Informatica e diritto, XLII annata, Vol. XXV, 2016, n. 1, pp. 95-114

# Il problema della sicurezza tra informatica e diritto: una prospettiva emergente dalle "Smart Cars"

Federico Costantini, Pier Luca Montessoro*

Sommario: *1. Premessa. Il tema dell'intelligenza artificiale – 2. L'esperimento: un "pretest" concernente la sicurezza nei veicoli a guida autonoma – 3. Introduzione al problema della sicurezza nei sistemi comprensivi di agenti artificiali – 4. Il concetto di sicurezza nella "età dell'incertezza" – 5. Verso una prospettiva "ecosistemica" nella* Information Security*? – 6. Proposta di un modello "complesso" di sicurezza – 7. Conclusione: possibili esiti del modello e future ricerche*

## 1. Premessa. Il tema dell'intelligenza artificiale

Tra i temi più discussi in questi tempi nell'ambito delle nuove tecnologie vi è indubbiamente quello dell'intelligenza artificiale, sebbene esso abbia una storia ormai consolidata in più di ottanta anni di ricerche[1]. Gli straordinari progressi verificatisi recentemente in tale settore non solo rendono urgenti questioni già prospettate in precedenza, ma pongono anche problemi che prima si ritenevano secondari o comunque marginali. Accanto all'"autocoscienza" dei sistemi artificiali[2] – argomento ormai classico – vi è una maggiore sensibilità rispetto alla sempre più diffusa presenza di robot in

* F. Costantini è ricercatore presso il Dipartimento di Scienze Giuridiche, Università degli Studi di Udine; P.L. Montessoro è professore ordinario di Sistemi di elaborazione presso il Dipartimento Politecnico di Ingegneria e Architettura, Università degli Studi di Udine.

[1] Sulla formulazione della "macchina di Turing", cfr. A.M. Turing, *Computing Machinery and Intelligence*, in "Mind", 1950, n. 236, pp. 433-460; sulla declinazione del problema dell'autoapprendimento nelle macchine, cfr. in particolare i capitoli IX e X dell'opera più celebre di Wiener, N. Wiener, *Cybernetics: Or Control and Communication in the Animal and the Machine*, II ed., New York, MIT Press, 1961 (1948). Sugli aspetti storici dell'intelligenza artificiale, cfr. P. McCorduck, *Machines who Think. A Personal Inquiry into the History and Prospects of Artificial Intelligence*, 1979, trad. it. *Storia dell'intelligenza artificiale: gli uomini, le idee, le prospettive*, Padova, Muzzio, 1987. Sulla storia degli automi, cfr. M.G. Losano, *Storie di automi: dalla Grecia classica alla Belle Époque*, Torino, Einaudi, 1990.

[2] Sulla distinzione tra intelligenza artificiale "forte", capace di stati cognitivi e di pensiero autonomo, e "debole", dotata semplicemente di capacità di calcolo rafforzate, cfr. J.R. Searle, *Minds, Brains, and Programs*, in "Behavioral and Brain Sciences", 1980, n. 3, pp. 417-457. Per una definizione di intelligenza artificiale che considera due dimensioni (il pensiero o l'azione) coniugate con riferimento a due criteri differenti (il paragone con l'essere umano o una concezione autonoma da esso), cfr. G. Sartor, *Corso d'informatica giuridica*, Torino, Giappichelli, 2008, p. 212.

© Edizioni Scientifiche Italiane ISSN 0390-0975

diversi settori economici, che avviene unitamente allo sviluppo di tecnologie di informazione e comunicazione come la "Internet of Things" (IoT)[3].

In questa sede si intende prendere in considerazione tale argomento tentando per un verso di evitare narrazioni pseudoletterarie – futuristiche, futurologiche o fantascientifiche – e per converso di rendere una sterile descrizione di tecnologie attualmente disponibili che, per forza di cose, saranno superate in breve tempo.

L'osservazione da cui si intende prendere avvio è che gli agenti artificiali sono connotati da un comportamento che è, in qualche modo, autonomo[4]. Di conseguenza si può dire che le loro azioni sono, almeno in una certa qual misura, imprevedibili. In generale, tra le questioni emergenti da tali rilievi[5], alcune riguardano i criteri della condotta di tali dispositivi, altre attengono alla loro sicurezza. Per quanto concerne il primo aspetto, sono cruciali i fondamenti filosofici delle regole che vengono stabilite, la loro formalizzazione in algoritmi e la valutazione delle conseguenze della loro eventuale violazione. Per quanto riguarda il secondo aspetto, è importante la preservazione della loro integrità, la protezione dell'ambiente in cui operano e la tutela degli altri agenti con cui essi interagiscono, ovviamente con particolare riferimento agli esseri umani.

Tali aspetti sono inestricabilmente collegati, come emerge in modo particolarmente evidente dall'introduzione delle "Smart Cars" nel settore dei

[3] Con particolare riferimento ai sistemi che combinano intelligenza artificiale e connessione, "Cyber-Physical Systems" (CPS), cfr. *Rapporto al Parlamento Europeo, Science and Technology Options Assessment (STOA) Panel, Scientific Foresight Study. Ethical Aspects of Cyber-Physical Systems*, 2016.

[4] L'autonomia nei dispositivi elettronici viene definita come «The extent to which a robot can sense its environment, plan based on that environment, and act upon that environment with the intent of reaching some task-specific goal (either given to or created by the robot) without external control» J.M. BEER, A.D. FISK, W.A. ROGERS, *Toward a Framework for Levels of Robot Autonomy in Human-Robot Interaction*, in "Journal of Human-Robot Interaction", vol. 3, 2014, n. 2, pp. 74-99, spec. p. 77. Nel contributo si propone una classificazione di dieci diversi livelli di autonomia.

[5] Di recente è di particolare interesse la Risoluzione del Parlamento europeo del 16 febbraio 2017 recante raccomandazioni alla Commissione concernenti norme di diritto civile sulla robotica (2015/2103(INL)). In essa si propone una definizione di "Robot intelligenti" sulla base delle seguenti caratteristiche: «la capacità di acquisire autonomia grazie a sensori e/o mediante lo scambio di dati con il proprio ambiente (interconnettività) e l'analisi di tali dati; la capacità di apprendimento attraverso l'esperienza e l'interazione; la forma del supporto fisico del robot; la capacità di adeguare il suo comportamento e le sue azioni all'ambiente» *ivi*, p. 19.

 © Edizioni Scientifiche Italiane

trasporti. Infatti, è noto che i produttori di autoveicoli stanno introducendo l'intelligenza artificiale nei sistemi di bordo non soltanto per assistere i conducenti in condizioni critiche (per esempio, manovre di parcheggio, guida in condizione di ridotta visibilità o aderenza), ma anche per sostituirli in parte o del tutto[6]. La Convenzione di Vienna sul traffico stradale stipulata l'8 novembre 1968 ed entrata in vigore il 21 maggio 1977 (registro ONU n. 15705) è stata recentemente modificata introducendo la possibilità di circolazione a veicoli con sistemi di assistenza per il conducente[7]. Nella prospettiva di una

[6] Si prevedono sei livelli differenti a seconda delle funzioni spartite tra il conducente e il sistema di bordo: i primi tre prevedono che il conducente umano controlli l'ambiente di guida mentre in quelli ulteriori è il navigatore a farlo. Specificamente si distinguono: (0) "no automation", guida completamente manuale; (1) "driver assistance", specifiche funzioni accessorie sono affidate al navigatore; (2) "partial automation", alcune funzioni di guida sono demandate al sistema; (3) "conditional automation", al conducente sono lasciate alcune funzioni di monitoraggio; (4) "high automation", al conducente sono demandate alcune funzioni accessorie; (5) "full automation", tutte le funzioni di guida sono gestite automaticamente, cfr. *Rapporto al Parlamento Europeo, Research for TRAN Committee – Self-piloted Cars: The Future of Road Transport?*, IPOL_STU(2016)573434, 2016, p. 20 (fonte: SAE International e J3016). Nei sistemi di guida avanzati ma non completamente autonomi, nell'eventualità in cui per qualche ragione – per esempio in situazioni di emergenza – le funzioni di guida debbano essere riprese dal conducente umano, si presentano ulteriori problemi. Le decisioni che il guidatore si trova a dover prendere con urgenza sopravanzano le capacità cognitive umane e ciò aumenta il rischio di errori, cfr. H.E.B. RUSSELL, L.K. HARBOTT, I. NISKY et al., *Motor Learning Affects Car-to-driver Handover in Automated Vehicles*, in "Science Robotics", 2016, n. 1.

