



[Diritto e innovazione](#)" class="voce">

Navigating the Grey Area: brevissime riflessioni su IA, “nuovi” dilemmi morali e responsabilità penale di Alice Giannini

di [Alice Giannini](#)

17 aprile 2023

Giustizia Insieme è felice di ospitare oggi un articolo di Alice Giannini, dottoranda di ricerca in diritto penale presso l'Università di Firenze e presso l'Università di Maastricht, **vincitrice della prima edizione del Premio Giulia Cavallone istituito presso la Fondazione Calamandrei.**

Giulia Cavallone era un giovane magistrato che, fino alla sua morte a soli 36 anni, ha prestato servizio presso il Tribunale di Roma, dove ha trattato tra l'altro il processo relativo ai depistaggi nel caso Cucchi, e che prima di entrare in magistratura aveva dedicato ampia parte delle sue energie intellettuali alla ricerca accademica in materia penalistica, con un percorso di ampio respiro internazionale. Aveva infatti effettuato periodi di tirocinio presso Istituzioni europee, vinto borse di studio e svolto periodi di ricerca presso importanti istituti stranieri, fino al dottorato di ricerca in cotutela internazionale tra l'Università La Sapienza di Roma e l'Università Paris II Patnhéon Assas di Parigi. La comparazione, lo sguardo rivolto a sé e all'altro con identico rispetto ed attenzione, è lo strumento che Giulia ha scelto, convinta che fosse l'unico utile, nel diritto come nella vita.

Il Premio voluto dalla famiglia Cavallone in memoria di Giulia è un ideale passaggio di testimone ad altre generazioni di giovani ricercatori che, con quello stesso spirito di apertura, intendano

inserire nel proprio percorso accademico un periodo presso università o istituti di ricerca esteri. E il testimone è stato raccolto da Alice Giannini, la cui vocazione internazionale è testimoniata dagli studi già compiuti presso la University of California Hastings College of the Law, San Francisco, e coronata con la cotutela internazionale attivata tra l'Università di Firenze e quella di Utrecht per il suo progetto di ricerca intitolato “*Criminal behaviour and accountability of Artificial Intelligence Systems*”, nell'ambito del Curriculum internazionale European and Transnational Legal Studies.

Accogliendo Alice tra i nostri autori accogliamo idealmente quel respiro, che sopravvive a Giulia grazie al filo invisibile che il Premio contribuisce a tessere, legando il suo percorso interrotto a quello di Alice, di tutte le altre e gli altri giovani ricercatori che costruiscono la propria cultura nell'apertura e nel confronto.

Per chi non abbia conosciuto Giulia, Giustizia Insieme l'ha ricordata qui: <https://www.giustiziainsieme.it/it/gli-attori-della-giustizia/1350-giulia-cavallone-un-ricordo>

Ancora di Giulia e del Premio Giulia Cavallone avevamo già parlato qui: <https://www.giustiziainsieme.it/it/diritto-e-societa/2541-premio-giulia-cavallone-anno-2022>

Navigating the Grey Area: brevissime riflessioni su IA, “nuovi” dilemmi morali e responsabilità penale

Alice Giannini*

Sommario: 1. Introduzione – 2. Un problema di “allineamento”? La formalizzazione di scelte morali – 3. Una vita vale più di mille? L'esempio di bilanciamento tedesco

1. Introduzione

I dilemmi morali costituiscono scenari ipotetici allettanti: sono in grado di suscitare una discussione accesa in qualsiasi tipo di auditorio. La loro attrattiva risiede anche nel fatto che contribuiscono a illustrare i valori fondamentali di una cultura giuridica in un modo accessibile a un pubblico più ampio.^[1] Spesso, le soluzioni proposte per risolvere tali dilemmi sono connesse alla discussione dell'attribuzione di responsabilità per le conseguenze dannose derivanti dall'una o dall'altra scelta. Il tema non è nuovo al diritto penale, ma ha riacquisito notorietà con il diffondersi di sistemi di intelligenza artificiale (IA) sempre più avanzati.

