reus* e della *mens rea*, il diritto penale mostra un certo grado di flessibilità, qualora per ragioni di politica criminale risulti necessaria e siano soddisfatte certe premesse dogmatiche. Pur ammettendo la possibilità di riconoscere personalità giuridica agli AI, dotandoli di un proprio patrimonio, e che certi agenti siano in grado di soddisfare i requisiti dell'elemento oggettivo e soggettivo del reato, a chi scrive sembra che manchi un pezzo del puzzle per poter applicare il diritto penale agli AI. Quale utilità avrebbe applicare sanzioni penali agli agenti autonomi intelligenti, nel modo in cui sono descritte da Hallevy? Sia che si abbracci una definizione di reato di natura puramente formale, inteso come fatto cui la legge ricollega una sanzione penale, cosicché il concetto di reato è determinato esclusivamente in funzione delle conseguenze giuridiche (pena o misura di sicurezza) che il legislatore riconnette a tali fatti; sia che si abbracci una definizione di reato di tipo sostanziale, secondo cui è reato un fatto che aggredisce un bene giuridico meritevole di protezione, sempreché l'aggressione sia tale da far apparire inevitabile il ricorso alla pena e che sanzioni di tipo non penale siano insufficienti a garantire un'efficace tutela, un giudizio di colpevolezza di un AI non raggiungerebbe alcun risultato utile per la società o per la vittima del crimine. L'AI non subirebbe alcun effetto deterrente della condanna, né potrebbe percepirne il disvalore, e a chi scrive sembra indispensabile affinché la sanzione penale possa esplicare le proprie funzioni. Dunque perché dovremmo applicare il diritto penale agli AI, venendo meno i presupposti politico criminali del sistema sanzionatorio?

Ad oggi, ci sembra che la soluzione migliore, per regolare il possibile emergere di un fenomeno criminoso legato alle condotte di agenti autonomi intelligenti, sia quella di adottare la teoria degli agenti normativi. Come abbiamo evidenziato, sia che si tratti di agenti cd. *norm follower*, sia che si tratti di agenti capaci di "violazioni intelligenti" delle norme, queste ultime hanno un ruolo e un'efficacia nella standardizzazione dei comportamenti di un AI. In base all'ambiente in cui gli agenti si troveranno ad operare sarà possibile decidere quale dei due approcci sia più adatto. Qualora si verifichi una violazione non giustificabile, e non sia riscontrabile un comportamento doloso o colposo del programmatore e/o dell'utente, non sarà necessario ricorrere a sanzioni di tipo penale, ma intervenire direttamente sul sistema ed evitare la possibilità che, nelle medesime circostanze, si verifichi nuovamente.

Sulla base di quanto esposto fin'ora e delle considerazioni svolte, proviamo a costruire e analizzare uno scenario in area medica e a delineare eventuali profili di responsabilità. Il sistema preso in esame è Watson, un sistema esperto utilizzato in campo diagnostico, le cui caratteristiche saranno approfondite nelle sezioni successive.

4.7 Il Dr. Watson: costruzione e analisi di uno scenario

Negli ultimi cinque anni, venticinque ricercatori dell'IBM hanno sviluppato Watson[43], un super-computer apparso per la prima volta nel 2011 all'interno di un programma televisivo americano di nome Jeopardy! Si tratta di un quiz televisivo, dove per poter rispondere alle domande è necessaria una piena comprensione del linguaggio naturale, compreso l'uso di regionalismi e giochi di parole[44]. Storicamente, il linguaggio naturale ha rappresentato un limite computazionale, a causa delle ambiguità e della complessità presenti nel linguaggio umano[45]. Watson non solo è in grado di elaborare e comprendere le domande poste in linguaggio naturale, ma anche di rispondere[46]. Al contrario di un motore di ricerca, come Google per esempio, che cerca parole chiave per indirizzare l'utente verso siti web e documenti in cui è possibile trovare la risposta a ciò che si cerca, Watson è in grado di comprendere domande poste e restituire all'utente risposte corrette[47]. Si tratta di un passo gigantesco per l'intelligenza artificiale, per la cosiddetta *deep analysis* e per la processazione del linguaggio[48].

Dal 2011 l'IBM collabora con diverse cliniche americane per la sperimentazione e l'uso di Watson come assistente decisionale del personale medico, in campo diagnostico e terapeutico.

[43] Watson prende il nome del fondatore dell'International Business Machines Corporation (IBM), Thomas J. Watson. Baker, S. *Final Jeopardy: man vs. machine and the quest to know everything*. Houghton Mifflin Harcourt New York, NY, 2011.

[44] Si vedano, Thompson, C. «What is IBM's Watson». In: *New York Times Magazine (June 2011). http://www. nytimes. com/2010/06/20/magazine/20Computer-t. html* (2010); Video, *Why Jeopardy!?* disponibile al link: https://www.youtube.com/watch?v=ZvDyE9Guwls e anche, http://www-07.ibm.com/innovation/in/watson/what-is-watson/why-jeopardy.html; https://vimeo.com/20106636; *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, http://blog.ted.com/experts-and-ibm-insiders-break-down-watsons-jeopardy-win/.

[45] Video, *The Science Behind an Answer*, https://www.ibm.com/developerworks/community/blogs/video-portal/entry/ibm_watson_the_science_behind_an_answer?lang=en; disponibile anche all'indirizzo, http://www.youtube.com/watch?v=DywO4zksfXw,1:07\T1\textendash1:12.

[46] Video,*The Science Behind an Answer*, https://www.ibm.com/developerworks/community/blogs/video-portal/entry/ibm_watson_the_science_behind_an_answer?lang=en.

[47] *The Science Behind an Answer*, vedi nota precedente, 1:13-1:23; Thompson, «What is IBM's Watson», cit.

[48] Ferrucci, D. et al. «Building Watson: An overview of the DeepQA project». In: *AI magazine* 31.3 (2010), pp. 59–79; Paper, I. W. «Watson – A System Designed for Answers The future of workload optimized systems design». In: *IBM Systems and Technology* (2011), pp. 1–6; Video, *A System Designed for Answers*, IBM, http://www-03.ibm.com/systems/power/advantages/watson/index.html, 0:15-0:31, e 1:45-1:54; disponibile anche all'indirizzo https://www.youtube.com/watch?v=cU-AhmQ363I.

L'impiego di Watson è l'ideale in area sanitaria per diverse ragioni e tuttavia pone quesiti molto importanti sul piano della responsabilità. Proviamo a capire meglio chi è il Dr. Watson, perché è così utile e che tipo di problemi possono insorgere in campo medico.

4.7.1 Chi è il Dr. Watson?

Sebbene l'IBM abbia sviluppato Watson per partecipare a Jeopardy!, un team di medici e operatori sanitari della Columbia University lo utilizza oggi per fare diagnosi, suggerire terapie, e rispondere a domande in campo medico[49]. Watson è il primo sistema d'intelligenza artificiale in grado di comprendere domande poste in linguaggio naturale e sfruttare l'intero corpo di conoscenze mediche e le informazioni contenute nei *personal record* dei pazienti, per sviluppare diagnosi e piani terapeutici, il tutto in meno di tre secondi[50].

Attualmente è in grado non solo di suggerire diagnosi e piani terapeutici ai medici, ma anche di calcolare le probabilità di successo della terapia, fornendo prove mediche a fronte di ciascuna opzione[51]. Presto sarà in grado di interfacciarsi con l'equipe medica e i dispositivi medici e somministrare direttamente il piano terapeutico ai pazienti[52]. Una volta connesso ad Internet, sarà in grado di relazionarsi con pazienti in qualsiasi parte del mondo[53]. Watson è in grado di

[49] Circa ogni dieci anni l'IBM lancia una sfida nel campo dell'intelligenza artificiale. Nel 1997, un computer chiamato *Deep Blue* fu il primo sistema d'intelligenza artificiale a battere l'allora campione del mondo di scacchi Garri Kimovič Kasparov Baker, *Final Jeopardy: man vs. machine and the quest to know everything*, cit., 20. La sfida successiva fu lo sviluppo di *Blue Gene*, il primo sistema d'intelligenza artificiale progettato per l'analisi e il sequenziamento del genoma umano ibid. Watson rappresenta, ad oggi, l'ultima grande sfida. I ricercatori dell'IBM decisero di progettare Watson affinché partecipasse a Jeopardy!, per i progressi che sarebbero stati necessari nel campo dell'elaborazione del linguaggio naturale e della *Deep analysis* ibid. Gareggiando a Jeopardy!, Watson ha sconfitto i due più grandi campioni che avessero mai partecipato al quiz televisivo, ha vinto un milione di dollari ed è stato incoronato re di Jeopardy, *IBM's Watson supercomputer crowned Jeopardy king*, BBC NEWS (17 Febbraio 2011), http://www.bbc.com/news/technology-12491688. Si veda anche, Darren Murph, *Columbia Doctors Turn to IBM's Watson for Patient Diagnosis, Clairvoyance*, https://www.engadget.com/2011/03/24/columbia-doctors-turn-to-ibms-watson-for-patient-diagnosis-cla/, 24 marzo 2011.

[50] Video, *Perspectives on Watson: Healthcare*, 1:44–2:10, IBM, http://www-03.ibm.com/innovation/us/watson/watson_in_healthcare.shtml, disponibile anche all'indirizzo https://www.youtube.com/watch?v=vwDdyxj6S0U

[51] *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, *supra*, 14:57–15:08, 17:40–19:15.

[52] *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, *supra*, 12:10–12:51, 15:53–16:55

[53] L'uso di Watson in telemedicina ha un grande potenziale, *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, *supra*, 12:10–12:51, 15:53–16:55; si veda anche Teich, J. M. et al. «The informatics response in disaster, terrorism, and war». In: *Journal of the American Medical Informatics Association: JAMIA* 9.2 (2002), p. 97.

avere a portata di mano un numero elevatissimo di informazioni che un singolo medico non sarebbe in grado di avere (attualmente,si stima che la letteratura medica in campo diagnostico raddoppi ogni sette anni[54]), di processare tali informazioni in una manciata di secondi, monitorare un numero elevatissimo di variabili relative alla salute del paziente e controllare costantemente il suo stato di salute, considerare le informazioni presenti nei *personal records* dei pazienti e infine individuare la terapia più corretta e indicata[55]. Ancora, Watson può essere usato come sistema di supporto alle decisioni e alla diagnostica differenziale, discriminando tra patologie analoghe, che vengono progressivamente eliminate in base alla presenza o assenza di altri sintomi e segni[56]. Calcolando tutte le possibili diagnosi e i cicli di trattamento, e le probabilità di successo, Watson consente a medici e pazienti di prendere decisioni informate sul tipo di cura da intraprendere[57], e ridurre possibili cause di errore da parte di medici, colmando eventuali *gap* di informazioni, durante il trattamento[58] Quando fu progettato per partecipare a Jeopardy!, utilizzava un database costruito dall'IBM, senza poter accedere a Internet[59]. Attraverso il collegamento alla rete, è il grado di consultare guidelines relative ai trattamenti, banche dati, come per esempio *Wolfram Alpha* e PubMed, espandendo sempre più la sua base di conoscenza[60]. Rispetto ai sistemi di tele-chirurgia e chirurgia robotica, che dipendono completamente dal medico per poter funzionare, Watson lavora autonomamente e restituisce ai medici il risultato dei processi e delle elaborazioni svolte.

Tuttavia, se da un lato i contributi che Watson è in grado di dare in campo medico sono sbalorditivi, dall'altro le potenziali complicazioni circa i profili di responsabilità sono altrettanto impressionanti. Prima di procedere all'esame di tali profili, cerchiamo di approfondire il funzionamento tecnico del sistema e di individuarne il livello di automazione.

[54] Video, *Perspectives on Watson: Healthcare*, 1:44–2:10, *supra*.

[55] Secondo la ricerca, la capacità di un essere umano di processare informazioni,si limita a poter prendere in considerazione contemporaneamente meno di quattro differenti variabili, si veda, Halford, G. S. et al. «How many variables can humans process?» In: *Psychological science* 16.1 (2005), pp. 70–76. Riguardo a Watson, ancora *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, *supra*, 17:40–19:15.

[56] *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, *supra*, 20:18–20:46

[57] *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, *supra*, 17:40–19:15.

[58] *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, *supra*, 28:49–29:20; *Perspectives on Watson: Healthcare*, 0:46–0:58, 1:44–2:10.