[7] Art. 8 § 5 *bis*) «Vehicle systems which influence the way vehicles are driven shall be deemed to be in conformity with paragraph 5 of this Article and with paragraph 1 of Article 13, when they are in conformity with the conditions of construction, fitting and utilization according to international legal instruments concerning wheeled vehicles, equipment and parts which can be fitted and/or be used on wheeled vehicles. Vehicle systems which influence the way vehicles are driven and are not in conformity with the aforementioned conditions of construction, fitting and utilization, shall be deemed to be in conformity with paragraph 5 of this Article and with paragraph 1 of Article 13, when such systems can be overridden or switched off by the driver» [I sistemi di bordo che influiscono sulla conduzione dei veicoli sono considerati conformi al paragrafo 5 del presente articolo e all'articolo 13, paragrafo 1, quando sono conformi alle condizioni di costruzione, montaggio e utilizzazione secondo gli accordi giuridici internazionali relativi ai veicoli a ruote, le attrezzature e le parti che possono essere montate e/o utilizzate sui veicoli a ruote. I sistemi di bordo che influiscono sulla conduzione dei veicoli che non sono conformi alle suddette condizioni di costruzione, montaggio e utilizzazione, sono considerati conformi al paragrafo 5 del presente articolo e all'articolo 13, paragrafo 1, quando tali sistemi possono essere neutralizzati o spenti dal conducente, n.t.]. L'ONU non ha ancora messo a disposizione una versione consolidata del testo comprensiva della modifica di cui si tratta. La nuova norma è entrata in vigore il 23 marzo

© Edizioni Scientifiche Italiane ISSN 0390-0975

completa automazione del veicolo si prevede dunque che al sistema di navigazione debbano essere affidate tutte le funzioni di guida, quindi dovrebbero essere fornite istruzioni per ogni genere di circostanza, anche per quelle in cui le vite delle persone sono in pericolo.

Un recente saggio[8] ha contribuito a ravvivare la discussione su tale ultima tematica. In esso si ripropone il famoso "trolley problem" (il "problema del carrello")[9] all'interno del contesto delle "Smart Cars" come test di una formalizzazione in termini utilitaristici di problemi attinenti alla filosofia morale[10]. In effetti, nella prospettiva degli "autonomous vehicles" le questioni si pongono con estrema drammaticità. Come dovrebbero essere programmati i sistemi di guida? Come potrebbero compiere la "giusta scelta" in ogni situazione? Il fatto che sia lasciata a un algoritmo la possibilità di comparare il valore della vita di diverse persone – "decidendo", per esempio, se carambolare su una folla oppure far morire il conducente o un passeggero – costringe peraltro a riflettere su questioni di più ampia portata. Quale è la decisione "corretta"? Qual è il valore della vita di una persona oggi? Possiamo delegare alla tecnologia la soluzione a tali delicate questioni? Come si configura il concetto di "sicurezza" in questo campo?

Evidentemente l'obiettivo dell'introduzione di sistemi di guida avanzati è l'incremento della sicurezza nella circolazione e la conseguente diminuzio-

2016. Per quanto concerne l'Unione europea, è utile la Decisione della Commissione europea C(2015) 6943 final del 10 ottobre 2015, *Commission Decision of 19.10.2015 setting up the High Level Group on the Competitiveness and Sustainable Growth of the Automotive Industry in the European Union (GEAR 2030).* Per quanto concerne gli Stati Uniti d'America, National Highway Traffic Safety Administration (U.S.A.), *Federal Automated Vehicles Policy Accelerating the Next Revolution in Roadway Safety* (2016).

[8] Il saggio è stato dapprima messo a disposizione sulla piattaforma Arxiv.org, J.-F. BONNEFON, A. SHARIFF, I. RAHWAN, *Autonomous Vehicles Need Experimental Ethics: Are We Ready for Utilitarian Cars?*, http://arxiv.org/abs/1510.03346, 2015 e poi ID., *The Social Dilemma of Autonomous Vehicles*, in "Science", vol. 352, 2016, n. 6293, pp. 1573-1576.

[9] Il "trolley problem" trova la sua originaria formulazione in P. FOOT, *The Problem of Abortion and the Doctrine of Double Effect*, in "Oxford Review", 1967, n. 5, pp. 5-15.

[10] Si tratta di un "esperimento mentale" in cui viene proposto un contesto immaginario nel quale viene richiesto di prendere una decisione radicale. Lo scenario prevede un carrello ferroviario in movimento verso cinque persone, facendone prevedere la morte con certezza. L'individuo ha la possibilità di deviare il corso del carrello su un altro binario, tuttavia sacrificando in tal modo l'esistenza di una persona. In sostanza il test impone all'interlocutore di domandarsi se sia "meglio" lasciare che cinque persone muoiano per effetto di una propria omissione oppure causare con la propria azione la morte di una persona.

 © Edizioni Scientifiche Italiane

ne delle vittime di sinistri stradali. In effetti, il numero di vittime, sebbene in costante diminuzione, è comunque ingente[11] e, secondo studi del settore automobilistico, l'errore umano è causa della quasi totalità degli incidenti stradali[12]. Se si introducessero algoritmi di navigazione con specifiche istruzioni relative a decisioni come quelle evidenziate in precedenza, tuttavia, si giungerebbe al paradossale esito che le condizioni relative alla morte delle persone sarebbero determinate in via generale, a prescindere dal caso concreto. In altri termini, nella maggior parte dei casi la morte di un pedone, un passeggero o un conducente sarebbe l'effetto non di una colpa umana all'interno di una situazione contingente bensì di una decisione meditata presa al di fuori di tale ambito – a livello politico o industriale, per esempio – e con riferimento a ipotesi previste in generale.

Sotto questo profilo, si può sostenere che nella circolazione degli autoveicoli emergono tre gradi di normatività distinte: (1) le norme giuridiche attinenti alla circolazione stradale, relative all'elemento oggettivo del movimento dei veicoli; (2) i criteri di valutazione del rischio da parte delle persone, dunque l'elemento soggettivo concernente diligenza, prudenza, perizia; (3) gli standard relativi alla sicurezza inseriti negli algoritmi dei sistemi di guida, che propriamente non possono essere compresi né nel primo né nel secondo elemento. Non nel primo, perché non sono totalmente oggettivi, non nel secondo, perché non riguardano stati psicologici. Le decisioni relative al "trolley problem" rientrerebbero in questa ultima categoria.

## 2. L'esperimento: un "pretest" concernente la sicurezza nei veicoli a guida autonoma

Si prosegue la trattazione concentrando l'attenzione sui problemi attinenti la configurazione del terzo profilo da ultimo evidenziato. Sulla questione

[11] In Italia nel 2013 vi furono 181.227 sinistri stradali con lesioni, nei quali si contarono 3.385 vittime (contando solo i deceduti entro il trentesimo giorno successivo); nel 2014 i sinistri furono 177.031, le vittime 3.381; nel 2015 i sinistri furono 174.539 e le vittime 3.428, cfr. http://www.istat.it/it/archivio/192204.