Probabilmente il dilemma morale più “famoso” è il c.d. “*trolley problem*” (problema del carrello ferroviario). Immaginate di essere il conducente di un carrello che si dirige verso cinque operai che stanno riparando i binari. Per evitare di investire i cinque operai, dovete fermare il carrello, ma vi rendete conto che i freni non funzionano. Vedete un binario laterale sulla destra, sul quale c’è un solo operaio. A questo punto vi si presentano due scelte: far virare il carrello a destra e uccidere un uomo, oppure continuare ad andare dritto e uccidere cinque operai.[\[2\]](#) Con lo sviluppo delle tecnologie di IA, in particolare nel settore dei veicoli a guida (semi) autonoma, e la conseguente attenzione dedicata ad una loro possibile regolamentazione, è nato anche un dibattito su come si dovrebbe comportare il sistema di guida in caso di un incidente stradale ricostruito in termini di dilemma morale, nonché su chi dovrebbe essere destinatario di responsabilità a causa dei danni causati in seguito alla scelta posta in essere da tale sistema.[\[3\]](#)

In via generale, vi è uno sbilanciamento tra l’emergere di nuove tecnologie e la loro regolamentazione. Non sorprende la lentezza nella regolamentazione dell’IA tramite *hard-law*, soprattutto in un campo delicato come quello del diritto penale. Come è stato osservato, “il ramo legislativo sembra muoversi a una velocità trascurabile rispetto ai progressi tecnologici, rafforzando la percezione che la normazione tradizionale non sia adatta a questa sfida”.[\[4\]](#) Tra le possibili cause, gli autori citano “la mancanza di una definizione esauriente e precisa di IA... aggravata dal fatto che la definizione cambia con l’evolversi della tecnologia”.[\[5\]](#) Rileva notare come il dibattito in materia sia polarizzato. Da un lato, vi è chi sostiene che la regolamentazione “soffochi” l’innovazione, dall’altro, vi è chi crede in una “politica anticipatrice” del diritto.[\[6\]](#) Senz’altro, nel campo dell’IA è possibile rilevare un divario enorme fra lo sviluppo di varie forme di linee guida/principi etici e quello di leggi (o altre disposizioni vincolanti ad esse assimilabili). Ciò è dimostrato dall’impennata nell’elaborazione di principi relativi all’ “IA etica” negli ultimi 5 anni.[\[7\]](#)

Il settore della guida (semi)automatizzata è stato di recente teatro di numerosi incidenti che hanno chiamato in causa l’applicazione del diritto penale. Si pensi ad esempio al sinistro avvenuto nel 2018 a Tempe, Arizona, che ha portato all’uccisione di una donna in seguito all’impatto con un veicolo di prova di Uber (classificato come livello 3 di automazione)[\[8\]](#) a bordo del quale si trovava un agente umano, oggi accusato di omicidio stradale colposo.[\[9\]](#) Ciò ha portato all’avanzare di primigenie forme di regolamentazione dell’IA specificamente attinenti ai profili di responsabilità penale.[\[10\]](#) Si pensi ad esempio alla recente riforma francese del *Code de la Route* che ha regolamentato i profili di responsabilità penale connessi all’uso di auto a guida autonoma,[\[11\]](#) prevedendo l’esclusione di responsabilità nel caso in cui il sistema di guida sia

attivo, sia utilizzato in conformità alle istruzioni per l'utilizzo e venga commessa un'infrazione delle norme del codice della strada o i reati di omicidio e lesioni stradali.[\[12\]](#)

2. Un problema di “allineamento”? La formalizzazione di scelte morali.

Se si presuppone che i sistemi di IA dovranno affrontare in un futuro prossimo situazioni paradossali, quali i dilemmi morali, è necessario porsi la seguente domanda: possiamo insegnare norme morali ad un sistema artificiale?[\[13\]](#) In particolare, “[l]a definizione dell'algoritmo ‘morale’ delle auto senza conducente ripropone il dilemma del carrello e le sue varianti”.[\[14\]](#)

Gli studi su come programmare sistemi di IA “etici” sono già iniziati.[\[15\]](#) Nello specifico si è assistito ad un'impennata di ricerche sulla creazione di agenti morali artificiali (c.d. “AMA”), ossia su come implementare principi etici e facoltà decisionali morali nelle macchine affinché il loro comportamento possa essere ritenuto “eticamente accettabile”, sia nei confronti degli utenti umani, che di altre macchine. Gli AMA dovrebbero affrontare, o addirittura sostituire, il giudizio umano in situazioni moralmente difficili o ambigue.[\[16\]](#)

