

TRA *IMPRINTING* ETICO E DILEMMI MORALI NEI VEICOLI A GUIDA AUTONOMA

Fabio Fernicola*

SOMMARIO: 1.- Introduzione: l'avvento dei VA e il dilemma etico; 2.- Il dilemma del tunnel e la "moral machine": esplorare le scelte etiche; 3.- Dalla teoria alla pratica: responsabilità e sfide reali; 4.- Prospettive future e futuristiche.

1.- Introduzione: l'avvento dei VA e il dilemma etico.

Può accadere, in una serata conviviale tra amici, di aver bevuto un bicchiere di troppo e di dover rinunciare a guidare per tornare alla propria dimora, optando per i trasporti pubblici, il taxi o magari la sollecitudine di uno tra i commensali, ovviamente sobrio. Questo scenario ipotetico, non tanto lontano dalla realtà quotidiana, potrebbe essere smentito dalla nuova frontiera dell'industria automobilistica e del trasporto, poiché alcuni colossi, tra cui BMW, Tesla, Uber e Google, stanno investendo grandi risorse per introdurre nel mercato veicoli a guida completamente autonoma (VA). In alcune grandi città statunitensi, invece, i robotaxi della società Waymo, che conta una flotta di 700 veicoli, stanno diventando davvero un'alternativa alla guida tradizionale. Se questo scenario dovesse presto realizzarsi su larga scala, il tasso alcolemico nel nostro sangue non costituirà più un problema per percorrere in auto il tragitto di casa, né influenzerà le decisioni dei VA, ma spianerà la strada a un nuovo, e problematico, aspetto: l'intersezione tra l'ambiente esterno e le scelte morali che i VA dovranno prendere in situazioni emergenziali. Il presente contributo, pertanto, si propone di esplorare il dilemma etico insito nella programmazione dei VA, dal momento che la complessità di alcuni scenari affrontati dai VA, che si basano su sofisticati sistemi di intelligenza artificiale (IA), impedisce una programmazione deterministica dell'azione da svolgere. In primo luogo, si analizzerà se sia possibile definire una gerarchia di valori da implementare nei sistemi decisionali dei VA. In secondo luogo, si discuteranno le questioni relative alla responsabilità degli eventuali danni causati dai VA. Secondo l'ultimo Rapporto DEKRA (DEKRA Road Safety Road 2024), l'essere umano, con la sua condotta al volante, è il principale attore in oltre il 90% degli incidenti stradali. È una statistica impietosa, e per questo l'industria automobilistica sta puntando ormai da anni su sistemi di assistenza alla guida in grado di riconoscere tempestivamente situazioni critiche del traffico o modalità di guida non sicure, grazie al continuo miglioramento della tecnologia (sensori, automazione, digitalizzazione) per ridurre significativamente il numero di sinistri, con il relativo costo in termini di vite umane. Parallelamente, la frontiera dei VA potrebbe, in un futuro non troppo lontano, diminuire sensibilmente quella percentuale: gli agenti autonomi non si distraggono per controllare le notifiche dello smartphone, non violano i limiti di velocità e, per tornare all'esempio iniziale, non alzano il gomito prima di mettersi al volante. A ogni buon conto, la complessità della guida in termini di elaborazione dati in tempo reale, associata alla dinamicità delle strade a elevata congestione del traffico (basti pensare alla presenza di incroci, semafori, strisce pedonali, ciclisti e così via), evidenzia come la progettazione dei VA debba fare i conti non solo con le norme del Codice della Strada, ma

* Docente a contratto in Storia dell'ontologia antica presso l'Università degli Studi di Salerno.