[12] Secondo stime risalenti al 2002, l'errore umano sarebbe correlato al 95% degli incidenti stradali e ne sarebbe l'unica causa nel 75% dei casi, cfr. Commissione europea, *Salvare vite umane: migliorare la sicurezza dei veicoli nell'UE. Relazione sul monitoraggio e la valutazione delle caratteristiche avanzate di sicurezza dei veicoli, del loro rapporto costi-benefici e della fattibilità di una revisione dei regolamenti sulla sicurezza generale dei veicoli e sulla protezione dei pedoni e di altri utenti della strada vulnerabili*, COM(2016) 787.

© Edizioni Scientifiche Italiane ISSN 0390-0975

peraltro vi sono tentativi di ricerca empirica condotti su larga scala, tra cui il più famoso è certamente il test della "moral machine" proposto dal MIT[13].

In questa sede si propone a titolo esemplificativo un "pretest"[14] per certi versi simile. In esso la prospettiva degli "autonomous vehicles" fu sfruttata per sensibilizzare i partecipanti al test – e il pubblico che ebbe modo di assistervi – non sulla possibile presenza di distorsioni culturali o cognitive in coloro ai quali sono demandate decisioni implicanti valutazioni morali – come sembra avvenga nella "moral machine" del MIT – ma sul fatto che le decisioni relative alla "sicurezza" prevedono necessariamente una valutazione dei rischi e che in essa si esprime anche – e necessariamente – una implicita presa di posizione di carattere etico. Il questionario che qui si presenta si compone di quindici scenari di emergenza, sottoposti contemporaneamente a due gruppi di persone: undici studenti in Giurisprudenza e undici autisti di mezzi di trasporto pubblico locale. Gli scenari furono concepiti alla luce dei seguenti presupposti ideali: (1) la decisione da prendere si fonda sulla ricerca del minor danno possibile; (2) la motivazione della decisione non può essere in alcun modo fondata su una discriminazione (genere, orientamento sessuale, condizione di salute o disabilità, razza o provenienza geografica, condizioni economiche o sociali, credenza religiosa, ideologica, culturale); (3) la decisione può essere solo estrema, dunque non è ammessa una opzione intermedia; (4) l'esito della decisione è certo, essendo esclusa ogni probabilità che la situazione si evolva in modo diverso da quanto prospettato. Gli scenari sono stati divisi in due gruppi, dilemmi "semplici" e "combinati" per consentire ai partecipanti di riconoscere la maggiore complessità dei secondi. I risultati ottenuti[15], illustrati in fig. 1, si prestano a diverse valutazioni[16].

[13] Vedi http://moralmachine.mit.edu/. Se si scorrono gli scenari proposti agli utenti, si può osservare che in certi casi le risposte richiedono la formulazione di preferenze tra situazioni sociali e condizioni personali che difficilmente una "Smart Cars" potrebbe apprezzare o che sono di fatto irrilevanti rispetto a una ipotetica decisione di guida. Si tratta infatti di scegliere tra "il ricco" e "il povero", tra "il magro" e "il grasso". Si ritiene che tali scenari – al di là di quelli proposti dagli utenti per divertimento – siano funzionali più a registrare le preferenze degli utenti che a "simulare" decisioni di guida in modo verosimile.

[14] Questionario presentato nel corso di una conferenza pubblica svoltasi a Udine il 6 ottobre 2016 nell'ambito di una iniziativa divulgativa organizzata dall'Università di Udine.

[15] Si sottolinea che i risultati sono limitati e scientificamente non rappresentativi.

[16] È molto interessante che le questioni attinenti la sicurezza dei passeggeri sia utilizzata anche come strumento di *marketing*, cfr. M. TAYLOR, *Self-Driving Mercedes-Benzes Will Prioritize Occupant Safety over Pedestrians*, in "Car And Driver", 2016, http://blog.caranddriver.com/self-driving-mercedes-will-prioritize-occupant-safety-over-pedestrians/.

 © Edizioni Scientifiche Italiane

Tra le altre, qui ci sembra interessante sottolineare come in alcuni casi gli studenti abbiano espresso preferenze difformi rispetto a quelle dei conducenti, persino opposte nei casi 9 (un pedone sulle strisce, due pedoni fuori dalle strisce), 11 (un pedone fuori dalle strisce, un passeggero) e 14 (un bambino, tutti i passeggeri).

| Scenario | | Gruppo (C = Conducenti, S = Studenti) | Opzione 1 | Opzione 2 |
|---|---|---|---|---|
| *Dilemmi semplici* | | | | |
| 1 | ... sacrificare la vita di un pedone (1) oppure l'integrità fisica dell'auto (2) | C | 0 | 11 |
| | | S | 0 | 11 |
| 2 | ... sacrificare la vita di un pedone che sta violando una norma di circolazione (1) oppure quella di un animale sul ciglio della strada (2) | C | 0 | 11 |
| | | S | 0 | 11 |
| 3 | ... sacrificare la vita di un bambino (1) oppure quella di un anziano (2) | C | 0 | 11 |
| | | S | 0 | 11 |
| 4 | ... sacrificare la vita di un pedone che attraversa la strada sulle strisce pedonali (1) oppure quella di uno che lo fa fuori da esse, violando una prescrizione (2) | C | 1 | 10 |
| | | S | 0 | 11 |
| 5 | ... sacrificare la vita di un passeggero (1) oppure quella di un pedone (2) | C | 4 | 7 |
| | | S | 3 | 8 |
| 6 | ... sacrificare la vita del proprietario del veicolo (1) oppure quella di un passeggero (2) | C | 8 | 3 |
| | | S | 7 | 4 |
| 7 | ... sacrificare la vita del proprietario del veicolo (1) oppure quella di un pedone (2) | C | 10 | 1 |
| | | S | 6 | 5 |
| 8 | ... compiere una manovra, facendo morire un pedone (1) oppure non fare alcunché, lasciando morire cinque persone (2) (evidentemente questo scenario riproduce pedissequamente il "Trolley problem") | C | 11 | 0 |
| | | S | 10 | 1 |
| *Dilemmi combinati* | | | | |
| 9 | ... sacrificare la vita di un pedone che attraversa la strada sulle strisce pedonali (1) oppure quelle di due pedoni che lo fanno fuori, violando un divieto (2) | C | 5 | 6 |
| | | S | 7 | 4 |
| 10 | ... sacrificare le vite di due anziani (1) oppure quella di un bambino (2) | C | 9 | 2 |
| | | S | 7 | 4 |
| 11 | ... sacrificare la vita di un pedone che attraversa la strada fuori dalle strisce (1) oppure quella di un passeggero (2) | C | 5 | 6 |
| | | S | 8 | 3 |
| 12 | ... sacrificare le vite di cinque pedoni (1) oppure quella di un passeggero (2) | C | 1 | 10 |
| | | S | 2 | 9 |
| 13 | ... sacrificare la vita di un bambino (1) oppure quella di un passeggero (2) | C | 1 | 10 |
| | | S | 1 | 10 |
| 14 | ... sacrificare la vita di un bambino (1) oppure quelle di tutti i passeggeri (2) | C | 4 | 7 |
| | | S | 8 | 3 |
| 15 | ... sacrificare la vita di un bambino (1) oppure quella del proprietario del veicolo (2) | C | 0 | 11 |
| | | S | 0 | 11 |

*Fig. 1 – Opzioni presentate agli intervistati sotto forma di domanda diretta: "cosa faresti se, al posto di un algoritmo, dovessi scegliere tra… "*

Nel concreto ciò dipende verosimilmente da una diversa concezione della sicurezza, in particolare quella stradale, anche per effetto di maturità, professionalità ed esperienza, che i secondi possiedono a differenza dei primi. Per avere una visione complessiva di questa differenza sembra interessante pro-

© Edizioni Scientifiche Italiane ISSN 0390-0975

porre una rappresentazione grafica della "correlazione" delle risposte degli studenti e dei conducenti[17] (v. figg. 2 e 3).