Uno degli esempi più noti è l'esperimento *Moral Machine* del MIT, realizzato come un gioco online in cui gli utenti si trovano di fronte a uno scenario che prevede un dilemma morale.[\[17\]](#) Nello specifico, gli utenti si devono confrontare con un incidente inevitabile e con due possibili scelte (mantenere la rotta o sterzare), che portano alla morte di un determinato numero di esseri umani o di animali (o di entrambi). Analizzando le scelte tra i “due mali” espresse dagli utenti, cioè il risultato da questi ritenuto preferibile, in combinazione con i dati demografici e il loro luogo di residenza, i creatori dell'esperimento sono stati in grado di identificare tre tendenze principali che, a loro avviso, rappresentano scorci di un' “etica universale delle macchine”: la preferenza a risparmiare vite umane, la preferenza a risparmiare il maggior numero di vite possibili e la preferenza a risparmiare vite di persone giovani.[\[18\]](#)

Ricerche come quella della *Moral Machine* (e altre ancora) incorrono in diverse difficoltà. In primo luogo, per insegnare una regola morale a una macchina (seguendo un approccio *top-down*), la regola deve essere formalizzata in un linguaggio che possa essere compreso dal sistema (cioè un codice). Formalizzare le decisioni etiche è un'operazione estremamente impegnativa. Si tratta, difatti, di “delineare una moralità di tipo operativa, che dovrebbe definire algoritmicamente le condotte morali da assumere, specie in scenari di danno inevitabile”.[\[19\]](#) Le regole morali sono per loro natura ambigue, quindi difficili da tradurre e da scomporre secondo una precisa struttura riproducibile tramite un algoritmo.[\[20\]](#) Inoltre, vi è la difficoltà nel

ricostruire la flessibilità della “mente morale umana”, capace non solo di fare scelte morali in situazioni mai affrontate prima (a differenza dei sistemi di IA attuali, che si basano sull’esperienza passata) , ma anche di decidere di *infrangere* regole prestabilite.[\[21\]](#) Ciò può portare al c.d. “*alignment problem*”, termine utilizzato per descrivere situazioni in cui vi è un disallineamento tra valori comuni agli esseri umani e quelli espressi da un sistema di IA.[\[22\]](#) Esempio eclatante di tale problema è stato Tay, chatbot rilasciato da Microsoft nel 2016. Una volta permessogli di interagire liberamente online con gli utenti Twitter, Tay ha diffuso contenuti razzisti e antisemiti, dopo sole 24 ore dalla sua “nascita”.[\[23\]](#) Da ultimo, è importante sottolineare come non vi sia un metodo per misurare l’ “eticità” di un sistema di IA. In altri termini, al momento non esiste un “sistema etico ottimale”.[\[24\]](#) Ciò si riflette anche sul fatto che “dietro” la formalizzazione della regola etica vi sia sempre un essere umano, che ha i propri principi, stereotipi e convinzioni. Da un punto di vista normativo, alcuni abbandonano del tutto la questione, posto che “non si può seriamente pensare che la progettazione di regole che governano un’area così critica del progresso tecnologico debba essere messa in attesa finché i filosofi non ‘risolvono’ il *trolley problem* o l’infinità di esperimenti mentali ad esso assimilabili”.[\[25\]](#)

3. Una vita vale più di mille? L'esempio di bilanciamento tedesco

Nel maggio 2021, il *Bundesrat* ha approvato una legge che modifica il Codice della Strada (*Straßenverkehrsgesetz - StVG*).[\[26\]](#) Anche se non disciplina direttamente aspetti legati alla responsabilità penale, è di estremo interesse per questa riflessione, in quanto tenta di fornire una soluzione giuridica a dilemmi morali simili a quelli esaminati nel *Moral machine Experiment* del MIT. Così facendo, la Germania parrebbe aver aperto la strada alla regolamentazione dei livelli più alti di guida autonoma.[\[27\]](#)

Tra le disposizioni più interessanti di questo provvedimento vi è quella contenuta nell'articolo 1 § 1e (2)2, che elenca i requisiti tecnici che devono essere soddisfatti da un veicolo con funzioni di guida autonoma (livello 4 di automazione):

(2) I veicoli con funzione di guida autonoma devono essere dotati di un sistema tecnico in grado di,

[...]