[59] Baker, *Final Jeopardy: man vs. machine and the quest to know everything*, cit., 30.

[60] *Experts and IBM Insiders Break Down Watson's Jeopardy! Win*, *supra*, 12:10–:51, 15:53–16:55.

4.7.2 Analisi del sistema e dei livelli di automazione

Watson è un'applicazione avanzata di elaborazione del linguaggio naturale, *information retrieval*, rappresentazione della conoscenza, ragionamento e apprendimento automatico nel campo del cd. *open domain question answering*. È un sistema di ottimizzazione basato sull'architettura IBM DeepQA, eseguita su un cluster di server basati su processori IBM POWER7, per la formulazione di ipotesi, raccolta massiva di controprove, analisi e *scoring*.

DeepQA è un'architettura per il calcolo probabilistico massivo basato su un sistema di prove ed evidenze. Nel corso della sua partecipazione a Jeopardy! sono state utilizzate più di cento tecniche differenti per l'analisi del linguaggio naturale, per identificare le fonti di informazione, trovare e generare ipotesi, trovare e valutare prove e infine riunire e classificare tali ipotesi. Più che le singole tecniche utilizzate, ciò che rileva è il modo in cui esse sono combinate all'interno di DeepQA, poiché la sovrapposizione di approcci e tecniche differenti fa sì che i loro singoli punti di forza contribuiscano al miglioramento della precisione, dell'affidabilità e della velocità del sistema.

Tale architettura è basata su quattro principi generali:

1. **Massive Parallelism**: calcolo parallelo e massivo per la valutazione multipla di interpretazioni e ipotesi.

2. **Many experts**: integrazione, applicazione e valutazione simultanea di una vasta gamma di domande probabilistiche debolmente accoppiate e argomentazioni analitiche.

3. **Pervasive confidence estimation**: collaborazione simultanea di tutti i componenti del sistema per valutare ipotesi e associare a ognuna di esse un determinato grado di sicurezza, attribuendo un punteggio alle domande e alle possibili interpretazioni. Un sistema sottostante di calcolo ed elaborazione del grado di sicurezza impara ad ampliare e combinare i punteggi assegnati.

4. **Integrate shallow and deep knowledge**: bilanciamento nell'uso di una semantica rigorosa e non rigorosa, attraverso l'uso di numerose ontologie.

Per interpretare, comprendere e rispondere alle domande Watson si avvale di un processo in quattro fasi[61].

- Fase 1: Watson scompone la domanda in pezzi di discorso e singole parole, per comprendere il tipo di domanda e il significato della domanda[62];

[61] *The Science Behind an Answer*, *supra*, 2:02-2:18.
[62] *The Science Behind an Answer*, *supra*, 2:18–2:48.

- Fase 2: dopo aver determinato il tipo di domanda, Watson fa una ricerca all'interno del suo database, facendo emergere migliaia di possibili risposte a tutte le possibili domande generate nella prima fase[63];
- Fase 3: Watson valuta una serie di ipotesi e testa possibili prove, valutando sia quelle positive che quelle negative, per tutte le risposte possibili generate nella seconda fase[64]. È in grado di testare e valutare le possibili risposte corrette, grazie al sistema POWER7, un computer dotato di grande potenza di elaborazione che l'IBM ha progettato appositamente per Watson. Tale sistema ha ampia applicazione sia nel campo della *deep analysis* che del *problem solving*[65];
- Fase 4: Watson raccoglie e classifica tutte le potenziali risposte corrette, utilizzando la propria esperienza passata per rispondere a domande simili, crea una classifica, calcola la probabilità che ogni singola risposta sia corretta[66], e infine risponde alla domanda.

Watson esegue l'intero processo in meno di tre secondi[67].

Sulla base delle informazioni raccolte e della tassonomia dei livelli di automazione elaborata da Save e Feuerberg, riportata nel primo capitolo[68], abbiamo evidenziato i livelli di automazione di Watson, come riportati nella tabella 4.1.

[63] *The Science Behind an Answer*, *supra*, 2:48–3:19.
[64] *The Science Behind an Answer*, *supra*, 3:19–4:19
[65] Paper, «Watson – A System Designed for Answers The future of workload optimized systems design», cit.; *A System Designed for Answers*, *supra*, 1:17–1:30, 1:45–1:54.
[66] *The Science Behind an Answer*, *supra*, 4:19–6:03.
[67] Id.
[68] Tabella 2.3.

TABELLA 4.1: Watson- Level of Automation (LOA)

A	**B**	**C**	**D**
Information Acquisition	**Information Analysis**	**Decision and Action Selection**	**Action Implementation**
A5 Full Automation Support of Information Acquisition	**B5 Full Automation Support of Information Analysis**	**C2 Automated Decision Support**	**D0 Manual Action and Control**
The system supports the human in acquiring information on the process s/he is following. The system integrates data coming from different sources and filters and/or highlights the information items which are considered relevant for the user. The criteria for integrating, filtering and highlighting the relevant info are predefined at design level and not visible to the user (transparent to the user in Computer Science terms)	The system performs comparisons and analyses of data available on the status of the process being followed based on parameters defined at design level	The system proposes one or more decision alternatives to the human, leaving freedom to the human to generate alternative options. The human can select one of the alternatives proposed by the system or her/his own one	The human executes and controls all actions manually

Le informazioni relative al livello di automazione di Watson saranno utilizzate, nella sezione successiva, per indagare i profili di responsabilità, sulla base del livello di automazione e della divisione dei compiti tra essere umano e macchina[69].

4.7.3 Analisi di uno scenario e profili di responsabilità

Abbiamo più volte evidenziato come, l'introduzione di automazione e sistemi d'intelligenza artificiale all'interno delle relazioni umane e delle organizzazioni

[69]L'uso della LOAT come strumento di analisi dei compiti e delle relative responsabilità, è oggi utilizzato nel settore aeronautico, in particolare si veda il metodo di analisi sviluppato daSchebesta, H. et al. «Design According to Liabilities: ACAS X and the Treatment of ADS-B Position Data». In: ().

complesse sia in grado di avere un effetto dirompente sui profili di responsabilità dei soggetti coinvolti. In questa sezione abbiamo provato a costruire uno scenario e ad evidenziare le potenziali criticità in grado di emergere dall'uso di Watson in area medica.

Scenario: Watson è utilizzato per analizzare i sintomi di un paziente, diagnosticare da quale patologia è affetto ed elaborare un piano terapeutico adeguato. Il sistema acquisisce le informazioni disponibili sul caso specifico, integrando i dati provenienti da diverse fonti (sintomi, *personal records* del paziente, letteratura medica rilevante, etc.) e fa emergere possibili diagnosi e trattamenti. Esegue confronti e analisi dei dati disponibili, valuta una serie di ipotesi e testa possibili prove positive e negative per tutte le diagnosi e i corrispondenti piani terapeutici, emersi in precedenza. Watson raccoglie e classifica tutte le potenziali diagnosi corrette e i rispettivi piani terapeutici, sulla base della propria esperienza, crea una classifica e calcola la probabilità che ogni singola risposta sia corretta. Infine restituisce un numero di alternative riguardanti diagnosi e relativi piani terapeutici, assegnando a ognuna di esse un valore percentuale indicativo del grado di certezza. A seguito del trattamento terapeutico, il paziente muore a causa di un errore verificatosi nella fase di (1) acquisizione delle informazioni e/o (2) analisi delle informazioni e/o (3) decisione e selezione dell'azione (patologia e relativo piano terapeutico), e/o (4) attuazione del piano di azione (somministrazione del piano terapeutico al paziente).

Prima di procedere con l'analisi dello scenario, è necessario sottolineare che non è stata presa in considerazione l'ipotesi in cui alcune delle informazioni utilizzate dal sistema, per elaborare una diagnosi e il relativo piano terapeutico, come per esempio i dati contenuti nella cartella clinica del paziente, fossero errate. Tale ipotesi non riguarda infatti un errore di acquisizione delle informazioni ma un errore contenuto della fonte di informazione, e come tale esclusivamente riconducibile ad un errore del medico o del personale sanitario. Sulla base di tutte le considerazioni fatte e delle informazioni fin qui acquisite, proviamo ad analizzare lo scenario che abbiamo costruito e ad evidenziare possibili profili di responsabilità.

La morte del paziente è riconducibile alla presenza di un errore in:

1. **Fase di acquisizione delle informazioni**

   Si tratta di un errore nella ricerca e selezione di tutte le possibili diagnosi e i relativi piani terapeutici. Dall'analisi LOA di Watson, è risultato un livello di automazione A5 nella fase di acquisizione delle informazioni. Il sistema fa una ricerca all'interno di banche dati, attingendo ad un numero elevatissimo di informazioni, integra i dati acquisiti dalle diverse fonti e seleziona tutte le possibili diagnosi e tutti i possibili piani terapeutici corrispondenti. I criteri di integrazione dei dati sono predefiniti a livello di

sistema e non sono visibili all'utente (medico/personale sanitario). Considerato il livello di automazione A5 la responsabilità penale potrà essere attribuita:

(a) al produttore, qualora sia riscontrabile un errore in fase di progettazione, legato ai criteri di selezione, filtro e integrazione delle informazioni, predefiniti a livello di sistema

(b) al sistema, qualora avesse personalità giuridica e fosse in grado di soddisfare i requisiti della *mens rea*, tenendo conto dei possibili scenari costruiti nella sezione 4.5.

L'utente (medico/personale sanitario) non potrà essere considerato penalmente responsabile, poiché egli non interviene nella fase di ricerca, acquisizione, integrazione e selezione delle informazioni.

2. **Fase di analisi delle informazioni**

   Si tratta di un errore attinente ai test e alla valutazione di prove e controprove di possibili diagnosi e piani terapeutici per tutte le possibili ipotesi generate nella fase di acquisizione delle informazioni. Dall'analisi LOA di Watson, è risultato un livello di automazione B5. Il sistema analizza le informazioni ed esegue i test sulla base di parametri definiti a livello di sistema. La responsabilità per un errore nella fase di analisi potrà essere attribuita:

   (a) all'utente (medico/personale sanitario), qualora egli abbia fornito al produttore i parametri per l'analisi e i test, successivamente definiti a livello di sistema;

   (b) al produttore, qualora i parametri siano stati impostati da quest'ultimo.

   (c) al sistema, qualora avesse personalità giuridica e fosse in grado di soddisfare i requisiti della *mens rea*, tenendo conto dei possibili scenari costruiti nella sezione 4.5.

3. **Fase di decisione e selezione dell'azione** Watson raccoglie e classifica tutte le potenziali diagnosi e piani terapeutici corretti, sulla base dei test svolti nella fase precedente e della propria esperienza, calcola la probabilità che ogni singola soluzione sia corretta e infine propone una o più diagnosi e piani terapeutici alternativi, associando ad ogni opzione una percentuale indicativa del grado di certezza. Dall'analisi LOA di Watson è risultato un livello di automazione C2; l'utente (medico/personale sanitario) può

scegliere una delle soluzioni proposte dal sistema o ignorarle ed elaborare autonomamente una diagnosi e un corrispondente piano terapeutico.

È necessario distinguere due ipotesi:

Ipotesi 1: Se il sistema ha generato diagnosi e/o piani terapeutici errati e l'utente (medico/personale sanitario) ha scelto una delle soluzioni restituite dal sistema, la responsabilità potrà essere attribuita:

(a) all'utente, solo nell'ipotesi in cui il sistema sia considerato non affidabile, per esempio nel caso in cui Watson abbia restituito diagnosi errate in un numero elevato di casi precedenti. Se il sistema è considerato affidabile, dovrebbe poter essere applicato il principio di affidamento, quale limite alla responsabilità del medico e del personale sanitario, in base a quanto esposto nelle sezioni precedenti[70];

(b) al produttore, per un errore nella fase di progettazione dei criteri di raccolta, classificazione e assegnazione di un grado percentuale di certezza delle diagnosi e relativi piani terapeutici.

(c) al sistema, qualora avesse personalità giuridica e fosse in grado di soddisfare i requisiti della *mens rea*, tenendo conto dei possibili scenari costruiti nella sezione 4.5.

Ipotesi 2: Se il sistema ha generato diagnosi e piani terapeutici corretti, la responsabilità potrà essere attribuita:

(a) esclusivamente all'utente (medico/personale sanitario), qualora il sistema sia considerato affidabile ed egli abbia ignorato le diagnosi e i piani terapeutici indicati.