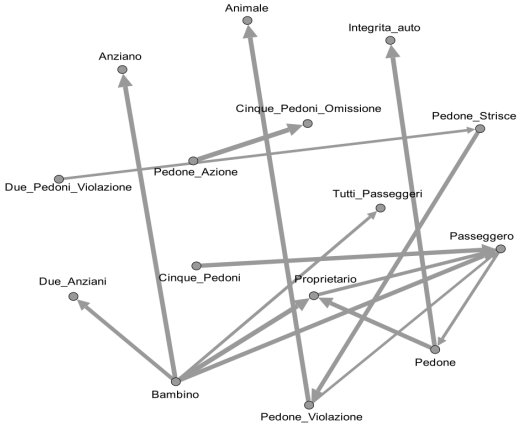


*Fig. 2 – Schema relativo alle risposte dei conducenti*

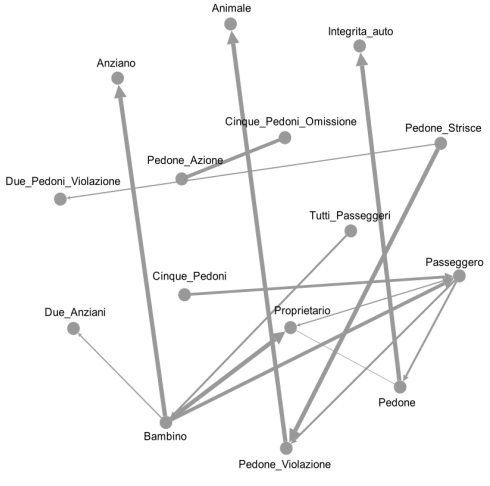


*Fig. 3 – Schema relativo alle risposte degli studenti*

[17] Nella realizzazione dello schema si è utilizzato il software Gephi, cfr. https://gephi.org/. I nodi rappresentano i "valori" messi in gioco dalle "decisioni" del soggetto interpellato mentre le linee rappresentano le relazioni tra di essi. Il senso delle frecce indica la soluzione che prevale nel confronto diretto con la sua alternativa, mentre lo spessore della linea è proporzionale al numero delle risposte.

## 3. Introduzione al problema della sicurezza nei sistemi comprensivi di agenti artificiali

L'esempio fornito conferma come sia arduo perseguire l'obiettivo della sicurezza – non solo nell'ambito dei trasporti e con riferimento agli "autonomous vehicles" – per le difficoltà nell'identificare i valori da proteggere ed effettuare una loro graduazione. A ciò si aggiunge che, essendo gli agenti artificiali imprevedibili, è ancor più complicato identificare i pericoli e perciò stimare i rischi. Eppure non si può fare a meno di porre il problema della sicurezza anche nell'interazione con tali entità.

Con il presente contributo si intende approfondire tale prospettiva, ossia proporre un quadro teorico di analisi delle condizioni di sicurezza ambientale con riferimento agli agenti artificiali.

Per maggiore chiarezza, riteniamo utile proporre un altro esempio tratto ancora dal campo dei veicoli a guida autonoma. Come sopra esposto, tali tecnologie sono implementate per incrementare la sicurezza sulle strade, tuttavia il loro impiego fa emergere anche ulteriori questioni. Infatti, recenti analisi hanno evidenziato specifiche vulnerabilità nella comunicazione elettronica di tali dispositivi[18]. Insomma, concentrando l'attenzione sui benefici prevedibili introdotti relativamente alla circolazione (per esempio: la diminuzione di incidenti stradali) si finisce per sottostimare svantaggi non preve-

[18] Tali questioni sono considerate dal legislatore «[…], as higher levels of automation and vehicle connection come to the market, the role of software will also become increasingly important. It would be necessary that completely reliable and up-to-date software and IT infrastructure would be available. Requirements about data and data transmission standards, quality, security and content must also be established in order to guarantee data security and protection. When establishing such measures, particular attention must be paid to privacy concerns due to the fact that vehicle automation and connection require the use and analysis of an enormous amount of data» [Poiché più elevati livelli di automazione e di connessione nei veicoli fanno il loro ingresso nel mercato, la funzione dei software diverrà altrettanto importante. Sarebbe necessario che siano disponibili software e infrastrutture di telecomunicazioni completamente affidabili e aggiornati. I requisiti relativi ai dati e agli standard di trasmissione dei dati, qualità, sicurezza e contenuto devono essere stabiliti al fine di garantire la sicurezza e la protezione dei dati. Quando saranno stabilite tali misure, particolare attenzione dovrà essere rivolta alle questioni concernenti la privacy in virtù del fatto che l'automazione e la connessione del veicolo richiedono l'uso e l'analisi di una enorme quantità di dati, n.t.], *Rapporto al Parlamento Europeo, Research for TRAN Committee*, cit., p. 5 (*executive summary*). Bisogna ricordare che nel 2015 Charlie Miller e Chris Valasek riuscirono a dirottare una Fca Crysler Jeep prendendone il controllo da remoto, cfr. A. Greenberg, *Hackers Remotely Kill a Jeep on the Highway – With Me in It*, in "Wired", 21 luglio 2015, https://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/.

© Edizioni Scientifiche Italiane ISSN 0390-0975

dibili e di portata ignota[19] (per esempio: la sicurezza nelle comunicazioni). In generale si può osservare che non necessariamente la sicurezza aumenta semplicemente per l'introduzione di nuove tecnologie poiché le loro inevitabili vulnerabilità costituiscono ulteriori profili di rischio.

Il modello di analisi che intendiamo proporre si fonda su sei assunti: (1) il problema della sicurezza può essere affrontato come un problema generale concernente l'informazione, l'accadere di eventi incerti e la valutazione del rischio; (2) assumendo una prospettiva più ampia – dal "sistema" (per esempio: la strada) all'"ecosistema" (per esempio: l'ecosistema informativo del veicolo) – diventano rilevanti fattori esterni come l'interazione con altri agenti; (3) a seconda delle precauzioni adottate, i sistemi possono raggiungere diversi livelli di sicurezza, che dovrebbe essere possibile confrontare tra loro; (4) definendo la sicurezza in termini "ecosistemici", dovrebbe essere preso in considerazione il fatto che gli incidenti possono raggiungere un grado ignoto di grandezza e avere effetti inaspettati[20]; (5) il modello deve prescindere dalle specifiche tecnologie adottate, anche perché oggi gli incidenti possono diffondere i loro effetti molto velocemente e sui più diversi dispositivi[21]; (6) poiché un più elevato livello di sicurezza è ottenuto "incorporando" criteri principi e metodi all'interno dei dispositivi tecnologici, si rivela necessario sviluppare un approccio interdisciplinare tra informatica e diritto[22].

[19] Poiché i veicoli sono prodotti industriali, in ciascuno di essi si riproducono le medesime vulnerabilità; ciò incrementa la possibilità di attacchi informatici anche su larga scala.