anche con aspetti di natura etica¹. Soprattutto se, secondo la classificazione dei sei livelli di automazione veicolare proposta nel 2014 dalla “Society of Automotive Engineers”, che vanno da 0 a 5, l’analisi viene condotta sulla versione più estrema di tali veicoli, in grado di operare senza supervisione dell’uomo, che da conducente si appresta a diventare passeggero. Ciò che mi preme sottolineare è come la progettazione di un veicolo di livello 5 di automazione (cosiddetto “fully automated”), debba necessariamente riflettere un particolare orientamento morale, una scelta valoriale che si mette in pratica quando un VA opta per quella soluzione e non altre parimenti razionali. Si pensi al caso delle collisioni inevitabili, situazioni imponderabili in cui il VA non può evitare lo scontro, trovandosi a decidere in pochi secondi se salvaguardare la salute dei propri passeggeri o quella degli individui esterni a esso. Il dilemma del tunnel, che proverò a descrivere con il seguente scenario, è un classico esempio nella letteratura di settore².

2.- Il dilemma del tunnel e la “moral machine”: esplorare le scelte etiche.

Immaginate di essere comodamente seduti nell’abitacolo della vostra auto autonoma, alle prese con un’attività rilassante mentre il VA entra in un tunnel autostradale a velocità sostenuta ma entro i limiti previsti dal Codice della Strada. Davanti a voi c’è uno scuolabus con numerosi bambini a bordo che procede alla vostra stessa velocità. Nella corsia di sorpasso c’è una vettura con due passeggeri di mezza età che vi sta affiancando per superavi. Per qualche imprevisto lo scuolabus inchioda e si presenta una situazione di estremo pericolo. Gli scenari possibili sono tre: a. il vostro VA frena ma inevitabilmente vi schiantate contro l’autobus, esponendo un numero imprecisato di bambini a lesioni gravi e potenzialmente letali; b. il vostro VA sterza bruscamente contro l’auto in transito a sinistra, la sperona e provoca il decesso della coppia; c. la vostra VA riconosce i due precedenti scenari, perciò decide di schiantarsi contro la parete laterale in cemento armato del tunnel, sacrificando la vostra vita ma salvandone altre. Come si può intuire, le prime due opzioni salvaguardano principalmente la vostra incolumità a scapito delle vite altrui, mentre il terzo scenario è il cosiddetto atto supererogatorio: si compie un atto di estremo sacrificio pur non essendo obbligatorio né espressamente richiesto. Le tre opzioni riflettono, dunque, tre diverse scelte valoriali che rispondono al seguente interrogativo: quando in gioco è la possibilità di salvare la vita di uno o più esseri umani, è più giusto lasciar morire oppure uccidere?

È da una riflessione su questa domanda che nacque nel 1967 il “trolley problem”, quando la filosofa inglese Philippa R. Foot ideò un esperimento mentale di etica applicata, sollevando la questione nei seguenti termini: nei pressi di uno scambio ferroviario, un vagone perde il controllo e si appresta a investire cinque persone legate sui binari; ma se qualcuno azionasse la leva che controlla il bivio, il vagone cambierebbe tragitto, investendo una sola persona, legata sull’altro binario. Cosa è moralmente giusto: azionare lo scambio oppure rimanere impassibili a guardare?³. Certo, si tratta di

¹ Sul punto, si rimanda a G. Tamburrini, *Etica delle macchine. Dilemmi morali per robotica e Intelligenza Artificiale*, Roma 2020; S.O. Hansson, M.A. Belin, B. Ludgren, *Self-Driving vehicles – An ethical overview*, in *Philosophy & Technology* 34.4 (2021) 1383ss. Per un inquadramento del problema dal punto di vista giuridico e filosofico, si rimanda a F. Coppola, *Intelligenza Artificiale, Metaverso e Sistema Penale: prevenzione, repressione, opportunità, rischi*, Milano 2025, 115ss.

² J. Gogoll, J. Mueller, *Autonomous cars: In favour of a mandatory setting*, in *Science and Engineering Ethics* 23.3 (2017) 681ss.