4. **Fase di attuazione del piano di azione** Si tratta di un errore nella fase di esecuzione dell'azione. Watson è un sistema di supporto alla diagnostica e come evidenziato in sede di analisi delle caratteristiche del sistema, attualmente non è in grado di somministrare autonomamente il piano terapeutico ai pazienti. Dall'analisi LOA, è risultato un livello di automazione D0. L'utente (medico/personale sanitario) esegue e controlla tutte le azioni, e nel caso specifico la somministrazione del piano terapeutico, manualmente. La responsabilità potrà essere attribuita:

   (a) esclusivamente all'utente (medico/personale sanitario).

Dall'analisi dello scenario e dei possibili profili di responsabilità è evidente che, il verificarsi di un errore in una qualsiasi delle fasi iniziali o intermedie

[70] In particolare si vedano le sezioni 3.3.2 e 3.4.

potrebbe rimanere latente lungo l'intero processo di diagnosi, selezione e attuazione del piano terapeutico, ed emergere come errore attivo solo in sede di manifestazione di danno al paziente, creando un effetto domino, difficilmente controllabile[71]. In sede di analisi delle responsabilità, sarà necessario individuare la fase esatta in cui si è verificato l'errore, e il rischio è diventato prevedibile da parte di uno dei soggetti coinvolti, prendendo in considerazione l'intero ciclo di vita del sistema[72].

Ancora, l'uso della LOAT per determinare i livelli di automazione del sistema e la successiva analisi dei profili di responsabilità hanno evidenziato, per ogni singola fase considerata, che: (1) se il livello di automazione è uguale a zero, la responsabilità sarà sempre dell'utente; (2) se il livello di automazione è massimo, e dunque l'utente non interviene in alcuna fase del processo, la responsabilità si sposta completamente sul produttore e/o sul sistema; (3) per le ipotesi di automazione intermedia, dove uomo e macchina collaborano attivamente nell'esecuzione di un compito, il livello di incertezza, relativo alla causa dell'errore, aumenta, determinando un aumento del rischio di responsabilità sia per l'utente che per il programmatore e/o il sistema.

Il grafico qualitativo, presentato nella figura 4.3 mostra quanto appena evidenziato. Tali considerazioni restano valide sia che si consideri ogni fase singolarmente, sia che si consideri il processo nel suo insieme.

Nello scenario esaminato, occorre chiedersi anche, in che modo il ruolo di Watson debba essere inquadrato in area medica e più in particolare quali sono le possibili ripercussioni sulla relazione che intercorre tra medico e paziente.

I più comuni profili di responsabilità penale del medico concernono reati di tipo colposo, nello specifico l'omicidio colposo e le lesioni personali colpose. Nella categoria generale dell'omicidio colposo sono comprese quelle evenienze ad elevata complessità che rientrano nel novero generico della colpa medica[73]. Attualmente, le lesioni personali colpose risultano essere le forme delittuose in ambito sanitario con la più elevata frequenza. Tali fattispecie si caratterizzano per la non volontarietà dell'evento pregiudizievole, che si realizza in conseguenza della violazione delle *leges artis* mediche, ossia dell'obbligo di agire con diligenza, prudenza e perizia (colpa generica) ed in osservanza di leggi, regolamenti, ordini e discipline (colpa specifica). In particolare, la diligenza cui

[71] Si veda in particolare quanto considerato nella sezione 2.7, riguardo alla distinzione tra errore latente ed errore attivo.

[72] Si veda quanto considerato nelle sezioni da 3.3 a 3.4, in relazione alla responsabilità da dispositivo medico difettoso e alle ripercussioni in tema di responsabilità di tutti i soggetti coinvolti nel ciclo di vita del prodotto, con particolare riguardo alla responsabilità medica.

[73] Per tutti, Gentilomo, A., Travainni, G. e Luca, D. *Medico e Giustizia*. Raffaello Cortina Editore, 2009.

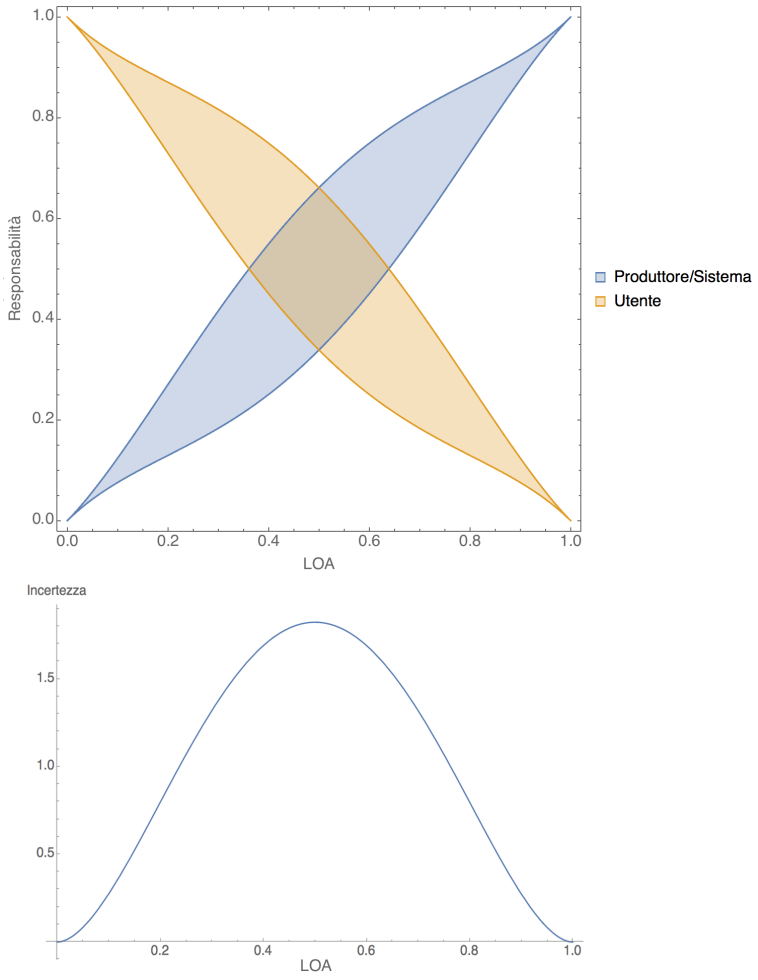


FIGURA 4.3: LOA, responsabilità e incertezza

è tenuto il medico nell'adempimento delle obbligazioni inerenti la propria attività professionale, non è semplicemente quella del buon padre di famiglia, ex art. 1176, comma 1, c.c., ma piuttosto quella qualificata, richiesta dalla natura dell'attività esercitata, ai sensi del secondo comma del medesimo articolo. L'espressione di tale diligenza qualificata, caratterizzata da un particolare sforzo tecnico-scientifico, è la perizia, intesa come conoscenza e applicazione di quel complesso di regole tecniche proprie della categoria professionale di appartenenza:le *leges* dell'arte medica, di natura cautelare, tese a perimetrare l'ambito del cosiddetto rischio consentito e, quindi, l'ambito di liceità dell'azione medica. La perizia di volta in volta si caratterizzerà in modi parzialmente diversi, sulla base del significato tecnico-qualitativo dello standard medio dello specialista di riferimento. In definitiva, ciò che si esige accanto ai generali doveri di diligenza e prudenza, è una perizia il cui contenuto è legato, da un lato, alle *leges artis* comuni a qualsiasi ramo della professione medica; dall'altro, alle regole di condotta specifiche del settore di specializzazione in cui opera il sanitario.

Lo standard di diligenza dovuto, diventa più complicato nel momento in cui un sistema ad elevata automazione, o d'intelligenza artificiale come Watson, si inserisce nella relazione medico-paziente[74]. In particolare, Watson è programmato per assumere il ruolo dello specialista a cui venga chiesto un consulto. La Corte di Cassazione, chiamata a giudicare sui profili di responsabilità del medico chiamato per un consulto specialistico, a fronte della responsabilità attribuita al medico che ha in cura il paziente, ha affermato che, i primi hanno gli stessi obblighi professionali dei medici che hanno in carico il paziente[75]. Poiché Watson è chiamato ad assumere il ruolo dello specialista a cui venga chiesta una

[74]Si veda quanto detto nella sezione 3.4 esaminando il caso delle cd. "valvole Killer", e in particolare, sentenza della sez. IV della Suprema Corte di Cassazione n. 40897/2011, in DeJure.

[75]In particolare, si vedano Cass., sezione IV, 18 dicembre 2009, n. 3365, in Cass. pen., 2011, 2586, con nota di A. Panetta. La disamina è stata condotta alla luce dei principi relativi alla responsabilità di equipe e del principio di affidamento. Secondo il primo principio, qualora ricorda l'ipotesi di cooperazione multidisciplinare, anche se svolta non contestualmente, come nel caso in esame, ciascun sanitario, oltre che al rispetto dei canoni di diligenza e prudenza connessi alle specifiche mansioni espletate, è tenuto ad osservare gli obblighi che derivano ad ogni membro dell'equipe dalla convergenza di tutte le attività verso il fine comune della cura del paziente. Secondo il principio di affidamento, ogni sanitario deve poter confidare che ciascuno si comporti adottando le regole precauzionali proprie del modello d'agente cui l'attività si riferisce. Si tratta di una regola finalizzata a contemperare il principio costituzionale della responsabilità personale con la specializzazione e la divisione dei compiti. In particolare, secondo la Cassazione "il principio di affidamento, non è, però, invocabile sempre e comunque, dovendo contemperarsi, con il concorrente principio della salvaguardia degli interessi del soggetto nei cui confronti opera la posizione di garanzia. Non è certamente invocabile, infatti, allorché l'altrui condotta colposa si innesti sull'inosservanza di una regola precauzionale proprio da parte di chi invoca il principio: ossia allorché l'altrui condotta colposa abbia la sua causa proprio nel non rispetto delle norme

consulenza, la questione del suo dovere verso i pazienti è particolarmente rilevante per eventuali profili di responsabilità. Watson è di vitale importanza per dare informazioni al medico curante, ma, come abbiamo evidenziato analizzando il sistema, non può ancora agire autonomamente, somministrando le terapie ai pazienti. Ciò considerato, potremmo escludere la presenza di una vera e propria relazione medico-paziente e una conseguente attribuzione di responsabilità per le ipotesi di lesioni colpose e/o omicidio colposo. Watson può solo raccomandare il piano terapeutico dopo la lettura e l'analisi della cartella clinica del paziente. La situazione diventerà più complicata, se e quando Watson potrà interfacciarsi direttamente con i pazienti e le attrezzature sanitarie. Diventerà necessario stabilire, se Watson debba essere considerato alla stregua del medico chiamato a consulto o meno, e nel primo caso sarà complicato stabilire un suo dovere professionale verso i pazienti.

Una questione ulteriore riguarda il consenso informato. Il dovere d'informazione, nell'ambito della professione medica, assume un rilievo fondamentale. Esso è intimamente legato al principio di autodeterminazione del paziente, quale regola fondamentale del rapporto tra lo stesso e il medico. Riguardo ad ogni intervento sanitario la manifestazione di volontà del paziente non può essere surrogata né disattesa, anche se per fini benefici. Lo specifico riferimento è all'art. 13 della Costituzione secondo cui "La libertà personale è inviolabile". Tale libertà, nel riconoscimento del diritto alla salute, si traduce in una pretesa, giuridicamente tutelata, di autodeterminazione del soggetto e di garanzia da ogni interferenza illegittima. Tale principio è valido nella maggior parte dei moderni ordinamenti. Watson, nel ruolo di medico di consulenza solleva questioni in relazione al consenso informato, poiché i pazienti dovranno essere informati che il medico sta utilizzando Watson come strumento diagnostico e che esso sta attivamente contribuendo alla diagnosi e alla scelta del piano terapeutico[76]. Il dovere di informazione del medico dovrebbe richiedere che il paziente sia informato dei risultati elaborati da Watson e della diagnosi e del piano terapeutico che il medico ha scelto di seguire. Ciò potrebbe portare a un maggiore disaccordo tra medico e paziente sulla valutazione della miglior linea terapeutica da seguire.

Comprendere come inquadrare giuridicamente Watson diviene fondamentale anche riguardo all'applicazione del principio di affidamento, quale limite alla responsabilità del medico. Qualora Watson sia considerato alla stregua di

cautelari, o specifiche o comuni, da parte di chi vorrebbe che quel principio operasse". Nello stesso senso, la sentenza pronunciata dalla IV sezione penale della Corte di Cassazione, n. 3365 del 26/01/2010.