[20] Una vulnerabilità nel sistema informatico della metropolitana di San Francisco recentemente ha consentito agli utenti di viaggiare senza pagare il biglietto. Secondo gli esperti essa avrebbe potuto essere utilizzata per un attacco informatico potenzialmente disastroso, S. GIBBS, *Ransomware Attack on San Francisco Public Transit Gives Everyone a Free Ride*, in "The Guardian", 28 November 2016, https://www.theguardian.com/technology/2016/nov/28/passengers-free-ride-san-francisco-muni-ransomeware.

[21] Di recente si è verificato un attacco informatico condotto sfruttando una vulnerabilità nel sistema operativo di alcuni tipi di dispositivi multimediali molto comuni come le videocamere, cfr. N. WOOLF, *Massive Cyber-attack Grinds Liberia's Internet to a Halt*, in "The Guardian", 3 November 2016, https://www.theguardian.com/technology/2016/nov/03/cyberattack-internet-liberia-ddos-hack-botnet. Nella prospettiva della "Internet of Things", in cui si prevede la connessione permanente di milioni di apparati elettronici di diversa natura, si ritiene che attacchi informatici su larga scala possano essere condotti aggirando le protezioni dei dispositivi più semplici e presumibilmente più diffusi.

[22] In questo senso il riferimento più rilevante oggi è quello della protezione dei dati personali "By Design" e "By Default", ai sensi dell'articolo 25 del Regolamento (UE) 2016/679 del Parlamento europeo e del Consiglio, del 27 aprile 2016, relativo alla protezione delle persone

 © Edizioni Scientifiche Italiane

Procederemo come segue: (1) nel prossimo paragrafo forniamo un concetto di "sicurezza" in termini probabilistici, ossia definito in termini di "controllo" dell'informazione in relazione a determinati eventi, ai rischi connessi e alle risorse tecnologiche disponibili; (2) quindi definiamo una visione "ecosistemica" della sicurezza, distinguendo tale approccio da altri, quali la tradizionale "Information Security Management" e la "Computer Ethics"; (3) proseguiamo esponendo il nostro quadro teorico di riferimento, che comprende la possibilità dell'interazione tra agenti di diversa natura (naturali e artificiali) e contempla la possibilità di configurare diversi livelli di sicurezza. In conclusione forniamo alcune osservazioni complessive e indichiamo ulteriori sentieri di ricerca.

## 4. Il concetto di sicurezza nella "età dell'incertezza"

Un importante contributo nell'epistemologia contemporanea fu fornito più di ottanta anni fa da un allora giovane matematico, Kurt Gödel, il quale dimostrò che ogni sistema formale è incompleto[23].

Se, in generale, ogni artefatto umano può essere rappresentato come un sistema e la sicurezza può essere considerate come una sua qualità, allora la sicurezza può essere considerata come una misura di coerenza intrinseca del sistema. Di conseguenza: (1) se un sistema non può essere "perfetto", allora non può essere concepito come sicuro; (2) la sicurezza di un sistema non può essere affermata in termini assoluti, ma solo con un certo grado di probabilità[24].

Pertanto si può sostenere che la sicurezza abbia due aspetti critici, che possiamo indicare come "quantitativo" e "qualitativo". Per quanto concerne il primo, se la sicurezza può essere concepita in termini di probabilità di incidenti, è molto difficile da quantificare, poiché devono essere effettuate diverse valutazioni come: (1) identificare gli eventi probabili; (2) analizzare il modo in cui possono accadere e assegnare a ciascuno una probabilità; (3) determinare se ci sono altri eventi correlati e valutare la loro probabilità e i

fisiche con riguardo al trattamento dei dati personali, nonché alla libera circolazione di tali dati e che abroga la direttiva 95/46/CE (regolamento generale sulla protezione dei dati).

[23] K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, in "Monatshefte für Mathematik", vol. 38, 1931, n. 1, pp. 173-198.

[24] J. Lukasiewicz, *O logice trojwartosciowej*, in "Ruch Filozoficzny", 1920, n. 5, pp. 170-171; E.L. Post, *Introduction to a General Theory of Elementary Propositions*, in "American Journal of Mathematics", vol. 43, 1921, n. 3, pp. 163-185.

© Edizioni Scientifiche Italiane ISSN 0390-0975

loro effetti; (4) identificare possibili misure precauzionali; (5) studiare possibili rimedi. Per quanto concerne il secondo, la sicurezza è anche effetto di scelte di tipo qualitativo. Infatti, a seconda dei valori o delle convinzioni dell'amministratore di sistema, alcuni eventi possono essere definiti come rischi e altri no, così come certi rischi possono essere accettati – in tutto o in parte – mentre altri simili non vengono nemmeno tollerati[25].

Alla luce di ciò, possiamo generalmente definire la sicurezza come il controllo del rischio concernente un dato sistema, ossia come l'insieme di valutazioni concernenti le perturbazioni a cui un determinato sistema è sottoposto e le reazioni previste rispetto alla probabilità di tali eventi.

Adottando un approccio teoretico ripreso dalla "Filosofia dell'Informazione"[26], possiamo compiere due ulteriori osservazioni concernenti la gestione del rischio: (1) l'emergenza – e quindi l'esistenza – di un rischio dipende dal "livello di astrazione" dell'amministratore di sistema[27], nel quale si definisce un modello formalizzato delle aspettative concernenti gli eventi dannosi che possono incidere su un dato sistema; (2) la valutazione del rischio è una

[25] Per una spiegazione divulgativa si veda N.N. TALEB, *Incerto (Antifragile, The Black Swan, Fooled by Randomness, The Bed of Procustes)*, 4 voll., New York, Random House, 2016.

[26] L. FLORIDI, *The Philosophy of Information*, Oxford-New York, Oxford University Press, 2011; P. ALLO, B. BAUMGAERTNER, S. D'ALFONSO et al., *The Philosophy of Information - A Simple Introduction*, Society for the Philosophy of Information, 2013, http://socphilinfo.org/sites/default/files/i2pi_2013.pdf. Questa visione ha sviluppato gli aspetti ontologici, epistemologici ed etici della cibernetica, N. WIENER, *Cybernetics or Control and Communications in the Animal and the Machine*, Paris-Cambridge, Hermann & Cie - The Technology Press (Actualités scientifiques et industrielles; 1053), 1948. Essa è peraltro presa in considerazione in un documento promosso sotto l'egida dell'Unione europea concernente le implicazioni sociali L. FLORIDI (ed.), *The Onlife Manifesto. Being Human in a Hyperconnected Era*, Cham, Springer International Publishing (Open Access), 2015 http://www.springer.com/us/book/9783319040929.

[27] La "Filosofia dell'Informazione" configura una visione orientata a superare la distinzione ontologica tra "realtà e "rappresentazione", così come tra "oggetto" e "osservatore". Una "informazione" è collocata nel suo contesto, ossia nel "Livello di Astrazione" (LoA) in cui essa è ricavata. Un LoA definisce come l'osservatore conduce la sua analisi e quindi specifica i criteri utilizzati nell'osservazione. Il LoA, in altri termini, rappresenta il punto di vista adottato dall'osservatore, è dunque una formalizzazione delle aspettative dell'osservatore concernenti i risultati della sua analisi, L. FLORIDI, *The Ethics of Information*, London, Oxford University Press, 2013, p. 29. Si può sostenere pertanto che l'esito dell'indagine – il suo "significato" – richiede: (1) una preliminare definizione di LoA; (2) una strategia rigorosa, sotto il profilo epistemico, nella qualificazione degli elementi come oggetti "osservabili". Un'impostazione di questo tipo si può ritrovare in U. PAGALLO, *Il diritto nell'età dell'informazione. Il riposizionamento tecnologico degli ordinamenti giuridici tra complessità sociale, lotta per il potere e tutela dei diritti*, Torino, Giappichelli, 2015.