³ Per un inquadramento del problema, con annessa bibliografia, si rimanda a J. Hacker-Wright, Philippa Foot, in E.N. Zalta (cur.), *The Stanford Encyclopedia of Philosophy*, Stanford, CA 2021, <https://plato.stanford.edu/archives/win2021/entries/philippa-foot/>.

una situazione abbastanza inverosimile, ma se al posto del vagone ferroviario sostituissimo il VA che è troppo vicino all'imprevisto per frenare, e se al posto delle persone legate ai binari avessimo da una parte un boy-scout adolescente che attraversa la strada con il semaforo rosso e dall'altra una signora anziana che passeggia sul marciapiede, il dilemma etico su chi investire sarebbe più attuale che mai. È chiaro che è stato il boy-scout a infrangere la legge passando col rosso, ma saremmo tutti disposti a spezzare una vita così giovane, quando con una sterzata si porrebbe fine alla vita di una persona che, al confronto, ha vissuto così a lungo? E se, invece della signora, ci ritrovassimo dieci persone anziane? La valutazione del quadro complessivo, basata sul compromesso costi-benefici, ci impone di prendere in considerazione il valore della vita umana non soltanto in termini assoluti, ma anche in rapporto alle complesse dinamiche di interazione sociale. E pone in risalto una questione cruciale, ossia che la tecnologia sofisticata alla base dei VA non può dirsi neutrale, ma incorpora e contiene valori morali⁴.

Allora per capire come vada impostato l'“imprinting” etico dei sistemi di IA nelle “self-driving cars”, un gruppo di ricercatori dello statunitense “Massachusetts Institute of Technology”, in collaborazione con università di altri paesi, ha realizzato un sondaggio a larga scala (disponibile al sito internet di riferimento <https://moralmachine.net>), che ha ottenuto oltre quaranta milioni di risposte da aree diverse del mondo⁵. Nel sondaggio veniva proposta ai partecipanti un'articolata serie di ambientazioni e di incidenti, e veniva chiesto di indicare quali persone fosse più giusto salvare e quali sacrificare, tenendo conto di diverse variabili, come età, genere, cultura di appartenenza, rispetto o violazione delle norme del codice stradale, numero di soggetti coinvolti nell'impatto e persino il decesso di animali da passeggio. D'altro canto, le combinazioni possono essere infinite e, come emerge anche da studi più recenti⁶, le preferenze degli individui possono anche essere molto incoerenti e poco razionali, frutto di coinvolgimenti emotivi e “bias” cognitivi. Ciò solleva significative problematiche riguardo all'implementazione di valutazioni morali casistiche in sistemi di IA, dove gli sviluppatori definiscono i parametri comportamentali delle macchine sulla base di una notevole quantità di dati allenanti, i cosiddetti “big data”.

Se infatti continuiamo ad analizzare i risultati della ricerca, ci rendiamo conto di quanto sia difficile operare una gerarchia di valori, sebbene dallo studio siano emerse tendenze generali, come la preferenza a risparmiare il maggior numero di vite, dare priorità agli esseri umani rispetto agli animali e salvaguardare maggiormente i giovani, in particolare i bambini. Tuttavia, sono emerse anche significative differenze regionali legate alle culture di appartenenza, che hanno permesso ai ricercatori di isolare tre grandi blocchi regionali che sono caratterizzati da un differente peso morale attribuito a diverse categorie di possibili vittime. Il blocco occidentale (che comprende il Nord America e gran parte dei paesi europei) si distingue per essere il raggruppamento di paesi in cui la preferenza accordata a particolari categorie di persone è minima. Il blocco orientale (che comprende molti paesi dell'Estremo oriente e buona parte dei paesi islamici) si caratterizza invece per una tendenza a risparmiare preferibilmente i pedoni e chi non ha infranto la legge. Per contro, in questi paesi è decisamente al di sotto della media la tendenza ad avere un occhio di riguardo per i giovani e per le donne. Infine, i ricercatori hanno anche notato che nei paesi in cui vi sono drammatiche

⁴ Sulla non-neutralità degli artefatti, si veda I.V.D. Poel, P. Kroes, *Can technology embody values?*, in P. Kroes, P.P. Verbeek (curr.), *The Moral Status of Technical Artefacts*, Dordrecht 2014, 103ss.

⁵ E. Awad, S. Dsouza, R. Kim, et al., *The Moral Machine experiment*, in *Nature* 563 (2018) 59ss.