2. rispettare autonomamente le norme del codice della strada rivolte al conducente del veicolo ed essere dotati di un sistema di prevenzione degli incidenti che

- (a) sia progettato per prevenire e ridurre i danni,
- (b) in caso di danni alternativi inevitabili a diversi beni giuridici, tenga conto dell'importanza dei beni giuridici, a condizione che la protezione della vita umana abbia la massima priorità, e
- (c) in caso di danno alternativo inevitabile alla vita umana, non tenga conto di caratteristiche personali.[\[28\]](#)

La legge non chiarisce quali siano tali caratteristiche personali, né afferma se i veicoli autonomi possano decidere di scontrarsi con individui che violino le regole del traffico nel caso di un incidente inevitabile. Un elenco delle suddette caratteristiche si trova nei principi elaborati nel 2017 dalla *Ethics Commission on Automated and Connected Driving* nominata dal *Bundesministerium für Digitales und Verkehr* ai quali si ispira la legge: si tratta di età, sesso, condizioni fisiche o psichiche.

La Commissione ha pubblicato un Report nel 2017, nel quale ha elaborato venti linee guida etiche dedicate al settore della guida automatizzata.[\[29\]](#) Non è possibile analizzare l'intero contenuto del Report della Commissione Etica in questa breve riflessione. Ci si limiterà dunque a portare all'attenzione brevemente ad alcuni dei principi ivi contenuti. Innanzitutto, preme soffermarsi sulle regole 7 ed 8, che affrontano i casi di “situazioni pericolose inevitabili”, ossia i dilemmi morali. La regola 7 stabilisce che, in tali scenari, la protezione della vita umana deve avere la massima priorità, se bilanciata con altri interessi giuridicamente protetti. Di conseguenza, entro i limiti di ciò che è tecnologicamente fattibile, i sistemi di IA preposti alla guida del veicolo devono essere programmati per accettare danni agli animali o alle proprietà in un conflitto di interessi, se ciò significa che così facendo di possono evitare danni alle persone.

La soluzione sembra semplice, a prima vista. Eppure, si pensi al caso in cui il danno alla proprietà - che, in astratto, non dovrebbe essere preferito alla protezione della vita - equivalga a causare la fuoriuscita di petrolio da un'autocisterna o, peggio ancora, il collasso della rete elettrica di un'intera area metropolitana.[\[30\]](#)

In effetti, la prima parte della regola 8 riconosce che la “normalizzazione” dei dilemmi morali, cioè l'invenzione di soluzioni generalizzate basate sul “male minore” non sia sempre fattibile, in quanto “i veri dilemmi etici, come quelli relativi alla scelta tra più vite umane, dipendono dalla situazione concreta, che comprende un comportamento ‘imprevedibile’ delle parti coinvolte. Non possono quindi essere chiaramente standardizzati, né possono essere programmati in modo da essere eticamente incontestabili”.[\[31\]](#) L'approccio proposto dalla Commissione Etica sarebbe simile a quello adottato con le vaccinazioni, dove “l'obbligo di vaccinazione imposto per legge si

traduce in una generale minimizzazione del rischio senza che sia possibile sapere in anticipo se la persona vaccinata apparterrà al gruppo dei (pochi) danneggiati (sacrificati)” e, nonostante ciò, “è nell’interesse di tutti essere vaccinati e ridurre il rischio complessivo di infezione”.^[32]

In conclusione, è possibile traslare tali riflessioni al campo di applicazione del diritto penale. Ci si potrebbe domandare, in ultima battuta, se così facendo la Germania abbia posto le basi per l’esclusione di una responsabilità penale nel caso di una situazione assimilabile ad un dilemma morale “stradale”. Ad esempio, ci si potrebbe chiedere se sia possibile determinare l’applicazione di una scriminante (applicabile al sistema stesso – qualora si apra alla responsabilità diretta del sistema di IA – e/o al soggetto umano destinatario della norma penale, sia questo il proprietario del veicolo o il produttore) nel caso in cui l’auto effettui il bilanciamento, secondo le regole prestabilite, e ciò causi un danno penalmente rilevante. A contrario, chi dovrebbe essere responsabile qualora tale bilanciamento sia svolto erroneamente? Tali quesiti verranno lasciati di proposito aperti. È importante ridimensionare, tuttavia, il valore dei dilemmi morali quali “guide normative”,^[33] soprattutto nell’ambito della diritto penale, basato su una responsabilità di natura *personale*, di conseguenza intrinsecamente collegata alla valutazione della singola situazione in concreto realizzatasi.