[76]Il medico infatti, è tenuto ad informare il paziente circa eventuali alternative diagnostiche e terapeutiche, e di rischi e benefici prevedibili e ad essi legati, così da permettere al paziente di fare una valutazione e una scelta consapevole.

un semplice dispositivo medico, dovrà essere soggetto alla relativa disciplina e in particolare, considerato il livello di automazione e di rischio, alla procedura qualificata di certificazione CE che, come evidenziato, dovrebbe permettere l'applicazione del principio di affidamento in caso di danni derivanti da dispositivo medico difettoso. Qualora Watson sia considerato alla stregua di un medico chiamato a consulto, il medico che ha in carico il paziente potrebbe invocare il principio di affidamento, sia in relazione alla certificazione CE, sia in relazione al ruolo di Watson come titolare di una posizione di garanzia, come tale giuridicamente tenuto a impedire il verificarsi di un evento dannoso, ed essere esentato da responsabilità quando il danno derivi dall'esclusiva condotta di Watson, sulla correttezza del cui operato il medico abbia fatto legittimo affidamento. In quest'ultimo caso, e sulla base dello scenario costruito in questa sezione, Watson potrebbe essere penalmente responsabile, nel caso in cui l'errore si verifichi nella fase di decisione e selezione dell'azione, e in particolare nella prima ipotesi, e nella fase di attuazione del piano di azione nel caso in cui si interfacci direttamente con il paziente e le attrezzature sanitarie.

Infine, il coinvolgimento di Watson in un caso di lesioni colpose o di omicidio colposo potrebbe mutare il criterio di diligenza e perizia applicabile. Come più volte ricordato, Watson è in grado di accedere ad una straordinaria quantità di informazioni e conoscenza in ambito medico e linee guida della pratica medica basata su prove di efficacia, utilizzando molte più informazioni di quanto potrebbe fare il migliore degli specialisti. I medici che si avvalgono di Watson potrebbero essere tenuti ad un maggiore livello di diligenza rispetto a quello normalmente richiesto. Se i sistemi d'intelligenza artificiale iniziano a praticare la medicina, i tradizionali standard di diligenza medica dovranno necessariamente evolvere poiché essi, costruiti per gli esseri umani, risultano inadeguati se applicati ai sistemi d'intelligenza artificiale.

Conclusioni

5.1 Conclusioni

Dal presente lavoro di ricerca emerge, dunque, una lunga serie di spunti.

Siamo partiti inquadrando il sistema sanitario come sistema socio-tecnico, costituito da componenti di natura diversa, quali quelle umane, tecnologiche e procedurali che interagiscono tra loro, capaci di generare problemi complessi che richiedono necessariamente risposte complesse.

È stato evidenziato come, prevenire gli errori significhi ridisegnare il sistema e i processi di lavoro, per renderli più sicuri, attraverso un approccio sistemico all'errore e al rischio clinico, superando una concezione dell'errore umano che vede l'attore potenzialmente libero di agire e di violare o meno le regole del sistema, e sottolineando la necessità di guardare all'interazione tra l'attore e il sistema, per capire le ragioni che lo hanno indotto a commettere l'errore. Il comportamento umano e la ricerca del colpevole smettono così di avere un ruolo centrale, per fare spazio alle condizioni in cui si verifica l'errore e alla ricerca delle cause di fallimento del sistema.

La riduzione dei danni derivanti dai processi di cura procede di pari passo al riconoscimento e alla prevenzione degli errori. Uno dei principali limiti del sistema di gestione del rischio clinico è quello di non riuscire a fornire analisi quantitative, in grado di garantire la rappresentatività del campione, variabile in base alla capacità degli operatori del settore di riconoscere gli eventi e alla volontà di renderli pubblici. Talvolta è molto difficile identificare anche le dimensioni del contesto cui riportare il numero di segnalazioni. Un elemento critico, relativo al *risk management* e ai sistemi di *Incident Reporting*, siano essi volontari o obbligatori, è l'incapacità di raggiungere a breve termine i propri obiettivi. Tale incapacità sembra essere riconducibile ai sistemi usati, raramente semplici da utilizzare, ambigui nella definizione degli eventi accaduti e di quelli da segnalare, spesso molto costosi in termini di tempo e denaro e a rischio di aumentare il contenzioso tra ente ospedaliero e paziente. La segnalazione degli eventi avversi è realmente praticabile solo qualora l'organizzazione che la

richiede dimostri di saperne cogliere la connotazione positiva. A tal fine, è certamente auspicabile un cambiamento culturale sul concetto di errore in sanità, riconoscendo in esso un'opportunità di apprendimento e miglioramento, contrastando il prevalente atteggiamento punitivo, che è uno dei principali motivi del fallimento delle politiche e delle strategie per la sicurezza nei sistemi sanitari. L'atteggiamento punitivo ostacola la segnalazione degli eventi avversi e dei *near misses*, impedendone di fatto la segnalazione "libera da rimprovero", in assenza di una politica coerente all'interno dell'organizzazione per la gestione confidenziale dei dati. Per questo motivo la legislazione vigente italiana dovrebbe essere aggiornata, sul modello di quanto fatto dai Governi australiano e danese, che vietano di utilizzare i dati delle segnalazioni sugli eventi avversi e i *near misses* a scopi giudiziari, a salvaguardia del principio generale del segreto professionale.

Dall'esame delle tecnologie, in uso o in fase di sperimentazione, nel settore sanitario, è stata evidenziata la difficoltà di ricomprenderle tutte all'interno di un'unica categoria concettuale, capace di definirne le caratteristiche in modo univoco. Alcune di esse si caratterizzano principalmente come strumenti di ausilio per l'operatore umano, altre, pur mantenendo tale caratteristica, si atteggiano come agenti che operano autonomamente all'interno dell'ambiente. Tali proprietà si riflettono inevitabilmente sia sui livelli di automazione dei singoli sistemi, sia sui profili di responsabilità che derivano dal loro impiego. Abbiamo avuto modo di notare come, una larga parte di tali tecnologie, già in uso o in fase di sperimentazione, non siano dispositivi nati, progettati e fabbricati per finalità primariamente mediche, e come da ciò possa derivare, talvolta, la difficoltà di ricomprenderli nell'alveo dei dispositivi medici, come classificati dalle direttive CEE, recepite nel nostro ordinamento ex d.lgs. 14 dicembre 1992, n. 507 e successive modifiche, e la necessità di adeguamento nel rispetto delle specifiche norme di settore, con particolare riguardo alla destinazione d'uso del prodotto che deve essere in ogni caso caratterizzata dalla finalità medica, secondo le linee guida comunitarie MEDDEV. Ancora, dall'analisi della Direttiva 93/42/CEE e del d.lgs di attuazione 24 febbraio 1997, n. 46, il cui obiettivo dichiarato risiede nell'esigenza di garantire un elevato livello di protezione e sicurezza contro i rischi per la salute di tutti coloro che vengono in contatto con dispositivi medici, è emersa la mancanza tra i criteri utilizzati per classificare i dispositivi medici in fasce di rischio, il livello di automazione dei DM. In particolare le variabili tecniche, che scaturiscono dalle modalità operative e applicative di cui si avvale il lavoro umano, e di cui fanno parte le tecnologie, non possono non essere prese in considerazione tra i criteri utilizzati per la classificazione in termini di rischio. Il rischio da ignoto tecnologico, è un rischio generalizzato che coinvolge tutte le componenti del sistema sanità è non può non rilevare in questa sede. Una visione socio-tecnica del rischio non può prescindere da un analisi che tenga conto

dell'interazione tra uomo e tecnologie e dunque di una tassonomia dei livelli di automazione, potenzialmente capace di incidere, a sua volta, sulla disciplina di certificazione, sul danno e sui profili di responsabilità.

Un aspetto centrale nell'analisi dei profili di responsabilità è certamente legato alle procedure di certificazione dei *medical device*, che variano in base alla classe di rischio del dispositivo medico considerato, e in particolare al sistema di certificazione di garanzia della qualità da parte dell'Organo Notificato. Tale sistema non può quindi che riverberarsi sul valore da attribuire alle garanzie di sicurezza del dispositivo che si riveli difettoso e fonte di danno. In particolare, in fase di accertamento di responsabilità per danni derivanti dal malfunzionamento del prodotto, dovrà tenersi necessariamente conto del ruolo dell'organo notificato e verificare l'eventuale sussistenza, in capo al medesimo, di eventuali profili di responsabilità penale, e valutare come l'apporto dell'Organismo Notificato e l'apposizione della certificazione CE, incidano suì profili di responsabilità altrui.

Abbiamo evidenziato come la responsabilità medica, le tecnologie ad elevata automazione e i pericoli ad essa intrinsechi, e il fenomeno del danno da prodotto difettoso si intersechino dando vita ad un'area che si caratterizza come una sorta di campo gravitazionale capace di attirare tutti i problemi in materia di responsabilità plurisoggettiva e colpevolezza. L'attività di produzione, distribuzione, commercio e utilizzo di dispositivi medici, coinvolge una pluralità di soggetti, dalla nascita alla morte del prodotto, e più in particolare, il produttore, l'organismo certificatore, e infine l'utilizzatore, confermando il potenziale apporto plurisoggettivo allo svolgimento di attività rischiose, che richiedono livelli di conoscenza molto elevati da parte degli operatori, e in cui l'evoluzione scientifica costante, può assumere un valore determinante in sede di giudizio.

Nei casi in cui sia rilevabile un difetto del *medical device* è proprio l'elemento della riconoscibilità della regola cautelare prima, e del difetto successivamente, ad assumere un ruolo discriminante nell'attribuzione della responsabilità penale a titolo di colpa. A ciò si aggiunga che la riconoscibilità del difetto dovrà essere declinata, almeno nel *quantum*, in relazione al soggetto di volta in volta considerato e alle competenze e capacità che tali soggetti hanno o dovrebbero avere in base alla relativa figura di agente modello.

In particolare, è stata approfondita la posizione del medico in relazione all'organismo notificato e al ruolo da attribuire alla certificazione di conformità CE, in sede di attribuzione di eventuali profili di responsabilità penale colposa, derivanti dal malfunzionamento del *medical device*. In base al principio dell'equivalenza delle cause vigente nel nostro ordinamento, un determinato fatto lesivo è ascrivibile a tutti coloro che pongono in essere una condotta che rappresenti,

nella progressione causale, un antecedente necessario al verificarsi del fatto medesimo. Il momento di inizio di tale progressione causale va individuato, certamente, nella commercializzazione dei prodotti, coinvolgendo tutti i soggetti cui sopra si è fatto riferimento.

In caso di danno derivante dall'uso di un dispositivo medico difettoso marcato CE, e con riferimento alla posizione del medico utilizzatore, la scelta tra pronuncia di condanna a titolo di concorso, cooperazione colposa, o assoluzione in virtù del principio di affidamento, quale limite all'imputazione dell'evento a titolo di colpa, dipenderà dal valore attribuito al parametro di riconoscibilità (a) del difetto, e/o (b) del pericolo e del dovere di prevedere e riconoscere inosservanze altrui. In particolare, è stato evidenziato come nelle fattispecie di omicidio o lesioni, che derivino da un difetto di un dispositivo certificato CE, debba potersi applicare il principio di affidamento, quale limite alla responsabilità penale del medico, salvi i casi in cui il difetto sia palese e riconoscibile. L'orientamento secondo cui la certificazione rilasciata da un Organismo notificato autorizzato avrebbe valenza puramente formale, e non sostanziale, escludendo così l'efficacia scusante della marcatura CE, nei confronti di distributori e medici che abbiano messo in commercio e utilizzato un dispositivo difettoso, è da escludere anche a fronte di alcune recenti sentenze pronunciate dalla Corte di Cassazione. Più precisamente, e in linea con quanto stabilito dalla Suprema Corte di Cassazione nella sentenza n. 40897/2011 e in conformità al principio di affidamento, si ritiene di poter attribuire efficacia sostanziale alla certificazione CE. Questa, per altro, è l'unica strada percorribile per poter garantire il rispetto del principio di responsabilità penale personale ex art. 27, comma 1 Cost., escludendo la sussistenza del dovere di riconoscere la negligenza altrui anche nei casi in cui ciò non sia in alcun modo prevedibile, per la sola interconnessione soggettiva che si crea in certi settori. In questo modo, si garantisce che il giudizio sul rischio e la prevedibilità ed evitabilità dell'evento, intimamente legato al principio di affidamento, come limite all'attribuzione di responsabilità a titolo di colpa, sia interpretato correttamente e in linea con il principio di colpevolezza. La responsabilità penale può aversi solo qualora siano posti in essere atti ed eventi, la cui realizzazione rientri nella sfera di controllo del soggetto agente.