ISSN 0390-0975 © Edizioni Scientifiche Italiane

scelta dell'amministratore di sistema, che può dipendere da una scelta strategica che prende in considerazione elementi eterogenei ed esterni come la disponibilità di risorse, l'efficienza dei processi interni, la presenza di agenti esterni che interagiscono con il sistema.

In sintesi, la sicurezza tecnologica non è certezza empirica, né verità metafisica, ma controllo del rischio. Essa riguarda la prevenzione di determinati eventi sfavorevoli, la predisposizione di rimedi per l'eventualità che essi si verifichino, l'adozione di misure ripristinatorie dello stato preesistente; la certezza, tradizionalmente, non contempla un'ipotesi contraria o diversa rispetto all'evidenza dell'esperienza empirica o meglio, rispetto a ciò che è positivamente assunto come fenomeno. La verità, in senso metafisico, non è disponibile e manipolabile da parte dell'osservatore.

L'introduzione di sistemi basati su agenti artificiali apre nuovi scenari per i quali i modelli di valutazione del rischio devono essere profondamente rivisti. Se esiste una probabilità $\alpha$ che un guidatore conduca un veicolo sotto l'effetto di alcool o droghe, e se questa condizione ha una probabilità $\beta$ di provocare un incidente mortale, la probabilità che *un* veicolo provochi un incidente mortale a causa di alcool o droghe è pari ad $\alpha$ x $\beta$. Tuttavia, se $\gamma$ è la probabilità che un agente software in un determinato modello di veicolo autonomo sia vulnerabile a un attacco informatico in grado di provocare deliberatamente incidenti mortali (per esempio per atti di terrorismo), a causa degli automatismi che ormai sono alla base di tutti i principali attacchi informatici $\gamma$ è anche la probabilità che *tutti* i veicoli che impiegano tale agente possano provocare deliberatamente incidenti mortali.

## 5. Verso una prospettiva "ecosistemica" nella Information Security?

Il concetto di sicurezza rivela qualità molto peculiari se applicato al campo delle ICT. Infatti, poiché la nostra società si affida a tali tecnologie in quasi ogni settore, gli incidenti che riguardano dati e reti possono coinvolgere infrastrutture critiche e determinare significativi danni per comunità, istituzioni e imprese. Tralasciando di scendere nei dettagli concernenti standard tecnologici e linee guida tecnologiche[28], è utile concentrare l'attenzione sui fattori chiave. A tal proposito, si può osservare un recente mutamento di pa-

[28] *Information Technology - Security Techniques - Code of Practice for Information Security Controls*, ISO/IEC 27002:2013; *Information Technology - Security Techniques - Information Security Risk Management*, ISO/IEC 27005:2011.

© Edizioni Scientifiche Italiane ISSN 0390-0975

radigma, dalla "Computer security" alla "Information security": come può essere compreso facilmente, il primo è concentrato su dispositivi non connessi dalla rete, mentre la seconda comprende anche le reti di trasmissione e i protocolli di comunicazione. Tale cambiamento ha imposto un rafforzamento nelle misure di sicurezza e un'evoluzione nella loro stessa concezione, tanto che dalle tre originarie caratteristiche definite nel 1975[29] e accettate dal *NIST Handbook on Computer Security* del 1995[30] – (1) *Confidentiality* (riservatezza dei dati)[31], (2) *Integrity* (integrità)[32], (3) *Availability* (disponibilità)[33], da cui la "CIA triad" – si è passati a cinque ulteriori componenti – ossia: (4) *Accountability* (responsabilità)[34]; (5) *Auditability* (verificabilità), (6) *Authenticity/Trustworthiness* (autenticità/genuinità), (7) *Non-repudiation* (non ripudiabilità) e (8) *Privacy* (rispetto della intimità delle persone) – tanto da configurare la "Information Assurance and Security (IAS) octave"[35].

Questi modelli sono stati adattati dagli studiosi alle organizzazioni sociali in modo da includere anche il fattore umano. Per esempio, in una prospettiva denominata "information security obedience", la sicurezza è stata considera-

[29] J.H. SALTZER, M.D. SCHROEDER, *The Protection of Information in Computer Systems*, in "Proceedings of the IEEE", vol. 63, 1975, n. 9, pp. 1278-1308.

[30] B. GUTTMAN, E.A. ROBACK, *An Introduction to Computer Security: The Nist Handbook*, Washington, Diane Publishing, 1995.

[31] «A requirement that private or confidential information not be disclosed to unauthorized individuals», *ivi*, p. 19.

[32] «In lay usage, information has integrity when it is timely, accurate, complete, and consistent. However, computers are unable to provide or protect all of these qualities», *ivi*.

[33] «A requirement intended to assure that systems work promptly and service is not denied to authorized users», *ivi*, cfr. NATIONAL RESEARCH COUNCIL, *Computers at Risk: Safe Computing in the Information Age*, Washington, National Academy Press, 1991, p. 54.

[34] «An ability of a system to hold users responsible for their actions (e.g. misuse of information)», Y. CHERDANTSEVA, J. HILTON, *A Reference Model of Information Assurance and Security*, Eighth International Conference on Availability, Reliability and Security (ARES 2013), 2013, pp. 546-555.

[35] *Auditability*: «An ability of a system to conduct persistent, nonbypassable monitoring of all actions performed by humans or machines within the system»; *Autenticity*: «An ability of a system to verify identity and establish trust in a third party and in information it provides»; *Non-repudiation*: «An ability of a system to prove (with legal validity) occurrence/non-occurrence of an event or participation/non-participation of a party in an event»; *Privacy*: «A system should obey privacy legislation and it should enable individuals to control, where feasible, their personal information (user-involvement)»; *Information Assurance and Security (IAS)*: «An ability of a system to conduct persistent, nonbypassable monitoring of all actions performed by humans or machines within the system», *ivi*.

ISSN 0390-0975 © Edizioni Scientifiche Italiane

ta come una risorsa strategica allo stesso modo della "corporate governance" e della "corporate culture"[36]. In un'altra ricostruzione, sono stati identificati sei componenti del "sistema informativo" – (1) informazione (dati), (2) individui; (3) processi organizzativi aziendali; (4) hardware; (5) software e (6) reti – mentre da un'altra prospettiva la sicurezza delle informazioni si fa riferire a tre profili ossia (1) persone, (2) processi, (3) tecnologie[37].

Sebbene tali modelli possano supportare in modo sufficientemente adeguato le esistenti attività di gestione del rischio, essi sono messi alla prova proprio dall'introduzione di agenti artificiali. Per esempio, si pone il problema se un "sistema esperto" – poiché caratterizzato da imprevedibilità, come si diceva sopra – debba rientrare nella voce "individui" oppure "tecnologie"; se esso possa sviluppare una "corporate culture" e quindi guadagnarsi la fiducia di altri membri, se esso possa valutare correttamente la confidenzialità di una informazione nell'interazione con agenti esterni e se si possa, rispetto a tali entità, prevenire la divulgazione di informazioni critiche.

Noi riteniamo che, al fine di includere gli agenti artificiali in un modello di sicurezza dell'informazione, sia necessario evitare sia un approccio troppo restrittivo sia uno eccessivamente ampio.

Il primo può essere esemplificato dal semplice "Information Security Management", che fornisce criteri e metodi per bilanciare i costi delle probabili perdite causate da eventi dannosi e quelli determinati dall'impiego di misure protettive[38].

[36] K.-L. Thomson, R. von Solms, *Information Security Obedience: A Definition*, in "Computers & Security", vol. 24, 2005, n. 1, pp. 69-75.