⁶ E. Award, S. Dsouza, A. Sharif, et al., *Universals and variations in moral decisions made in 42 countries by 70,000 participants*, in *Proceedings in the National Academy of Science* 117.5 (2019) 2332ss.

disuguaglianze di reddito (Perù, Nigeria, Indonesia) un fattore decisivo è lo *status* sociale delle potenziali vittime, sacrificando più facilmente le persone di ceto sociale più basso. I risultati del MIT sono molto intriganti perché mostrano come le scelte degli individui possano essere molto diversificate a seconda dei contesti socio-culturali di provenienza, prediligendo un codice etico deontologico oppure un approccio utilitaristico non facilmente negoziabili, probabilmente per via del pregiudizio di essere nella parte buona della curva statistica (e di avere la presunzione di pensare di guidare meglio degli altri)⁷.

3.- Dalla teoria alla pratica: responsabilità e sfide reali.

Vanno tuttavia fatte due precisazioni. In primo luogo, come si è evidenziato in apertura, l'errore umano rappresenta il fattore principale, per non dire esclusivo, dei sinistri stradali. Se i VA di livello 5 riuscissero, come sembra, a essere sempre prudenti e concentrati nella guida, eliminando in tutto, o in gran parte, i decessi su strada, ciò costituirebbe un traguardo fondamentale per la sicurezza della mobilità stradale. In secondo luogo, immaginare tutta una serie di rompicapi mentali è attività più filosofica che pratica, dal momento che le collisioni inevitabili non sono affrontate nei normali corsi di scuola guida. Si tratta di fenomeni estremamente improbabili in cui non si potranno mai conoscere con esattezza le percentuali di rischio di ogni opzione, anzi la scelta meno rischiosa sarebbe quella più ovvia, ossia frenare il più rapidamente possibile.

In linea di principio, si potrebbe adottare un quadro concettuale a partire dalle considerazioni svolte da Tommaso d'Aquino in merito alla dottrina della proporzione del bene sul male come risultato delle conseguenze delle nostre azioni volontarie⁸. Proprio ciò che il filosofo e teologo medievale diceva circa la relazione tra la bontà o malvagità di un'azione iniziale e un *eventus sequens* dovrebbe indurci a riflettere sul fatto che qualsiasi scelta, nella collisione inevitabile, proprio in quanto è costitutivamente animata da buone intenzioni, non può che risolversi in una situazione di aporia. Pertanto, come è stato osservato, in situazioni imprevedibili e complesse, cambiare direzione all'improvviso significherebbe avere un'alta probabilità di perdere il controllo del veicolo, provocando non solo danni a terzi ma persino reazioni a catena che non avevamo affatto preventivato, con il serio rischio, per usare le parole di Tommaso, di aggiungere *plura mala*⁹.

Ripensando al dilemma etico del tunnel che ho descritto in precedenza, non è affatto detto che il VA di livello 5, equipaggiato con sensori all'avanguardia e innovativi sistemi di frenata, una volta che abbia rispettato la distanza di sicurezza (o magari abbia deciso di aumentarla perché riconosce un potenziale pericolo nella eventuale collisione con uno veicolo adibito al trasporto di minori), si schianti contro lo scuolabus. Mentre è statisticamente più probabile che, sterzando bruscamente a destra o a sinistra per evitare lo scontro frontale, il VA inneschi una serie di danni collaterali persino

⁷ M.D. Garasic, *Droni, robot e visioni: perché abbiamo bisogno dell'algopolitica*, in M. Galletti, S. Zipoli Caiani (curr.), *Filosofia dell'Intelligenza Artificiale*, Bologna 2024, 227ss.: 239.