* Alice Giannini è dottoranda di ricerca in diritto penale presso l’Università di Firenze e la Maastricht University. Le riflessioni proposte in questo articolo fanno parte della sua tesi di ricerca, che si concentra sulla responsabilità penale dei sistemi di intelligenza artificiale. È stata la vincitrice della prima edizione del premio Giulia Cavallone, istituito dalla Fondazione Calamandrei e dalla famiglia Cavallone per onorare la memoria della dott.ssa Cavallone, giudice del Tribunale di Roma e dottoressa di ricerca in diritto e procedura penale.

[1] H. Hilgendorf, *The dilemma of autonomous driving: Reflections on the moral and legal treatment of automatic collision avoidance systems*, in E. Hilgendorf & J. Feldle (Eds), *Digitization and the Law*, Robotik und Recht, vol. 15, Nomos, 2018, p. 60.

[2] L’invenzione di questo tipo di *dilemma* è attribuita a Philippa Foot. Il termine “*Trolley Problem*” è stato coniato più tardi da Judith Jarvis Thomson. Cfr. P. Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, in *Oxford Review*, Vol. 51978; J. J. Thomson, *The Trolley Problem*, in *The Yale Law Journal*, n. 66, 1985, pp. 1395-1415. Si veda inoltre E. Hilgendorf, *The dilemma of autonomous driving: Reflections on the moral and legal treatment of automatic collision avoidance systems*, cit., p. 60.

[3] Sul rapporto fra dilemmi morali e diritto penale, si veda O. Di Giovine, *Dilemmi morali e diritto penale. Istruzioni per un uso giuridico delle emozioni*, il Mulino, 2022.

[4] P. Gomes Rêgo de Almeida, C. Denner dos Santos & J. Silva Farias, *Artificial Intelligence Regulation: a framework for governance*, in *Ethics and Information Technology*, vol. 23, 2021, p. 507.

[5] Ivi, p. 508.

[6] N. Boucher, European Parliamentary Research Service - Scientific Foresight Unit, *What if AI regulation promoted innovation?*, PE 729.515, 2022, p. 2.

[7] Si pensi ad esempio a; AI-HILEG, *Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*, 2019; OECD, [Recommendation of the Council on Artificial Intelligence](#), OECD/LEGAL/0449; UNESCO, [Recommendation on the Ethics of Artificial Intelligence](#), SHS/BIO/REC-AIETHICS/2021; Council of Europe - CEPEJ, [European Ethical Charter on the Use of AI in Judicial Systems Council of Europe](#), 2018.

[8] Ai sensi degli standard sviluppati dalla Society for Automotive Engineers International (SAE), “J3016 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles”, Aprile 2021.

[9] National Transportation Safety Board (NTSB), [Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018](#), Highway Accident Report NTSB/HAR-19/03. Si veda anche: A. DeArman, “The Wild, Wild West: A Case Study of Self Driving Vehicle Testing in Arizona”, *Arizona Law Review*, Vol. 61, 2019.

[10] Si veda, su questa *Rivista* sul medesimo tema, C. Corridori, [Machina delinquere non potest](#), 19 maggio 2022. Si veda *ex multis*: E. Hilgendorf, *The dilemma of autonomous driving: Reflections on the moral and legal treatment of automatic collision avoidance systems*, in E. Hilgendorf & J. Feldle (Eds), *Digitization and the Law*, in *Robotik und Recht*, vol. 15, Nomos, 2018; A. Cappellini, *Machina delinquere potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018; L. D’Amico, *Intelligenza artificiale e auto a guida autonoma. Tra prevenzione primaria, colpa penale e rischio consentito*, in *Riv. It. Med. Leg.*, vol. 3, 2022.

[11] *Ordonnance n° 2021-443 du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation* (TRAT2034523R, JORF n°0089 du 15 avril 2021, Texte n° 36), 2021.

[12] M. Giuca, *Disciplinare l'intelligenza artificiale. La riforma francese sulla responsabilità penale da uso di auto a guida autonoma*, in *Archivio Penale*, vol. 2, 2022

[13] [Ezra Klein Interviews Alison Gopnik](#), The New York Times, 16 April 2021.

[14] M. B. Saponaro, *L'insostenibile leggerezza del tragico*, in *Scelte tragiche - Atti del III convegno "Medicina e diritto penale"*, Taranto, 11 dicembre 2020 a cura di Giuseppe Losappio, p. 89.