Come già evidenziato analizzando le tecnologie in uso nel settore sanitario, sono emerse proprietà estremamente eterogenee e differenti tra i sistemi utilizzati in ambito medico. Alcune di tali tecnologie possono certamente essere ricomprese nella nozione di robot, inteso come sistema dotato di alcune funzioni essenziali, quali la capacità di agire su stimoli ambientali in combinazione con rilevamento e ragionamento logico, la cui funzione principale risiede nell'automazione del lavoro fisico, come per esempio i robot chirurgici. Tuttavia restano

fuori da tale definizione tutti i sistemi privi di caratteristiche meccatroniche, come per esempio sistemi esperti e altri sistemi d'intelligenza artificiale. Inoltre, la maggior parte dei sistemi d'intelligenza artificiale è in grado di operare autonomamente senza l'ausilio di esseri umani, nello svolgimento dei propri compiti, per esempio Watson, mentre la maggior parte dei sistemi robotici in area sanitaria, per esempio il Da Vinci, hanno bisogno di un operatore umano per poter espletare tali compiti. Alcune tecnologie si caratterizzano principalmente come strumenti di ausilio per l'operatore umano, altri, pur mantenendo tale caratteristica, si atteggiano come agenti capaci di operare autonomamente all'interno di un ambiente predeterminato. In tema di responsabilità penale legata all'uso di sistemi d'intelligenza artificiale, abbiamo rilevato come, il rapido sviluppo di tali tecnologie ponga la necessità urgente di adottare misure, capaci di proteggere e garantire la sicurezza della società da possibili danni agli interessi protetti dal diritto penale, siano essi individuali o collettivi. Abbiamo esaminato diversi possibili approcci elaborati dalla dottrina, rilevando come, in alcuni contesti specifici, ci riferiamo in particolare a quello delle armi autonome intelligenti, l'autonomia degli AI debba necessariamente essere limitata, a fronte dei rischi che essa comporta e delle potenziali ripercussioni disastrose in termini di vite umane. Il modello zoologico, pur valido per inquadrare alcune tipologie di sistemi d'intelligenza artificiale, appare in certi casi inadeguato poiché, a fronte di livelli di autonomia elevati e capacità cognitive e volitive sviluppate, non propone soluzioni efficaci per arginare la possibile commissione di reati da parte di agenti autonomi intelligenti. Tale modello si limita a regolare i rapporti tra AI e persone fisiche o giuridiche e a far ricadere gli effetti di eventuali condotte criminose all'interno della sfera giuridica di programmatori e proprietari e tuttavia, possono verificarsi casi in cui un AI ponga in essere una condotta astrattamente riconducibile a una fattispecie criminosa, senza che utente e produttore possano essere ritenuti penalmente responsabili. Come evidenziato, anche se in linea teorica la legge potrebbe disciplinare il comportamento degli AI allo stesso modo in cui disciplina quello degli animali, sembra ormai evidente la necessità di prepararsi ad una nuova classe di azioni giuridicamente rilevanti e difficilmente riconducibili agli animali.

Ancora, la responsabilità penale delle persone giuridiche è senza dubbio un enorme passo in avanti nell'area del diritto penale e i modelli utilizzati per sostenere tale responsabilità rivelano preziosi indizi per costruire un quadro dogmatico plausibile per un eventuale responsabilità penale degli AI.

Come abbiamo evidenziato cercando di stabilire la possibilità che certi agenti autonomi siano capaci di soddisfare i requisiti dell'*actus reus* e della *mens rea*, il diritto penale mostra un certo grado di flessibilità, qualora per ragioni di politica criminale risulti necessaria e siano soddisfatte certe premesse dogmatiche. Pur

ammettendo la possibilità di riconoscere personalità giuridica agli AI, dotandoli di un proprio patrimonio, e constatando che certi agenti siano in grado di soddisfare i requisiti dell'elemento oggettivo e soggettivo del reato, sembra che manchi un pezzo del puzzle per poter applicare il diritto penale agli AI. Quale utilità avrebbe applicare sanzioni penali agli agenti autonomi intelligenti? Sia che si abbracci una definizione di reato di natura puramente formale, inteso come fatto cui la legge ricollega una sanzione penale, cosicché il concetto di reato è determinato esclusivamente in funzione delle conseguenze giuridiche (pena o misura di sicurezza) che il legislatore riconnette a tali fatti; sia che si abbracci una definizione di reato di tipo sostanziale, secondo cui è reato un fatto che aggredisce un bene giuridico meritevole di protezione, sempreché l'aggressione sia tale da far apparire inevitabile il ricorso alla pena e che sanzioni di tipo non penale siano insufficienti a garantire un'efficace tutela, un giudizio di colpevolezza di un AI non raggiungerebbe alcun risultato utile per la società o per la vittima del crimine. L'AI non subirebbe alcun effetto deterrente della condanna, né potrebbe percepirne il disvalore. Tali aspetti sono indispensabili affinché la sanzione penale possa esplicare le proprie funzioni. Dunque perché dovremmo applicare il diritto penale agli AI, venendo meno i presupposti politico criminali del sistema sanzionatorio? Ad oggi, ci sembra che la soluzione migliore, per regolare il possibile emergere di un fenomeno criminoso legato alle condotte di agenti autonomi intelligenti, sia quella di adottare la teoria degli agenti normativi. Come abbiamo evidenziato, sia che si tratti di agenti cd. *norm follower*, sia che si tratti di agenti capaci di "violazioni intelligenti" delle norme, queste ultime hanno un ruolo e un'efficacia nella standardizzazione dei comportamenti di un AI. In base all'ambiente in cui gli agenti si troveranno ad operare sarà possibile decidere quale dei due approcci sia il più adatto. Qualora si verifichi una violazione non giustificabile, e non sia riscontrabile un comportamento doloso o colposo del programmatore e/o dell'utente, non sarà necessario ricorrere a sanzioni di tipo penale, ma intervenire direttamente sul sistema ed evitare la possibilità che, nelle medesime circostanze, si verifichi nuovamente.

Infine, dall'analisi dello scenario costruito intorno a Watson, è emerso come il verificarsi di un errore in una qualsiasi delle fasi iniziali o intermedie, sia in grado di rimanere latente lungo l'intero processo di diagnosi, selezione e attuazione del piano terapeutico, ed emergere come errore attivo solo in sede di manifestazione di danno al paziente, creando un effetto domino, difficilmente controllabile. L'uso della LOAT per determinare i livelli di automazione del sistema e la successiva analisi dei profili di responsabilità hanno evidenziato, per ogni singola fase considerata, che: se il livello di automazione è pari a zero, la responsabilità sarà sempre dell'utente; se il livello di automazione è massimo,

e dunque l'utente non interviene in alcuna fase del processo, la responsabilità si sposta completamente sul produttore e/o sul sistema; e per le ipotesi di automazione intermedia, dove uomo e macchina collaborano attivamente nell'esecuzione di un compito, il livello di incertezza, relativo alla causa dell'errore, aumenta, determinando un aumento del rischio di responsabilità sia per l'utente che per il programmatore e/o il sistema. Ci siamo chiesti in che modo il ruolo di Watson debba essere inquadrato in area medica e più in particolare quali siano le possibili ripercussioni sulla relazione che intercorre tra medico e paziente, quali le ripercussioni derivanti dall'uso di tale sistema sugli standard di diligenza e perizia richiesti nell'attività medica. Stabilire lo standard di diligenza dovuto, diventa certamente più complicato nel momento in cui un sistema d'intelligenza artificiale come Watson, si inserisce nella relazione medico-paziente. Come più volte ricordato, Watson è in grado di accedere ad una straordinaria quantità di informazioni e conoscenza in ambito medico e linee guida della pratica medica basata su prove di evidenza, utilizzando molte più informazioni di quanto potrebbe fare il migliore degli specialisti. I medici che si avvalgono di Watson potrebbero essere tenuti ad un maggiore livello di diligenza rispetto a quello normalmente richiesto. Se i sistemi d'intelligenza artificiale iniziano a praticare la medicina, i tradizionali standard di diligenza medica dovranno necessariamente evolvere, poiché essi, costruiti per gli esseri umani, risultano inadeguati se applicati ai sistemi d'intelligenza artificiale.

Abbiamo evidenziato come, la necessità di inquadrare giuridicamente Watson divenga fondamentale anche riguardo all'applicazione del principio di affidamento, quale limite alla responsabilità del medico. Qualora Watson sia considerato alla stregua di un semplice dispositivo medico, dovrà essere soggetto alla relativa disciplina e in particolare, considerato il livello di automazione e di rischio, alla procedura qualificata di certificazione CE che, come evidenziato, dovrebbe permettere l'applicazione del principio di affidamento in caso di danni derivanti da dispositivo medico difettoso. Qualora Watson sia considerato alla stregua di un medico chiamato a consulto, il medico che ha in carico il paziente potrebbe invocare il principio di affidamento, sia in relazione alla certificazione CE, sia in relazione al ruolo di Watson come titolare di una posizione di garanzia, come tale giuridicamente tenuto a impedire il verificarsi di un evento dannoso, ed essere esentato da responsabilità, quando il danno derivi dall'esclusiva condotta di Watson, sulla correttezza del cui operato il medico abbia fatto legittimo affidamento.

Chi scrive, spera in questo modo di aver contribuito, almeno in parte, all'avanzamento dello stato dell'arte in materia di responsabilità penale e uso di sistemi d'intelligenza artificiale nell'e-health.

Bibliografia

Abbott, J. J., Hager, G. D. e Okamura, A. M. «Steady-hand teleoperation with virtual fixtures». In: *Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on*. IEEE. 2003, pp. 145–151.

Agarwal, R. et al. «The RoboConsultant: telementoring and remote presence in the operating room during minimally invasive urologic surgeries using a novel mobile robotic interface». In: *Urology* 70.5 (2007), pp. 970–974.

Alcobendas-Maestro, M. et al. «Lokomat robotic-assisted versus overground training within 3 to 6 months of incomplete spinal cord lesion randomized controlled trial». In: *Neurorehabilitation and neural repair* 26.9 (2012), pp. 1058–1063.

Aleo, S. «Causalità, complessità e funzione penale». In: *Per un'analisi funzionalistica della responsabilità penale* (2003).

Aleo, S., Centonze, A. e Lanza, E. *La responsabilità penale del medico*. Giuffrè Editore, 2007.

Allen, C., Varner, G. e Zinser, J. «Prolegomena to any future artificial moral agent». In: *Journal of Experimental & Theoretical Artificial Intelligence* 12.3 (2000), pp. 251–261.

Alpa, G. «La responsabilità medica». In: *Resp. civ. prev* (1999), pp. 315–336.

Amati, E. *Introduzione al diritto penale internazionale*. Vol. 2. Giuffrè Editore, 2010.

AMBROSE, R. et al. «WTEC Panel Report on International Assessment of Research and Development in Robotics.» In: (2007).

Amirabdollahian, F. et al. «Multivariate analysis of the Fugl-Meyer outcome measures assessing the effectiveness of GENTLE/S robot-mediated stroke therapy». In: *Journal of NeuroEngineering and Rehabilitation* 4.1 (2007), pp. 1–16.

Arthanat, S., Desmarais, J. M. e Eikelberg, P. «Consumer perspectives on the usability and value of the iBOT® wheelchair: findings from a case series». In: *Disability and Rehabilitation: Assistive Technology* 7.2 (2012), pp. 153–167.

ASR, R. E. R. *FMEA-FMECA analisi dei modi di errore/guasto e dei loro effetti nelle organizzazioni sanitarie*. 2002.

Association, C. M., Association, C. H. et al. *Report on the medical insurance feasibility study*. California Medical Association, 1977.

Azzali, G. «La responsabilità penale del produttore per danni alla salute». In: *Rivista trimestrale di diritto penale dell'economia* (1991), p. 848.

Baker, S. *Final Jeopardy: man vs. machine and the quest to know everything*. Houghton Mifflin Harcourt New York, NY, 2011.

Ballantyne, G. H. e Moll, F. «The da Vinci telerobotic surgical system: the virtual operative field and telepresence surgery». In: *Surgical Clinics of North America* 83.6 (2003), pp. 1293–1304.