⁸ Thom. de Aquino, *S.Th.* I-II, q. 20, art. 5: *Si est praecogitatus, manifestum est quod addit ad bonitatem vel malitiam. Cum enim aliquis cogitans quod ex opere suo multa mala possunt sequi, nec propter hoc dimittit, ex hoc apparet voluntas eius esse magis inordinata. Si autem eventus sequens non sit praecogitatus, tunc distinguendum est. Quia si per se sequitur ex tali actu, et ut in pluribus, secundum hoc eventus sequens addit ad bonitatem vel malitiam actus, manifestum est enim meliorem actum esse ex suo genere, ex quo possunt plura bona sequi; et peiorem, ex quo nata sunt plura mala sequi. Si vero per accidens, et ut in paucioribus, tunc eventus sequens non addit ad bonitatem vel ad malitiam actus, non enim datur iudicium de re aliqua secundum illud quod est per accidens, sed solum secundum illud quod est per se.*

⁹ R. Davnall, *Solving the single-vehicle self-driving car trolley problem using risk theory and vehicle dynamics*, in *Science and Engineering Ethics* 26.1 (2020) 431ss.

peggiori. Questa non è un'implicita ammissione di abbandonare il dibattito sulla valutazione morale della condotta dei VA, ma vuole evidenziare come altre siano le sfide all'ordine del giorno, a partire dalla realizzazione di una tecnologia futuristica che riconosca in una frazione di secondo tutte le variabili del contesto critico: dalle condizioni meteorologiche avverse, alla decifrazione della segnaletica orizzontale, verticale e luminosa, per non parlare del numero complessivo dei ciclisti o la presenza di una donna in gravidanza sul ciglio della strada, solo per fare pochi esempi. Intendo dire che, in un contesto storico nel quale le “self-driving cars” di livello 5 non facciano registrare incidenti grazie alla celere ed efficiente diagnostica, o quantomeno facciano in modo che i sinistri stradali siano soltanto un'ipotesi remota e residuale, legata alle collisioni inevitabili, l'asse della discussione si debba spostare sull'individuazione dei profili di responsabilità derivanti dall'impiego dei veicoli a elevata automatizzazione¹⁰.

Nel nostro ordinamento, se l'assenza di una normativa specifica costringe ad applicare per analogia le norme esistenti, attribuendo la responsabilità amministrativa al proprietario del veicolo, un simile approccio potrebbe risultare ben presto obsoleto nel caso dei VA di livello 5 che si affidano a “software” avanzati: il malfunzionamento di un VA con una capacità decisionale completamente autonoma implica responsabilità da parte di chi produce e progetta questi oggetti. In primo luogo, dichiarare esplicitamente dove ricade la responsabilità in caso di incidente ridefinisce il modo in cui i rischi dell'assicurazione auto sono stabiliti, poiché l'errore del conducente è sostanzialmente rimosso. In secondo luogo, oltre alla responsabilità delle case automobilistiche che producono i VA, esistono anche altre responsabilità che derivano dalla progettazione degli artefatti tecnologici, e vedono coinvolti tutti coloro che lavorano allo sviluppo del “software”, per cui è difficile individuare un contributo determinante, da un punto di vista causale, da parte di uno di loro. È il cosiddetto problema delle molte mani, che può sorgere in contesti organizzativi dove è richiesta una maggiore profusione di sforzi in fase progettuale, grazie all'ausilio di diverse figure professionali (ingegneri, informatici, filosofi, ecc.)¹¹.

A ciò va aggiunto un'ulteriore complicazione, a cui accenno solo in maniera cursoria. I “software” dei VA si basano su algoritmi computazionali di “deep-learning” che sono costruiti come delle scatole nere, cioè non sono dissezionabili poiché le informazioni sono disperse in milioni di connessioni sinaptiche artificiali¹². In altre parole, non sarebbe possibile spiegare perché il veicolo avrebbe preso “quella” decisione, rendendo impossibile imputare *ex post* la responsabilità da prodotto difettoso: al “software engineer”, al “project manager”, al “design lead” o al filosofo della mente? O a tutti solidalmente?