[15] Si veda *ex multis*: E. Awad et al., *Computational ethics*, in *Trends in Cognitive Sciences*, vol. 26, Issue 5, 2022; D. Leslie, The Alan Turing Institute, *Understanding artificial intelligence ethics and safety. A guide for the responsible design and implementation of AI systems in the public sector*, 2019; J. Ganascia, *Ethical System Formalization using Non-Monotonic Logics*, in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 29, 2007; S. Serafimova, *Whose morality? Which rationality? Challenging artificial intelligence as a remedy for the lack of moral enhancement*, in *Humanities and Social Sciences Communication*, 2020; P. Schramowski et al., *The Moral Choice Machine*, in *Frontiers in Artificial Intelligence*, vol. 3, 2020; V. Charisi, *Towards Moral Autonomous Systems*, in *arXiv:1703.04741v3*, 2017.

[16] A. Martinho, *Perspectives about artificial moral agents*, in *AI and Ethics*, vol. 1, 2021, p. 481.

[17] L'esperimento è accessibile presso il sito <https://www.moralmachine.net/>. I risultati dell'esperimento sono discussi in: E. Awad et al., *The Moral Machine experiment*, in *Nature*, Vol. 563, 2018; E. Awad et al., *Crowdsourcing Moral Machines*, in *Communications of the ACM*, vol. 63, n. 3, 2020, pp. 48-55; E. Awad et al., *Universals and variations in moral decisions made in 42 countries by 70,000 participants*, in *PNAS*, vol. 117, n. 5, 2020, pp. 2332-2337.

[18] Awad et al., *The Moral Machine experiment*, cit., p. 63.

[19] M. B. Saponaro, cit., p. 89.

[20] E. Mokhtarian, *The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence*, in *Vanderbilt Journal of Entertainment and Technology Law*, vol. 21, 2020, p. 173.

[21] Awad et. al, *When Is It Acceptable to Break the Rules? Knowledge Representation of Moral Judgement Based on Empirical Data*, in *arXiv:2201.07763*, 2022.

[22] I. Gabriel, *Artificial Intelligence, Values, and Alignment*, in *Minds and Machines*, vol. 30, 2020, pp. 411-437

[23] E. Hunt, *Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter*, *The Guardian*, 24 March 2016. t

[24] Mokhtarian, cit., p. 173.

[25] A. Guerra, F. Parisi & D. Pi, *Liability for robots I: legal challenges*, in *Journal of Institutional Economics*, vol. 18, n. 3, 2021, p.10.

[26] Bundestag, Gesetz zur Änderung des Straßenverkehrsgesetzes, 2021.

[27] A. Kriebitz, R. Max & C. Lütge, *The German Act on Autonomous Driving: Why Ethics Still Matter*, in *Philosophy & Technology*, vol. 35, 2022, p. 11.

[28] “(2) Kraftfahrzeuge mit autonomer Fahrfunktion müssen über eine technische Ausrüstung verfügen, die in der Lage ist,

[...] 2. selbstständig den an die Fahrzeugführung gerichteten Verkehrsvorschriften zu entsprechen und die über ein System der Unfallvermeidung verfügt, das

a) auf Schadensvermeidung und Schadensreduzierung ausgelegt ist,

b) bei einer unvermeidbaren alternativen Schädigung unterschiedlicher Rechtsgüter die Bedeutung der Rechtsgüter berücksichtigt, wobei der Schutz menschlichen Lebens die höchste Priorität besitzt, und

c) für den Fall einer unvermeidbaren alternativen Gefährdung von Menschenleben keine weitere Gewichtung anhand persönlicher Merkmale vorsieht”.

[29] Il report è disponibile presso: https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile. Gli stessi ideatori della *Moral Machine* fanno riferimento al lavoro della Commissione etica tedesca e affermano che il loro esperimento rappresenta il primo e unico tentativo di fornire linee guida ufficiali per le scelte etiche dei veicoli autonomi.

[30] L’esempio è fatto dalla Commissione Etica a p. 17 del Report.

[31] Report, cit., p.11.

[32] Ibid.

[33] M. B. Saponaro, cit., p. 90.