Barbieri, B. et al. «An expert system for the oral anticoagulation treatment». In: *Innovations in Applied Artificial Intelligence*. Springer, 2005, pp. 773–782.

Beck, U. «La società del rischio, trad. it». In: *Roma, Carocci* (2000).

Bengio, Y. et al. «Neural probabilistic language models». In: *Innovations in Machine Learning*. Springer, 2006, pp. 137–186.

Berkelmans, R., Arns, M. e Duysens, J. «The development of a hybrid outdoor FES bike». In: *Proc. of the 8th Annual Conference of the International Functional Electrical Stimulation Society (IFESS 2003)*. 2003.

Bernardi, A. «La responsabilità da prodotto nel sistema italiano: profili sanzionatori». In: *Rivista Trimestrale di Diritto Penale dell'Economia. Padova* 1-2 (2003), pp. 1–45.

Bilancetti Mauro e Bilancetti, F. *La responsabilità civile e penale del medico*. Cedam, Padova, 2010.

Billings, C. E. «Human-centered aircraft automation: A concept and guidelines». In: (1991).

Bloss, R. «Mobile hospital robots cure numerous logistic needs». In: *Industrial Robot: An International Journal* 38.6 (2011), pp. 567–571.

Boden, M. A. *The Philosophy of Artificial Intelligence*. 1990.

Boella, G., Van Der Torre, L. e Verhagen, H. «Introduction to Normative Multiagent Systems». In: *Dagstuhl Seminar Proceedings*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2007.

Booster, C. «ROOT CAUSE ANALYSIS». In: *Joint Commission Perspectives on Patient Safety* 3.5 (2003).

Borruso, R. *Computer e diritto*. Computer e diritto v. 1. A. Giuffrè, 1988.

Bratman, M. «Intention, plans, and practical reason». In: (1987).

Bricola, F. «Responsabilità penale per il tipo e per il modo di produzione». In: *La responsabilità dell'impresa per i danni all'ambien-te e ai consumatori, Milano* 87 (1978).

Bright, J. R. *Automation and management*. Division of Research, Graduate School of Business Administration, Harvard University Boston, 1958.

Bucchi, M. «M. Catino," Da Chernobyl a Linate. Incidenti tecnologici o errori organizzativi?", 2002». In: *Rassegna Italiana di Sociologia* 45.2 (2004), pp. 299–300.

Buchanan, B. G., Shortliffe, E. H. et al. *Rule-based expert systems*. Vol. 3. Addison-Wesley Reading, MA, 1984.

Burgar, C. G. et al. «Development of robots for rehabilitation therapy: the Palo Alto VA/Stanford experience». In: *Journal of rehabilitation research and development* 37.6 (2000), pp. 663–674.

Butera, F. *Il castello e la rete. Impresa, organizzazioni e professioni nell'Europa degli anni'90*. Vol. 1. FrancoAngeli, 2005.

— «Note sulla storia dell'automazione. Dall'impatto sociale dell'automazione alla progettazione congiunta di tecnologia, organizzazione e sviluppo delle persone». In: *STUDI ORGANIZZATIVI* (2014).

Butera, F. e De Witt, G. «Valorizzare il lavoro per rilanciare l'impresa». In: *La storia delle isole di produzione alla Olivetti negli anni'70* (2011).

Casa, F. «Dalle scienze cognitive alle applicazioni giuridiche dell'intelligenza artificiale / Federico Casa.» In: ().

Castelfranchi, C. et al. «Deliberative normative agents: Principles and architecture». In: *Intelligent agents VI. Agent theories, architectures, and languages*. Springer, 1999, pp. 364–378.

CASTRONUOVO, D. «Responsabilità da prodotto e struttura del fatto colposo». In: *Riv. it. dir. proc. pen*. Vol. 301. 2005.

Centonze, F. *La normalità dei disastri tecnologici: il problema del congedo dal diritto penale*. Giuffrè, 2004.

Cherns, A. «The Principles of Sociotechnical Design1». In: *Human relations* 29.8 (1976), pp. 783–792.

Cheung, E. et al. «A new endoscopic microcapsule robot using beetle inspired microfibrillar adhesives». In: *Advanced Intelligent Mechatronics. Proceedings, 2005 IEEE/ASME International Conference on*. IEEE. 2005, pp. 551–557.

Cinotti, R., Basini, V. e Di Denia, P. «Il sistema di incident reporting nelle organizzazioni sanitarie». In: *Collana Dossier* 86 (2003).

Cipolla, P. «Profili penali del contrassegno CE». In: *Giurisprudenza di merito* 10, 2133 (2012).

Clark, J. V. e Krone, C. G. «Towards an overall view of organizational development in the early seventies». In: *Thomas, JM/Bennis, WG: The Management of change and conflict. Harmondsworth. S. 284f* (1972).

Cobb, J. et al. «Hands-on robotic unicompartmental knee replacement A PROSPECTIVE, RANDOMISED CONTROLLED STUDY OF THE ACROBOT SYSTEM». In: *Journal of Bone and Joint Surgery, British Volume* 88.2 (2006), pp. 188–197.

Cohn, J. «ntroduction to special issue: Robotic assistance in neuro-motor therapy». In: *Robotica* 21 (2003).

Connell, L. «Statement before the subcommittee on oversight and investigations, Committee on Veterans' Affairs». In: *Washington, DC: US House of Representatives* (2000).

CONTI, D. «SOCIALLY ASSISTIVE ROBOTICS UNA POSSIBILE UNIONE TRA ROBOTICA E PSICOLOGIA». In: ().

Coote, S. et al. «The effect of the GENTLE/s robot-mediated therapy system on arm function after stroke». In: *Clinical rehabilitation* 22.5 (2008), pp. 395–405.

Crossman, E. *Automation and skill, dsir, Problems of Progress in Industry (9), London. Reprinted in Edwards et Lees F.(eds), The Human Operator in Process Control*. 1974.

Dario, P. e Bergamasco, M. «An advanced robot system for automated diagnostic tasks through palpation». In: *IEEE Transactions on Biomedical Engineering* 35.2 (feb. 1988), pp. 118–126.

Dautenhahn, K. «Robots as social actors: Aurora and the case of autism». In: *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco*. Vol. 359. 1999, p. 374.

Davies, B. «A review of robotics in surgery». In: *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 214.1 (2000), pp. 129–140.

Davis, L. E. «Evolving alternative organization designs: their sociotechnical bases». In: *Human Relations* 30.3 (1977), pp. 261–273.

Degani, A. *Taming HAL: Designing interfaces beyond 2001*. Palgrave Macmillan, 2004.

Delp, S. L. et al. «Computer assisted knee replacement.» In: *Clinical orthopaedics and related research* 354 (1998), pp. 49–56.

Dennett, D. C. «Evolution, error, and intentionality». In: *Contemporary Materialism* (1986), p. 254.

DeRosier, J. et al. «Using health care failure mode and effect analysis^{TM}: the VA National Center for Patient Safety's prospective risk analysis system». In: *The Joint Commission Journal on Quality and Patient Safety* 28.5 (2002), pp. 248–267.

DIRECTIVA. «90/385/CEE e successiva modifica 2007/47/CE». In: *Dispositivo medico classe II a* (1990).

— «93/42/CEE e successiva modifica 2007/47/CE». In: *Dispositivo medico classe II a* (2005).

Dolghi, O. et al. «Miniature in vivo robot for laparoendoscopic single-site surgery». In: *Surgical endoscopy* 25.10 (2011), pp. 3453–3458.

Dressler, J. *Cases and materials on criminal law*. West Group Publishing, 2007.

Dreyfus, H. L. *What computers can't do: The limits of artificial intelligence*. Vol. 1972. Harper & Row New York, 1979.

Duff, R. A. «Intention, agency and criminal liability: Philosophy of action and the criminal law». In: *Intention, Agency and Criminal Liability: Philosophy of Action and the Criminal Law, Blackwell* (1990).

Duni, G. «L'obbligo di prevedere le condotte altrui». In: *Rivista giuridica della circolazione e dei trasporti* (1964).

Eliakim, R. et al. «Evaluation of the PillCam Colon capsule in the detection of colonic pathology: results of the first multicenter, prospective, comparative study.» In: *Endoscopy* 38.10 (2006), pp. 963–970.

Emery, F. e Trist, E. «Sistemi socio-tecnici». In: *Progettazione e sviluppo delle organizzazioni* (1974).

Emery, F. E., Trist, E. L. et al. «The causal texture of organizational environments». In: *Human relations* 18.1 (1965), pp. 21–32.

Endsley Mica R e Garald, D. «Theoretical underpinnings of situation awareness: A critical review». In: *Situation awareness analysis and measurement* (2000), pp. 3–32.

Endsley, M. R. «Automation and situation awareness». In: *Automation and human performance: Theory and applications* (1996), pp. 163–181.

— «Level of automation effects on performance, situation awareness and workload in a dynamic control task». In: *Ergonomics* 42.3 (1999), pp. 462–492.

— «The application of human factors to the development of expert systems for advanced cockpits». In: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Vol. 31. 12. SAGE Publications. 1987, pp. 1388–1392.

Endsley, M. R. e Kiris, E. O. «The out-of-the-loop performance problem and level of control in automation». In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.2 (1995), pp. 381–394.

Enzo, F. G.-.-.-M. *Diritto penale, parte generale*. 2009.

Esteves, J. e Joseph, R. C. «A comprehensive framework for the assessment of eGovernment projects». In: *Government information quarterly* 25.1 (2008), pp. 118–132.

Ferrari, E., Robins, B. e Dautenhahn, K. «Therapeutic and educational objectives in robot assisted play for children with autism». In: *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*. IEEE. 2009, pp. 108–114.

Ferrucci, D. et al. «Building Watson: An overview of the DeepQA project». In: *AI magazine* 31.3 (2010), pp. 59–79.

Fiandaca Giovanni e Musco, E. «Diritto penale. Parte generale». In: *Bologna, Zanichelli* (2004).

Finocchiaro, G. *I contratti ad oggetto informatico*. Cedam, 1993.

Fitts, P. M. «Human engineering for an effective air-navigation and traffic-control system.» In: (1951).

Forti, G. *Colpa ed evento nel diritto penale*. Giuffrè, 1990.

Frank, S. J. «Tort Adjudication and the Emergence of Artificial Intelligence Software». In: *Suffolk UL Rev.* 21 (1987), p. 623.

Freitas, P. M., Andrade, F. e Novais, P. «Criminal Liability of Autonomous Agents: From the Unthinkable to the Plausible». In: *AI Approaches to the Complexity of Legal Systems*. Springer, 2014, pp. 145–156.

Gabbai, J. M. «Complexity and the aerospace industry: Understanding emergence by relating structure to performance using multi-agent systems». Tesi di dott. Citeseer, 2005.

Garcia, E., Sater, J. M. e Main, J. «Exoskeletons for Human Performance Augmentation (EHPA): A Program Summary.» In: *The Robotics Society of Japan* 20.8 (2002), pp. 822–826.

Gentilomo, A., Travainni, G. e Luca, D. *Medico e Giustizia*. Raffaello Cortina Editore, 2009.

Gerstner, M. E. «Liability issues with artificial intelligence software». In: *Santa Clara L. Rev.* 33 (1993), p. 239.

Ghahramani, Z. «Probabilistic machine learning and artificial intelligence». In: *Nature* 521.7553 (2015), pp. 452–459.

Gillies, P. *The Law of criminal complicity*. Law Book Company, 1980.

Girone, M. et al. «Orthopedic rehabilitation using the" Rutgers ankle" interface». In: *Studies in health technology and informatics* (2000), pp. 89–95.

Giulianotti, P. C. et al. «Robotics in general surgery: personal experience in a large community hospital». In: *Archives of surgery* 138.7 (2003), pp. 777–784.

HALEY, R. W. et al. «The efficacy oe infection surveillance and control programs in preventing nosocomial infections in us hospitals». In: *American journal of epidemiology* 121.2 (1985), pp. 182–205.

Halford, G. S. et al. «How many variables can humans process?» In: *Psychological science* 16.1 (2005), pp. 70–76.

Hallevy, G. *The Criminal Liability of Artificial Intelligence Entities*. Akron Intellectual Property, 2010.

Hallevy, G. «Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control». In: *Akron Intell. Prop. J.* 4 (2010), p. 171.

— «I, Robot—I, Criminal"—When Science Fiction Becomes Reality: Legal Liability of AI Robots Committing Criminal Offenses». In: *Syracuse Sci. & Tech. L. Rep.* 22 (2010), pp. 1–9.