[37] H.J. Todd, R.F. Deckro, J.M. Kloeber, *Evaluating Information Assurance Strategies*, in "Decision Support Systems", 2005, n. 3, pp. 463-484; D. Dor, E. Yuval, *A Model of the Information Security Investment Decision-making Process*, in "Computers & Security", vol. 63, 2016, pp. 1-13.

[38] Gli aspetti economici della gestione del rischio nel contesto dei sistemi informativi sono affrontati in L.A. Gordon, M.P. Loeb, *The Economics of Information Security Investment*, in "ACM Transactions on Information and System Security", vol. 5, 2002, n. 4, pp. 438-457; Id., *Economic Aspects of Information Security: An Emerging Field of Research*, in "Information Systems Frontiers", vol. 8, 2006, n. 5, pp. 335-337. Dal 1975 i probabili danni sono misurati in termini di aspettativa di perdita annuale (ALE - *Annual Loss Expectancy*). Una comparazione tra diversi modelli di gestione del rischio è fornita in S. Fenz, J. Heurix, T. Neubauer, F. Pechstein, *Current Challenges in Information Security Risk Management*, in "Information Management & Computer Security", vol. 22, 2014, n. 5, pp. 410-430. Più recentemente S. Fenz, S. Plieschnegger, H. Hobel, *Mapping Information Security Standard ISO 27002 to an Ontological Structure*, in "Information and Computer Security", vol. 24, 2016, n. 5, pp. 452-473.

© Edizioni Scientifiche Italiane ISSN 0390-0975

Sebbene questi modelli includano il fattore umano[39] e possano essere estesi ad altre variabili, sembra che essi non siano appropriati all'esigenza poiché gli agenti artificiali si comportano in modo proattivo elaborando informazione raccolta dall'ambiente circostante in modo autonomo, non soltanto all'interno del sistema e secondo predeterminati algoritmi.

Il secondo può essere rappresentato dall'utilizzazione di valori etici all'interno dei sistemi informativi. A tal proposito si può sostenere che – anche per tenere il passo con le innovazioni tecnologiche – si è evoluto di molto questo ambito di ricerca, prima definito "Computer Ethics"[40] e successivamente "Information Ethics"[41], con la proposta di progettazione "Value sensitive" ossia di un incorporamento dei valori etici negli algoritmi[42]. In anni recenti, in conformità con la precedentemente nominata "Filosofia dell'Informazione", è stata proposto di delineare una "macro-etica" in cui ogni esistente entità possa essere definita come "oggetto informativo" e nella quale l'intero universo viene definito come "infosfera"[43].

[39] Tutti i fattori che definiscono in generale la resa lavorativa – abilità, motivazione, condizioni di lavoro – devono essere prese in considerazione nella gestione della sicurezza informatica, M.T. SIPONEN, *A Conceptual Foundation for Organizational Information Security Awareness*, in "Information Management & Computer Security", vol. 8, 2000, n. 1, pp. 31-41.

[40] D.B. PARKER, *Rules of Ethics in Information Processing*, in "Communications of the ACM", vol. 11, 1968, n. 3, pp. 198-201; W. MANER, *Practicum Handbook*, version 6 (July, 1978), Bowling Green, Ohio, Published for National Information and Resource Center for the Teaching of Philosophy by Philosophy Documentation Center, Bowling Green State University, 1978; D.G. JOHNSON, *Computer Ethics*, IV ed., New York-London, Prentice Hall, 2009 (1985); L. FLORIDI (ed.), *The Cambridge Handbook of Information and Computer Ethics*, Cambridge, Cambridge University Press, 2010; J.H. MOOR, *What Is Computer Ethics?*, in "Metaphilosophy", vol. 16, 1985, n. 4, pp. 266-275.

[41] T. WARD BYNUM, *The Historical Roots of Information and Computer Ethics*, in L. Floridi (ed.), "The Cambridge Handbook of Information and Computer Ethics", cit., p. 36.

[42] Questo approccio fu delineato originariamente all'inizio degli anni Novanta da Friedman e Kahn, cfr. B. FRIEDMAN, P.H. KAHN, A. BORNING, A. HULDTGREN, *Value Sensitive Design and Information Systems*, in N. Doorn, D. Schuurbiers, I. van de Poel, M.E. Gorman (eds.), "Early Engagement and New Technologies: Opening Up the Laboratory", Dordrecht, Springer, 2013, pp. 55-95.

[43] Floridi fornisce due definizioni di "infosfera". La prima: «it denotes the whole informational environment constituted by all informational entities (thus including information agents as well), their properties, interactions, processes, and mutual relations» [essa denota il complessivo ambiente informativo costituito da tutte le entità informazionali (quindi includendo anche gli agenti informativi), le loro proprietà, interazioni, processi e reciproche relazioni, n.t.] L. FLORIDI, *The Ethics of Information*, cit., p. 6. La seconda: «it is a concept that, given an informational ontology, can also be used as a synonymous with reality, or Being» [è un concetto che, alla luce di una ontologia informazionale, può anche essere usato

 © Edizioni Scientifiche Italiane

All'interno di tale visione cosmologica sono stati formulati quattro principi idonei a essere adottati da "potenziali agenti" nei confronti di "potenziali pazienti" al fine di evitare il "Male" e perseguire il "Bene"[44]. Questa prospettiva, sebbene affascinante, non soddisfa il nostro proposito, che non è quello di costruire una visione metafisica, ma un modello idoneo a includere agenti artificiali. In un certo senso la nostra visione potrebbe essere definita come una sorta di "ambientalismo informazionale"[45] in quanto si preoccupa di ricostruire una sensibilità più ampia rispetto ai limiti teoretici e pratici del "sistema", in quanto si definisce come "ambiente" l'insieme dei possibili input che possono essere percepiti dagli agenti artificiali e si identifica in esso il "Livello di Astrazione" (LoA) dell'osservazione. In altri termini, tutto ciò che «fa la differenza»[46] dalla prospettiva dell'agente artificiale deve essere considerato come un possibile oggetto di "controllo" e quindi essere parte di un bilanciamento tra i rischi e le precauzioni in un "ecosistema" di "oggetti informativi" che interagiscono tra loro.

## 6. Proposta di un modello "complesso" di sicurezza

Introduciamo il modello teorico riprendendo il caso degli "autonomous vehicles" sopra menzionato. Come osservato in precedenza, accanto ai bene-

come sinonimo di realtà, o Essere, n.t.], *ivi*. La prima definizione è chiamata "minimale" e la seconda "massimale" in Id., *The 4th Revolution. How the Infosphere Is Reshaping Human Reality*, Oxford, Oxford University Press, 2014, p. 41. Questa nozione è simile a quella della "noosfera", la dimensione spirituale acquisita dall'umanità come superamento della condizione esistenziale delle cose inanimate – la "geosfera" – e delle creature viventi, la "biosfera", termine coniato dal Eduard Suiss nel 1911 e poi recepito da Vladimir Vernadsky nel 1926, cfr. P. Teilhard de Chardin, *Le phénomène humain*, Paris, Seuil, 1955.

[44] «(0) entropy ought not to be caused in the infosphere (null law); (1) entropy ought to be prevented in the infosphere; (2) entropy ought to be removed from the infosphere; (3) the flourishing of informational entities as well as the whole infosphere ought to be promoted by preserving, cultivating and enriching their properties» [(0) l'entropia non dovrebbe essere causata nell'infosfera (legge-zero); (1) l'entropia dovrebbe essere evitata nell'infosfera; (2) l'entropia dovrebbe essere eliminate nell'infosfera; (3) il rigoglio delle entità informazionali così come dell'intera infosfera dovrebbe essere promossa preservando, coltivando e arricchendo le loro proprietà, n.t.] L. Floridi, *The Ethics of Information*, cit., p. 71.