* Alice Giannini è dottoranda di ricerca in diritto penale presso l’Università di Firenze e la Maastricht University. Le riflessioni proposte in questo articolo fanno parte della sua tesi di

ricerca, che si concentra sulla responsabilità penale dei sistemi di intelligenza artificiale. È stata la vincitrice della prima edizione del premio Giulia Cavallone, istituito dalla Fondazione Calamandrei e dalla famiglia Cavallone per onorare la memoria della dott.ssa Cavallone, giudice del Tribunale di Roma e dottoressa di ricerca in diritto e procedura penale.

[1] H. Hilgendorf, *The dilemma of autonomous driving: Reflections on the moral and legal treatment of automatic collision avoidance systems*, in E. Hilgendorf & J. Feldle (Eds), *Digitization and the Law*, Robotik und Recht, vol. 15, Nomos, 2018, p. 60.

[1] L'invenzione di questo tipo di *dilemma* è attribuita a Philippa Foot. Il termine “*Trolley Problem*” è stato coniato più tardi da Judith Jarvis Thomson. Cfr. P. Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, in *Oxford Review*, Vol. 51978; J. J. Thomson, *The Trolley Problem*, in *The Yale Law Journal*, n. 66, 1985, pp. 1395-1415. Si veda inoltre E. Hilgendorf, *The dilemma of autonomous driving: Reflections on the moral and legal treatment of automatic collision avoidance systems*, cit., p. 60.

[1] Sul rapporto fra dilemmi morali e diritto penale, si veda O. Di Giovine, *Dilemmi morali e diritto penale. Istruzioni per un uso giuridico delle emozioni*, il Mulino, 2022.

[1] P. Gomes Rêgo de Almeida, C. Denner dos Santos & J. Silva Farias, *Artificial Intelligence Regulation: a framework for governance*, in *Ethics and Information Technology*, vol. 23, 2021, p. 507.

[1] Ivi, p. 508.

[1] N. Boucher, European Parliamentary Research Service - Scientific Foresight Unit, *What if AI regulation promoted innovation?*, PE 729.515, 2022, p. 2.

[1] Si pensi ad esempio a; AI-HILEG, *Ethics Guidelines for Trustworthy Artificial Intelligence (AI)*, 2019; OECD, [Recommendation of the Council on Artificial Intelligence](#), OECD/LEGAL/0449; UNESCO, [Recommendation on the Ethics of Artificial Intelligence](#), SHS/BIO/REC-AIETHICS/2021; Council of Europe - CEPEJ, [European Ethical Charter on the Use of AI in Judicial Systems Council of Europe](#), 2018.

[1] Ai sensi degli standard sviluppati dalla Society for Automotive Engineers International (SAE), “J3016 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles”, Aprile 2021.

[1] National Transportation Safety Board (NTSB), [*Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, Tempe, Arizona, March 18, 2018*](#), Highway Accident Report NTSB/HAR-19/03. Si veda anche: A. DeArman, “The Wild, Wild West: A Case Study of Self Driving Vehicle Testing in Arizona”, *Arizona Law Review*, Vol. 61, 2019.

[1] Si veda, su questa *Rivista* sul medesimo tema, C. Corridori, [*Machina delinquere non potest*](#), 19 maggio 2022. Si veda *ex multis*: E. Hilgendorf, *The dilemma of autonomous driving: Reflections on the moral and legal treatment of automatic collision avoidance systems*, in E. Hilgendorf & J. Feldle (Eds), *Digitization and the Law*, in *Robotik und Recht*, vol. 15, Nomos, 2018; A. Cappellini, *Machina delinquere potest? Brevi appunti su intelligenza artificiale e responsabilità penale*, in *Criminalia*, 2018; L. D’Amico, *Intelligenza artificiale e auto a guida autonoma. Tra prevenzione primaria, colpa penale e rischio consentito*, in *Riv. It. Med. Leg.*, vol. 3, 2022.

[1] *Ordonnance n° 2021-443 du 14 avril 2021 relative au régime de responsabilité pénale applicable en cas de circulation d'un véhicule à délégation de conduite et à ses conditions d'utilisation* (TRAT2034523R, JORF n°0089 du 15 avril 2021, Texte n° 36), 2021.

[1] M. Giuca, *Disciplinare l'intelligenza artificiale. La riforma francese sulla responsabilità penale da uso di auto a guida autonoma*, in *Archivio Penale*, vol. 2, 2022

[1] [*Ezra Klein Interviews Alison Gopnik*](#), The New York Times, 16 April 2021.