— *Liability for Crimes Involving Artificial Intelligence Systems*. Springer, 2014.

— «Unmanned vehicles–Subordination to criminal law under the modern concept of criminal liability». In: (2012).

Hallevy, G. *When Robots Kill: Artificial Intelligence Under Criminal Law*. Northeastern University, 2013.

Harwin, W. S., Patton, J. L. e Edgerton, V. R. «Challenges and Opportunities for Robot-Mediated Neurorehabilitation». In: *Proceedings of the IEEE* 94.9 (set. 2006), pp. 1717–1726.

Harwin, W. e Hillman, M. «Introduction». In: *Robotica* 21 (01 gen. 2003), pp. 1–1.

Heckerman, D. E. e Shortliffe, E. H. «From certainty factors to belief networks». In: *Artificial Intelligence in Medicine* 4.1 (1992), pp. 35–52.

Herring, J. *Criminal law: text, cases, and materials*. Oxford University Press, USA, 2014.

Herron, D., Marohn, M. et al. «A consensus document on robotic surgery». In: *Surgical endoscopy* 22.2 (2008), pp. 313–325.

Hinton, G. et al. «Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups». In: *Signal Processing Magazine, IEEE* 29.6 (2012), pp. 82–97.

Hollander, C. D. e Wu, A. S. «The current state of normative agent-based systems». In: *Journal of Artificial Societies and Social Simulation* 14.2 (2011), p. 6.

Hollnagel, E. *Cognitive reliability and error analysis method (CREAM)*. Elsevier, 1998.

— «From function allocation to function congruence». In: *Coping with computers in the cockpit(A 00-40958 11-54), Aldershot, United Kingdom and Brookfield, VT, Ashgate Publishing, 1999*, (1999), pp. 29–53.

— «The Human in Control: Modelling What Goes Right Versus Modelling What Goes Wrong». In: *Human Modelling in Assisted Transportation*. Springer, 2011, pp. 3–7.

Hsu, F.-H. *Behind Deep Blue: Building the computer that defeated the world chess champion*. Princeton University Press, 2002.

Ingebretsen, M. «Where's My Personal Robot?» In: *Intelligent Systems, IEEE* 24.6 (2009), pp. 90–93.

Italia. «Legge 23 dicembre 1978, n. 833: Istituzione del Servizio Sanitario Nazionale». In: *Gazzetta Ufficiale* 360 ().

Jakopec, M. et al. «The hands-on orthopaedic robot" Acrobot": Early clinical trials of total knee replacement surgery». In: *Robotics and Automation, IEEE Transactions on* 19.5 (2003), pp. 902–911.

Jennings, N. R. «Commitments and conventions: The foundation of coordination in multi-agent systems». In: *The knowledge engineering review* 8.03 (1993), pp. 223–250.

Kalan, S. et al. «History of robotic surgery». In: *Journal of Robotic Surgery* 4.3 (2010), pp. 141–147.

Karagozler, M. E. et al. «Miniature endoscopic capsule robot using biomimetic micro-patterned adhesives». In: *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*. IEEE. 2006, pp. 105–111.

Karnow, C. E. «Liability for distributed artificial intelligences». In: *Berkeley Technology Law Journal* (1996), pp. 147–204.

Kastner, J., Nimmervoll, R. e Wagner, I. «What are the benefits of the Otto Bock C-leg? A comparative gait analysis of C-leg, 3R45 and 3R80». In: *MEDIZINISCH ORTHOPADISCHE TECHNIK* 119 (1999), pp. 131–137.

Kawamoto, H. et al. «Power assist method for HAL-3 using EMG-based feedback controller». In: *Systems, Man and Cybernetics, 2003. IEEE International Conference on*. Vol. 2. IEEE. 2003, pp. 1648–1653.

Kazerooni, H. «Springer Handbook of Robotics». In: a cura di Siciliano, B. e Khatib, O. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. Cap. Exoskeletons for Human Performance Augmentation, pp. 773–793.

Kessel, C. J. e Wickens, C. D. «The transfer of failure-detection skills between monitoring and controlling dynamic systems». In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 24.1 (1982), pp. 49–60.

Kinny, D., Georgeff, M. e Rao, A. «A methodology and modelling technique for systems of BDI agents». In: *Agents breaking away*. Springer, 1996, pp. 56–71.

Kohn, L. T., Corrigan, J. M., Donaldson, M. S. et al. *To err is human: building a safer health system. A report of the Committee on Quality of Health Care in America, Institute of Medicine*. 2000.

Kronreif, G. et al. «Playrob-robot-assisted playing for children with severe physical disabilities». In: *Rehabilitation Robotics, 2005. ICORR 2005. 9th International Conference on*. IEEE. 2005, pp. 193–196.

Kronreif, G. et al. «Robot assistance in playful environment-user trials and results». In: *Robotics and Automation, 2007 IEEE International Conference on*. IEEE. 2007, pp. 2898–2903.

Lamma, E. et al. «An expert system for microbiological data validation and surveillance». In: *Medical Data Analysis*. Springer, 2001, pp. 153–160.

Lamma, E. et al. «Artificial intelligence techniques for monitoring dangerous infections». In: *Information Technology in Biomedicine, IEEE Transactions on* 10.1 (2006), pp. 143–155.

Lang, J. et al. «Robotic systems in orthopaedic surgery». In: *Journal of Bone & Joint Surgery, British Volume* 93.10 (2011), pp. 1296–1299.

Leape, L. L. «Reporting of adverse events». In: *N Engl J Med*. Citeseer. 2002.

LEGAL-IST. «Report on Legal Issues of Software Agents». In: *Doc.* D14.Rev. 2 (2006), pp. 1–140.

Lehman-Wilzig, S. N. «Frankenstein unbound: towards a legal definition of artificial intelligence». In: *Futures* 13.6 (1981), pp. 442–457.

Lessig, L. *Code and other laws of cyberspace*. Vol. 3. Basic books New York, 1999.

Lgs, D. «settembre 2000, n. 332». In: *Attuazione della direttiva* 98 (8), p. 79.

Liang, B. A. e Storti, K. «Creating problems as part of the" solution": the JCAHO Sentinel Event Policy, legal issues, and patient safety.» In: *Journal of health law* 33.2 (1999), pp. 263–285.

Lindholm, R. e Norstedt, J.-P. *The Volvo Report*. Swedish Employers' Confederation, 1975.

Lo, H. S. e Xie, S. Q. «Exoskeleton robots for upper-limb rehabilitation: state of the art and future prospects». In: *Medical engineering and physics* 34.3 (2012), pp. 261–268.

Loureiro, R. et al. «Upper Limb Robot Mediated Stroke Therapy—GENTLE/s Approach». In: *Autonomous Robots* 15.1 (2003), pp. 35–51.

Lum, P. S. et al. «MIME robotic device for upper-limb neurorehabilitation in subacute stroke subjects: A follow-up study». In: *Journal of rehabilitation research and development* 43.5 (2006), p. 631.

Maeso, S. et al. «Efficacy of the Da Vinci surgical system in abdominal surgery compared with that of laparoscopy: a systematic review and meta-analysis». In: *Annals of surgery* 252.2 (2010), pp. 254–262.

Mantovani, F. «Diritto penale». In: *Parte generale, Padova: CEDAM* (2001).

Mantovani, M. *Il principio di affidamento nella teoria del reato colposo*. A. Giuffrè, 1997.

Marchal-Crespo, L. e Reinkensmeyer, D. J. «Review of control strategies for robotic movement training after neurologic injury». In: *Journal of neuroengineering and rehabilitation* 6.1 (2009), p. 20.

Marescaux, J. «Nom de code:«Opération Lindbergh»». In: *Annales de chirurgie*. Vol. 127. 1. Elsevier Masson. 2002, pp. 2–4.

Marescaux, J. et al. «Transatlantic robot-assisted telesurgery». In: *Nature* 413.6854 (2001), pp. 379–380.

Marinucci, G. e Dolcini, E. *Manuale di diritto penale: parte generale*. Giuffrè editore, 2012.

Martinelli, T. et al. «Robot-based tele-echography clinical evaluation of the TER system in abdominal aortic exploration». In: *Journal of Ultrasound in Medicine* 26.11 (2007), pp. 1611–1616.

Marttos, A. et al. «Usability of telepresence in a level 1 trauma center». In: *Telemedicine and e-Health* 19.4 (2013), pp. 248–251.

McFarland, D. *Guilty robots, happy dogs: the question of alien minds*. OUP Oxford, 2008.

McFarland, T. e McCormack, T. «Mind the Gap: Can Developers of Autonomous Weapons Systems be Liable for War Crimes». In: *Int'l L. Stud. Ser. US Naval War Col.* 90 (2014), p. i.

Miller, E. J. «Technology, territory, and time.» In: *Human Relations* 12 (1959), pp. 245–272.

Minsky, M. «Steps toward artificial intelligence». In: *Computers and thought* 406 (1963), p. 450.

Mirbagheri, M. M. «Comparison between the therapeutic effects of robotic-assisted locomotor training and an anti-spastic medication on spasticity». In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Ago. 2015, pp. 4675–4678.

Moray, N., Inagaki, T. e Itoh, M. «Adaptive automation, trust, and self-confidence in fault management of time-critical tasks.» In: *Journal of Experimental Psychology: Applied* 6.1 (2000), p. 44.

Mukai, T. et al. «Development of the tactile sensor system of a human-interactive robot RI-MAN». In: *IEEE Transactions on Robotics* 24.2 (2008), pp. 505–512.

Mumford, E. «The story of socio-technical design: Reflections on its successes, failures and potential». In: *Information Systems Journal* 16.4 (2006), pp. 317–342.

Nagakubo, A., Kuniyoshi, Y. e Cheng, G. «Etl-humanoid-a high-performance full body humanoid system for versatile actions». In: *Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on*. Vol. 2. IEEE. 2001, pp. 1087–1092.

— «The ETL-Humanoid system—a high-performance full-body humanoid system for versatile real-world interaction». In: *Advanced robotics* 17.2 (2003), pp. 149–164.

Naville, P. *Vers l'automatisme social?: problèmes du travail et de l'automation*. Gallimard, 1963.

Nejat, G., Sun, Y. e Nies, M. «Assistive robots in health care settings». In: *Home health care management and practice* 21.3 (2009), pp. 177–187.

Neumann, M. «Norm internalisation in human and artificial intelligence». In: *Journal of Artificial Societies and Social Simulation* 13.1 (2010), p. 12.

Newman, W. S. «Automatic obstacle avoidance at high speeds via reflex control». In: *Robotics and Automation, 1989. Proceedings., 1989 IEEE International Conference on*. IEEE. 1989, pp. 1104–1109.

Niechwiadowicz, K. e Khan, Z. «Robot based logistics system for hospitals-survey». In: *IDT Workshop on Interesting Results in Computer Science and Engineering*. Citeseer. 2008.

Onishi, M. et al. «Generation of human care behaviors by human-interactive robot RI-MAN». In: *Robotics and Automation, 2007 IEEE International Conference on*. IEEE. 2007, pp. 3128–3129.

Orendurff, M. S. et al. «Gait efficiency using the C-Leg». In: *Journal of rehabilitation research and development* 43.2 (2006), p. 239.

Ormerod, D. *Smith and Hogan: Criminal Law*. 2008.

Padhy, N. *Artificial intelligence and intelligent systems*. Vol. 337. Oxford University Press Oxford, 2005.

Pagallo, U. *The Laws of Robots*. Vol. 200. Springer, 2013.

Pagliaro, A. «Il diritto penale tra norma e società. Scritti 1956-2008». In: *Giuffrè* II (2009).

Palep, J. H. et al. «Robotic assisted minimally invasive surgery». In: *Journal of Minimal Access Surgery* 5.1 (2009), p. 1.

Paliero, C. «L'autunno del patriarca. Rinnovamento o trasmutazione del diritto penale dei codici». In: *Riv. it. dir. proc. pen*. Vol. 4. 1994, p. 1220.

Palma, V. et al. *Paro therapy: potenzialità di un robot zoomorfo come mediatore sociale nel trattamento non farmacologico di bambini con sindrome autistica*.

Paper, I. W. «Watson – A System Designed for Answers The future of workload optimized systems design». In: *IBM Systems and Technology* (2011), pp. 1–6.

Parasuraman, R. E. e Mouloua, M. E. *Automation and human performance: Theory and applications.* Lawrence Erlbaum Associates, Inc, 1996.