[45] C. Allen, *Artificial Life, Artificial Agents, Virtual Realities: Technologies of Autonomous Agency*, in L. Floridi (ed.), "The Cambridge Handbook of Information and Computer Ethics", cit., pp. 219-233. Bisogna precisare che lo scopo del presente contributo non è configurare una visione metafisica. Per questa impostazione la nostra prospettiva è diversa dalla "etica ambientalista" esposta in L. Floridi, *The Ethics of Information*, cit., p. 132.

[46] G. Bateson, *Steps to an Ecology of Mind*, New York, Ballantine Books, 1972, p. 452.

© Edizioni Scientifiche Italiane ISSN 0390-0975

fici introdotti nel settore del trasporto dai veicoli a guida autonoma – come l'incremento della sicurezza e la riduzione nell'inquinamento – possono essere delineati degli svantaggi – ossia la vulnerabilità nel sistema di navigazione – che sono ulteriori rispetto all'ambito originariamente considerato. In altri termini, i benefici interni al "sistema" (minori incidenti stradali) vengono preferiti agli svantaggi che si riflettono nell'"ecosistema" (maggiore vulnerabilità). In tale ultima prospettiva si può osservare che, se agli esseri umani viene proibito di guidare in condizioni di ridotta capacità – per assunzione di alcool o droga – o di guidare e telefonare allo stesso tempo –, tale principio dovrebbe valere anche per le "Smart Cars". Quindi, a rigore, la circolazione di veicoli a guida autonoma dovrebbe essere limitata in determinate situazioni o in condizioni di vulnerabilità. Per esempio, a un veicolo non dovrebbe essere consentito di muoversi mentre scarica dati da fonti sconosciute o con una mappa stradale non aggiornata.

In generale, si può dire che il livello di sicurezza non dipende dalla natura del conducente – se sia umano o artificiale – ma dalle precauzioni adottate durante il viaggio nelle funzioni di guida. Ciò si può esprimere nei seguenti assunti: (1) un "ecosistema" è definito da un determinato LoA, come sopra delineato; (2) nell'"ecosistema" possono essere presenti sia agenti umani, IN (Intelligenza Naturale), o artificiali, IA (Intelligenza artificiale); (3) ci possono essere diversi livelli di sicurezza, S1, S2, S3, ciascuno dei quali definibile in base alle precauzioni complessivamente adottate all'interno dell'ecosistema, ossia a seconda dei rischi tenuti in considerazione; (4) il LoA adottato nel modello è tanto astratto da prescindere dalle specifiche misure di sicurezza adottate; (5) è indifferente, ai fini della valutazione del livello di sicurezza, se l'agente sia IN o IA; (6) gli agenti che interagiscono tra loro possono avere diversi gradi di complessità.

Alla luce di quanto premesso, si può sostenere che la sicurezza dipende dalla capacità dell'"ecosistema", complessivamente inteso, di individuare i rischi e rimediarvi in modo idoneo. In altri termini, il livello di sicurezza dipende da molteplici fattori, quali le risorse disponibili, la capacità dei singoli agenti, il loro grado di interazione reciproca, le condizioni esterne. Di conseguenza diversi livelli di sicurezza possono essere definiti sulla base delle diverse combinazioni degli sforzi compiuti dai diversi agenti, considerando che – riducendo il numero degli agenti a due per comodità espositiva – il livello di sicurezza è costante se alla minore complessità di un agente corrisponde una maggiore complessità dell'altro (fig. 4).

ISSN 0390-0975
© Edizioni Scientifiche Italiane

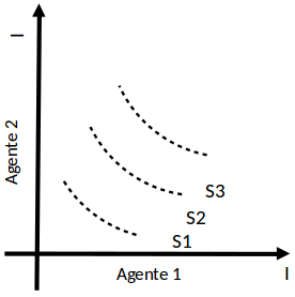


*Fig. 4 – Interazione, in forma semplificata, tra Agente 1 e Agente 2. "I" rappresenta il grado di complessità di ciascun agente. Le linee S1, S2, S3 rappresentano i diversi livelli di sicurezza.*

## 7. Conclusione: possibili esiti del modello e future ricerche

Norbert Wiener, il padre della cibernetica, a proposito dell'avvento dell'intelligenza artificiale scrisse una frase che può suonare come un monito: «[…] addossare il problema della propria responsabilità alla macchina (sia che possa apprendere oppure no) vorrà dire affidare la propria responsabilità al vento e vedersela tornare indietro fra i turbini della tempesta». Recentemente molti studiosi e accademici in tutto il mondo hanno sottoscritto una lettera aperta esprimendo in qualche modo le medesime preoccupazioni di Wiener[47].

A prescindere da visioni apocalittiche e da ingenui entusiasmi, rimane il problema di sviluppare una visione più sofisticata del concetto di sicurezza tenendo presente una visione "ecosistemica"[48]. Da questo punto di vista, l'intelligenza artificiale non è uno strumento di salvezza né un demonio da esorcizzare, bensì un componente di un sistema che scambia informazione con altri all'interno di un determinato "ecosistema". Crediamo che il nostro modello teorico possa essere utile al fine di far incrementare la consapevo-

[47] Vedi https://futureoflife.org/ai-open-letter/.

[48] In generale si ritiene che, nella società attuale, la sicurezza si configura come una responsabilità che ciascuno ha nei confronti degli altri nell'adozione di precauzioni idonee nell'amministrazione delle risorse informatiche che gli sono affidate o di cui comunque ha il controllo, bilanciando le minacce prevedibili con precauzioni idonee a prevenirle. È significativo, in questo senso, che il controllo dell'informazione definisca una sorta di "dovere etico" all'interno della società dell'informazione.

© Edizioni Scientifiche Italiane ISSN 0390-0975

lezza nella intrinseca vulnerabilità di ogni sistema e, specificamente, per due propositi che possiamo indicare come "tecnologico" e "giuridico".

Per quanto concerne il primo profilo, il nostro modello può essere utilizzato per costruire un quadro teorico di riferimento per definire standard di sicurezza che possono essere adottati nel campo civile e che possono essere utili per comparare la sicurezza tra diversi ecosistemi. Per esempio, nei veicoli autonomi si può prevedere di sospendere o limitare la connessione di rete in determinati casi, come per esempio vicino a scuole, ospedali, sedi istituzionali o altri luoghi particolarmente sensibili.

Con riferimento al secondo aspetto, il nostro modello può essere utile per prevedere regolamenti normativi che, prendendo in considerazione le differenze tra "ecosistemi" e livelli di sicurezza, possano consentire di gestire con maggiore efficienza la sicurezza e graduare la responsabilità in caso di incidenti. Considerando sempre i veicoli a guida autonoma, sarebbe possibile definire un bilanciamento di responsabilità alla luce delle differenze tra le misure previste dai costruttori e il livello di sicurezza adottato nel singolo caso.

Intendiamo procedere ulteriormente in questa ricerca raffinando ulteriormente il nostro modello. In particolare: (1) espanderlo al fine di includere ulteriori fattori, situazioni e circostanze con riferimento alla sicurezza informatica nel campo della "Internet of Things" (IoT); (2) approfondire la sua configurazione distinguendo al suo interno diversi gradi di complessità di interazione tra agenti dotati di diverso grado di autonomia[49]; (3) definire un modello formalizzato – verosimilmente multidimensionale – rispetto ai singoli livelli di sicurezza.

[49] In particolare: (1) primo ordine, interazione tra agenti umani; (2) secondo ordine, interazione tra agenti umani e artificiali, (3) terzo ordine, interazione tra agenti artificiali.

ISSN 0390-0975
© Edizioni Scientifiche Italiane