E non si possono affatto escludere, nelle delicate fasi di sviluppo, le pressioni commerciali sul “team”, che di fatto potrebbero limitare la piena libertà d'azione degli sviluppatori, componente rilevante per individuare una piena responsabilità morale. I veicoli, infatti, dovranno essere programmati per prendere decisioni eticamente delicate che riflettono un particolare insegnamento morale, e siano al contempo appetibili per il mercato: chi comprerebbe mai VA che mettono di “default” a repentaglio la vita di chi guida? Di converso, sarebbe giusto accettare che il singolo utente imponga le proprie preferenze di preservazione del conducente, come se stesse impostando la temperatura dell'aria

¹⁰ L. La Via, *Innovazione e diritto: i veicoli a guida autonoma tra responsabilità civile, penale e amministrativa*, in *Cammino Diritto* 3 (2025) 1ss.

¹¹ V. Tripodi, *Etica delle tecniche. Una filosofia per progettare il futuro*, Milano 2020.

¹² A. Schubbach, *Judging Machines. Philosophical Aspects of Deep Learning*, in *Synthese* 198.2 (2019) 1807ss.

climatizzata, secondo un approccio egoistico-utilitarista? Sulla base delle indagini sociologiche condotte dagli studi che, come abbiamo visto, hanno evidenziato un relativismo su base geografica, sarebbe lecito avere opinioni diverse sulla moralità e su come metterle in pratica, così da produrre VA che rispondono ai valori di un particolare mercato di riferimento?

4.- Prospettive future e futuristiche.

Forse, una soluzione a questi annosi problemi potrebbe giungere dallo stesso progresso tecnologico, nella misura in cui l'elevato grado di sofisticazione e l'enorme potenza di calcolo dei "software" implementati nel VA saranno realmente capaci di effettuare una disamina olistica di tutte le variabili coinvolte in un evento imponderabile. Nel caso delle collisioni inevitabili si potrebbe, pertanto, invocare lo stato di necessità ex art. 54 c.p., esonerando sviluppatori e produttori dall'imputabilità, dal momento che il VA ha agito in forza di un evento inevitabile e involontario, sulla base di una valutazione prognostica che deriva dalla pura analisi di dati complessi¹³. Ciò lascerebbe al sofisticatissimo sistema di IA, capace di elaborare in tempo reale una notevole quantità di dati, il compito di selezione il miglior "output" in relazione allo scenario contingente, secondo un modello di "decision-making" fluido e adattivo che riproduce l'autonomia di giudizio dell'uomo, ma senza l'indecisione, l'emotività e le distorsioni cognitive che lo affliggono nella tragica circostanza.

Se gli studi condotti dal MIT ci hanno insegnato qualcosa, è che qualsiasi forma di "imprinting" etico per i VA è legata a una peculiare "Weltanschauung", a una visione del mondo ancorata a dinamiche culturali difficili da armonizzare in un catalogo di regole gerarchicamente strutturate. Pur esaminando i meccanismi cognitivi che rendono le false credenze così resistenti alla confutazione, uno studio del 2012 ha mostrato quanto sia difficile, per la mente umana, abbandonare convinzioni già acquisite e interiorizzate, soprattutto se intervengono fattori esterni come il contesto sociale, le convinzioni pregresse e la reiterazione dello stimolo informativo¹⁴. Il valore assoluto della vita di un uomo dovrebbe essere uguale in ogni parte del pianeta, e non può essere subordinato a una valutazione comparativa tra fattori soggettivi e culturalmente determinati, come l'aspettativa di vita, lo status sociale e, dettaglio da non trascurare, una latente superiorità dei valori e dei principi della cultura di appartenenza. Mi rendo conto che il dilemma etico della scelta viene, in questo frangente, aggirato, e anzi si finisce per scegliere di non scegliere. Ma è proprio qui che si potrebbe avanzare una proposta interpretativa che fissi una sorta di "Grundnorm" valida per tutti proprio per il suo grado di astrazione e generalità, lasciando che il VA valuti caso per caso. Certo, qualcuno potrebbe obiettare giustamente che la decisione se salvare la vita di un pedone oppure quella di due ciclisti non può essere demandata al freddo calcolo di una macchina, allineandosi alla tesi del filosofo Günther Anders il quale, già negli anni Cinquanta del secolo scorso, denunciava i rischi di una pervasiva tecnologizzazione nella vita dell'uomo, perché avrebbe privato l'agire umano di un *tèlos*, svilito alla condizione del puro e semplice fare¹⁵. Tuttavia, così ragionando, non dovremmo usufruire degli enormi benefici che l'intelligenza artificiale sta apportando in settori cruciali dello scibile umano, primo fra tutti il campo medico-sanitario, e dovremmo affidarci alle mani e agli occhi dell'uomo per le operazioni chirurgiche più delicate e complesse.