Parasuraman, R. e Riley, V. «Humans and automation: Use, misuse, disuse, abuse». In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39.2 (1997), pp. 230–253.

Parasuraman, R., Sheridan, T. B. e Wickens, C. D. «A model for types and levels of human interaction with automation». In: *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 30.3 (2000), pp. 286–297.

Parasuraman, R. et al. «Adaptive function allocation reduces performance cost of static automation». In: *7th International Symposium on Aviation Psychology*. DTIC Document. 1993, pp. 37–42.

Partridge, D. e Wilks, Y. *The foundations of artificial intelligence: A sourcebook*. Cambridge University Press, 1990.

Perrow, C. *Normal accidents: Living with high risk technologies*. Princeton University Press, 2011.

— *Normal Accidents: Living with High Risk Technologies (Updated)*. Princeton University Press, 1999.

Piergallini, C. *Danno da prodotto e responsabilità penale: profili dommatici e politico-criminali*. Giuffrè, 2004.

Plaisant, C. et al. «A storytelling robot for pediatric rehabilitation». In: *Proceedings of the fourth international ACM conference on Assistive technologies*. ACM. 2000, pp. 50–55.

Pulitanò, D. *Diritto penale*. Giappichelli, 2009.

Putman, H. «I robot: macchine o vita creata artificialmente?» In: *Mente, Linguaggio e Realtà* (1987), pp. 416–438.

Reason, J. «Combating omission errors through task analysis and good reminders». In: *Quality and Safety in Health Care* 11.1 (2002), pp. 40–44.

— *Human error*. Cambridge university press, 1990.

Reason, J. T. *Managing the risks of organizational accidents*. Vol. 6. Ashgate Aldershot, 1997.

Rebora, G. «Organizzazione aziendale». In: *Teorie e strumenti per l'analisi e la progettazione, Carocci Editore* (1998).

Reynolds, C. W. «Flocks, herds and schools: A distributed behavioral model». In: *ACM SIGGRAPH computer graphics*. Vol. 21. 4. ACM. 1987, pp. 25–34.

Ropohl, G. «Philosophy of socio-technical systems». In: *Techné: Research in Philosophy and Technology* 4.3 (1999), pp. 186–194.

Rosenthal, J., Riley, T. e Booth, M. *State reporting of medical errors and adverse events: results of a 50-state survey*. National Academy for State Health Policy, 2000.

Runciman, W. «Lessons from the Australian Patient Safety Foundation: setting up a national patient safety surveillance system—is this the right model?» In: *Quality and Safety in Health Care* 11.3 (2002), pp. 246–251.

Russel, S. e Norvig, P. «Artificial Intelligence: A Modern Approach, 2003». In: *EUA: Prentice Hall* ().

Salcudean, S. E. et al. «Medical Image Computing and Computer-Assisted Intervention – MICCAI'99: Second International Conference, Cambridge, UK, September 19-22, 1999. Proceedings». In: a cura di Taylor, C. e Colchester, A. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999. Cap. Robot-Assisted Diagnostic Ultrasound – Design and Feasibility Experiments, pp. 1062–1071.

Salute, M. della. «Risk management in Sanità». In: *Il problema degli errori. Allegato* 4 (2004).

— *Sicurezza dei pazienti e gestione del rischio clinico: Manuale per la formazione degli operatori sanitari*. 2007.

Sankai, Y. «HAL: Hybrid assistive limb based on cybernics». In: *Robotics Research*. Springer, 2010, pp. 25–34.

Sarter, N. B. e Woods, D. D. «How in the world did we ever get into that mode? Mode error and awareness in supervisory control». In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1 (1995), pp. 5–19.

Sartor, G. «Gli agenti software: nuovi soggetti del ciberdiritto». In: *Contratto e impresa* 2 (2002), pp. 57–91.

— *Intelligenza artificiale e diritto: un'introduzione*. A. Giuffrè, 1996.

Save, L., Feuerberg, B. e Avia, E. «Designing human-automation interaction: a new level of automation taxonomy». In: *2012), Proc. Human Factors of Systems and Technology* (2012).

Schaeffer, C. e May, T. «Care-o-bot-a system for assisting elderly or disabled persons in home environments». In: *Assistive technology on the threshold of the new millenium* (1999).

Schank, R. C. «What is AI anyway?» In: *The foundation of artificial intelligence—a sourcebook*. Cambridge University Press. 1990, pp. 3–13.

Schebesta, H. et al. «Design According to Liabilities: ACAS X and the Treatment of ADS-B Position Data». In: ().

Searle, J. R. «Minds, brains, and programs». In: *Behavioral and brain sciences* 3.03 (1980), pp. 417–424.

Sermanet, P. et al. «Overfeat: Integrated recognition, localization and detection using convolutional networks». In: *Proc. International Conference on Learning Representations* (2014).

Serpelloni, G. e Simeoni, E. «Principi sull'organizzione dell'azienda socio-sanitaria pubblica». In: *Quality management* (1995).

Sheridan, T. B. «Human supervisory control of robot systems». In: *Robotics and Automation. Proceedings. 1986 IEEE International Conference on*. Vol. 3. IEEE. 1986, pp. 808–812.

Sheridan, T. B. e Verplank, W. L. *Human and computer control of undersea teleoperators*. Rapp. tecn. DTIC Document, 1978.

Sheridan, T. B. «Task analysis, task allocation and supervisory control». In: *Handbook of human-computer interaction* (1997), pp. 87–105.

Shoham, Y. e Tennenholtz, M. «Emergent conventions in multi-agent systems: Initial experimental results and observations». In: *KR-92* (1992), pp. 225–231.

— «On the synthesis of useful social laws for artificial agent societies (preliminary report)». In: *AAAI*. 1992, pp. 276–281.

Shortell, S. M. e Kaluzny, A. D. *Health care management: a text in organization theory and behavior*. Albany, New York: Delmar Thomson Learning, 1988., 2013.

Shute, S. «Knowledge and Belief in the Criminal Law». In: *Criminal law theory – doctrines of the general part* (2002).

Sieg, A., Friedrich, K. e Sieg, U. «Is PillCam COLON Capsule Endoscopy Ready for Colorectal Cancer Screening? A Prospective Feasibility Study in a Community Gastroenterology Practice». In: *The American journal of gastroenterology* 104.4 (2009), pp. 848–854.

Sitti, M. «Micro-and nano-scale robotics». In: *American Control Conference, 2004. Proceedings of the 2004*. Vol. 1. IEEE. 2004, pp. 1–8.

Stella, F. «Giustizia e modernità». In: *La protezione dell'innocente e la tutela delle vittime* (2003).

— *Leggi scientifiche e spiegazione causale nel diritto penale*. Vol. 22. Giuffrè, 1990.

Stoianovici, D. «URobotics—urology robotics at Johns Hopkins». In: *Computer Aided Surgery* 6.6 (2001), pp. 360–369.

Stoianovici, D. et al. «A modular surgical robotic system for image guided percutaneous procedures». In: *Medical Image Computing and Computer-Assisted Interventation—MICCAI'98*. Springer, 1998, pp. 404–410.

Storari, S. et al. «Validation of biochemical laboratory results using the DNSev expert system». In: *Expert systems with applications* 25.4 (2003), pp. 503–515.

Stortoni, L. «Angoscia tecnologica ed esorcismo penale». In: *Riv. it. dir. proc. pen*. Vol. 71. 2004.

Sugano, N. «Computer-assisted orthopedic surgery». In: *Journal of Orthopaedic Science* 8.3 (2003), pp. 442–448.

Sung, G. T. e Gill, I. S. «Robotic laparoscopic surgery: a comparison of the da Vinci and Zeus systems». In: *Urology* 58.6 (2001), pp. 893–898.

Susskind, R. E. «Expert systems in law: A jurisprudential approach to artificial intelligence and legal reasoning». In: *The modern law review* 49.2 (1986), pp. 168–194.

Suzuki, K. et al. «Intention-based walking support for paraplegia patients with Robot Suit HAL». In: *Advanced Robotics* 21.12 (2007), pp. 1441–1469.

Szecsi, T. et al. «Hospital robot module development in the iward project». In: *6th CIRP International Conference on Intelligent Computation in Manufacturing Engineering. Naples, Italy*. 2008.

Tang, V. C. e Stacey, N. C. «Robotic surgery». In: *Annals of The Royal College of Surgeons of England* 89.4 (2007), p. 447.

Taylor, R. et al. «A steady-hand robotic system for microsurgical augmentation». In: *The International Journal of Robotics Research* 18.12 (1999), pp. 1201–1210.

Teich, J. M. et al. «The informatics response in disaster, terrorism, and war». In: *Journal of the American Medical Informatics Association: JAMIA* 9.2 (2002), p. 97.

Teubner, G. «Rights of Non-humans? Electronic Agents and Animals as New Actors in Politics and Law». In: *Journal of Law and Society* 33.4 (2006), pp. 497–521.

Thompson, C. «What is IBM's Watson». In: *New York Times Magazine (June 2011). http://www. nytimes. com/2010/06/20/magazine/20Computer-t. html* (2010).

Thorsrud, E. «Policy-making as a learning process in working life». In: *Working life* (1981), pp. 313–327.

Tondu, B. e Lopez, P. «The McKibben muscle and its use in actuating robot-arms showing similarities with human arm behaviour». In: *Industrial Robot: An International Journal* 24.6 (1997), pp. 432–439.

Trist, E. e Bamforth, K. «Some social and psychological consequences of the Longwall method». In: *Human relations* 4.3 (1951), pp. 3–38.

Üneri, A. et al. «New steady-hand eye robot with micro-force sensing for vitreoretinal surgery». In: *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*. IEEE. 2010, pp. 814–819.

Upbin, B. «IBM's Watson Gets Its First Piece Of Business In Healthcare». In: *Forbes Tech* (2013).

Veruggio, G. «The EURON Roboethics Roadmap.» In: *Humanoids*. Citeseer. 2006, pp. 612–617.

Vieyres, P. et al. «A tele-operated robotic system for mobile tele-echography: The OTELO project». In: *M-Health*. Springer, 2006, pp. 461–473.

Vincent, C. *Patient safety*. John Wiley & Sons, 2011.

Vineis, P. *Modelli di rischio: epidemiologia e causalità*. Einaudi, 1990.

Von Bertalanffy, L. «The history and status of general systems theory». In: *Academy of Management Journal* 15.4 (1972), pp. 407–426.

Von Bertalanffy, L. et al. «The theory of open systems in physics and biology». In: *Science* 111.2872 (1950), pp. 23–29.

Wang, H. e Kasagami, F. «A patient transfer apparatus between bed and stretcher». In: *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on* 38.1 (2008), pp. 60–67.

Weitzenboeck, E. *Electronics agents: some other legal issues. ECLIP 2nd Summer School, Palme De Mallorca*. 2001.

Weyns, D., Steegmans, E. e Holvoet, T. «Towards active perception in situated multi-agent systems». In: *Applied Artificial Intelligence* 18.9-10 (2004), pp. 867–883.

Weyns, D. et al. «Environments for multi-agent systems». In: *First International Workshop (E4MAS 2004), New York, NY, July 19, 2004, Revised Selected Papers*. Vol. 3374. 2004.

Wickens, C. D. e Kessel, C. «The effects of participatory mode and task workload on the detection of dynamic system failures». In: *Systems, Man and Cybernetics, IEEE Transactions on* 9.1 (1979), pp. 24–34.

Winograd, T. «Thinking machines: Can there be? Are we». In: *The boundaries of humanity: Humans, animals, machines* (1991), pp. 198–223.

Wittgenstein, L. et al. *Philosophische Untersuchungen: Philosophical Investigations*. Wiley-Blackwell, 2009.

Woods, D. D. *Behind human error*. Ashgate Publishing, Ltd., 2010.

Wooldridge, M. J. *Reasoning about rational agents*. MIT press, 2000.

Wright, J., Hill, P. e Favaretti, C. *La governance clinica*. McGraw-Hill, 2005.

Yonemitsu, S., Fujimoto, T. e Tamura, T. «Research for practical use of rehabilitation support equipment for severe dementia». In: *Gerontechnology* 2.1 (2002), p. 91.

Zoss, A. B., Kazerooni, H. e Chu, A. «Biomechanical design of the Berkeley lower extremity exoskeleton (BLEEX)». In: *Mechatronics, IEEE/ASME Transactions on* 11.2 (2006), pp. 128–138.