[1] M. B. Saponaro, *L'insostenibile leggerezza del tragico*, in *Scelte tragiche - Atti del III convegno “Medicina e diritto penale”*, Taranto, 11 dicembre 2020 a cura di Giuseppe Losappio, p. 89.

[1] Si veda *ex multis*: E. Awad et al., *Computational ethics*, in *Trends in Cognitive Sciences*, vol. 26, Issue 5, 2022; D. Leslie, The Alan Turing Institute, *Understanding artificial intelligence ethics and safety. A guide for the responsible design and implementation of AI systems in the public sector*, 2019; J. Ganascia, *Ethical System Formalization using Non-Monotonic Logics*, in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 29, 2007; S. Serafimova, *Whose morality? Which rationality? Challenging artificial intelligence as a remedy for the lack of moral enhancement*, in *Humanities and Social Sciences Communication*, 2020; P. Schramowski et al., *The Moral Choice Machine*, in *Frontiers in Artificial Intelligence*, vol. 3, 2020; V. Charisi, *Towards Moral Autonomous Systems*, in *arXiv:1703.04741v3*, 2017.

[1] A. Martinho, *Perspectives about artificial moral agents*, in *AI and Ethics*, vol. 1, 2021, p. 481.

[1] L’esperimento è accessibile presso il sito <https://www.moralmachine.net/>. I risultati dell’esperimento sono discussi in: E. Awad et al., *The Moral Machine experiment*, in *Nature*, Vol.

563, 2018; E. Awad et al., *Crowdsourcing Moral Machines*, in *Communications of the ACM*, vol. 63, n. 3, 2020, pp. 48-55; E. Awad et al., *Universals and variations in moral decisions made in 42 countries by 70,000 participants*, in *PNAS*, vol. 117, n. 5, 2020, pp. 2332-2337.

[1] Awad et al., *The Moral Machine experiment*, cit., p. 63.

[1] M. B. Saponaro, cit., p. 89.

[1] E. Mokhtarian, *The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence*, in *Vanderbilt Journal of Entertainment and Technology Law*, vol. 21, 2020, p. 173.

[1] Awad et. al, *When Is It Acceptable to Break the Rules? Knowledge Representation of Moral Judgement Based on Empirical Data*, in *arXiv:2201.07763*, 2022.

[1] I. Gabriel, *Artificial Intelligence, Values, and Alignment*, in *Minds and Machines*, vol. 30, 2020, pp. 411-437

[1] E. Hunt, *Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter*, *The Guardian*, 24 March 2016. t

[1] Mokhtarian, cit., p. 173.

[1] A. Guerra, F. Parisi & D. Pi, *Liability for robots I: legal challenges*, in *Journal of Institutional Economics*, vol. 18, n. 3, 2021, p.10.

[1] Bundestag, *Gesetz zur Änderung des Straßenverkehrsgesetzes*, 2021.

[1] A. Kriebitz, R. Max & C. Lütge, *The German Act on Autonomous Driving: Why Ethics Still Matter*, in *Philosophy & Technology*, vol. 35, 2022, p. 11.

[1] “(2) Kraftfahrzeuge mit autonomer Fahrfunktion müssen über eine technische Ausrüstung verfügen, die in der Lage ist,

[...] 2. selbstständig den an die Fahrzeugführung gerichteten Verkehrsvorschriften zu entsprechen und die über ein System der Unfallvermeidung verfügt, das

a) auf Schadensvermeidung und Schadensreduzierung ausgelegt ist,

b) bei einer unvermeidbaren alternativen Schädigung unterschiedlicher Rechtsgüter die Bedeutung der Rechtsgüter berücksichtigt, wobei der Schutz menschlichen Lebens die höchste Priorität besitzt, und

c) für den Fall einer unvermeidbaren alternativen Gefährdung von Menschenleben keine weitere Gewichtung anhand persönlicher Merkmale vorsieht”.

[1] Il report è disponibile presso: https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile. Gli stessi ideatori della *Moral Machine* fanno riferimento al lavoro della Commissione etica tedesca e affermano che il loro esperimento rappresenta il primo e unico tentativo di fornire linee guida ufficiali per le scelte etiche dei veicoli autonomi.

[1] L’esempio è fatto dalla Commissione Etica a p. 17 del Report.

[1] Report, cit., p.11.

[1] Ibid. [1] M. B. Saponaro, cit., p. 90.