¹³ Più diffusamente, cf. Coppola, *Intelligenza Artificiale* cit. 126ss.

¹⁴ Cf. S. Lewandowsky, U.K.H. Ecker, C.M. Seifert, N. Schwarz, J. Cook, *Misinformation and Its Correction: Continued Influence and Successful Debiasing*, in *Psychological Science in the Public Interest* 13.3 (2012), 106ss.

¹⁵ G. Anders, *Die Antiquitäten der Mensch.* Bd. I: *Über die Seele im Zeitalter der zweiten industriellen Revolution*, München 1956.

E allora forse il miglior approccio resta ancora quello tramandatoci da Tommaso, e cioè che qualsiasi decisione presa dal VA in situazioni impreviste ed emergenziali deve essere mossa da buone intenzioni, laddove l'aggettivo buono qualifica una condotta finalizzata alla prioritaria salvaguardia della vita umana e, secondariamente, alla minimizzazione dei danni alle cose: un'assiologia che detta poche regole, ma chiare e universalmente condivisibili. Provando a essere più circostanziati, mi sia concessa un'incursione nel mondo della letteratura fantascientifica: la rielaborazione della prima legge della robotica avanzata dallo scrittore Isaac Asimov più di ottant'anni fa potrebbe fare meglio al caso nostro. I programmatori potrebbero scrivere, nel codice sorgente dell'algoritmo, che un VA non debba mai danneggiare un essere umano né permettere, tramite l'inazione, che ciò accada. Al suddetto veicolo, poi, l'ardua scelta del *quomodo*. Ma sappiamo bene come queste valutazioni non sono soltanto morali o giuridiche, ma sono soprattutto politiche, perché hanno un impatto significativo sul tipo di società che intendiamo costruire e affidare alle generazioni future.

Abstract.- L'avvento dei veicoli a guida autonoma (VA) solleva questioni etiche e giuridiche significative, soprattutto negli scenari di collisione inevitabile che richiedono decisioni rapide e moralmente complesse. Attraverso il riferimento al “trolley problem” e all'esperimento globale del MIT “Moral Machine” si evidenziano le divergenze culturali e i limiti nell'impostare un “imprinting” morale di portata universale nei sistemi di intelligenza artificiale dei VA. Parallelamente, si affronta la questione della responsabilità in caso di sinistri: nei contesti in cui il controllo umano è assente, occorre ridefinire i criteri per imputare eventuali danni, considerando anche la natura opaca degli algoritmi di “deep learning” e il coinvolgimento di molteplici attori nello sviluppo tecnologico. Viene quindi proposto un approccio che coniughi progresso tecnologico, criteri etici minimali e universalmente condivisibili, e una chiara attribuzione di responsabilità giuridica.

The advent of autonomous vehicles (AVs) gives rise to a number of significant ethical and legal questions, especially in scenarios of unavoidable collisions that require rapid and morally complex decisions. By referring to the “trolley problem” and the MIT's global “Moral Machine” experiment, cultural differences and limitations in establishing a universal moral “imprinting” in the artificial intelligence systems of AVs are highlighted. Concurrently, the issue of liability in the event of accidents is addressed: in contexts where human control is absent, the criteria for attributing liability for damages need to be redefined, taking into account the opaque nature of deep learning algorithms and the involvement of multiple actors in technological development. The proposed approach integrates technological progress, minimal and universally acceptable ethical criteria, and a clear attribution of legal responsibility.