*Leonardo Cracco*

# Machine ethics: stato dell'arte e osservazioni critiche

Abstract: *This essay aims to introduce the field of study known as machine ethics, viz. the aspect of the ethics of artificial intelligence concerned with the moral behavior of AI systems. I discuss the present potential of this technology and put forward some ethical considerations about the benefits and perils deriving from the development of this field of research. There is an urge of debate, considering that there is an increase in machine usage in ethically-sensitive domains, to ensure that future technology becomes advantageous for humans and their well-being.*

Keywords: *Machine ethics, machine morality, friendly AI, artificial intelligence.*

## 1. Introduzione

L'*etica automatica* – o *machine ethics*[1] – ha come obiettivo principale quello di fornire alle macchine la capacità di prendere decisioni etiche in maniera esplicita, così da renderle autonome in contesti moralmente sensibili.

La disciplina non si occupa di *software* o di sistemi di intelligenza artificiale che 'agiscono' in maniera eticamente accettabile perché le loro funzioni interne implicano di *default* un comportamento che esternamente giudicheremmo morale, senza fare riferimento a valori o teorie etiche. L'obiettivo non è, in altri termini, quello di creare macchine eticamente allineate, ma quello di fornire alla macchina una capacità di formulare giudizi morali complessi in autonomia e in linea di principio 'nuovi '[2].

Nell'ottica per cui la supervisione umana è destinata a diminuire e l'autonomia delle macchine ad aumentare, potrebbe non bastare generare algoritmi nei limiti

1 Nota anche come *machine morality* o *friendly AI*. W. Wallach, S. Franklin, C. Allen, *A conceptual and computational model of moral decision making in human and artificial agents*, in "Topics in Cognitive Science", 2010, 2, pp. 454-485.

2 La plausibilità e la moralità di questa impresa sono questioni che vengono poste all'interno della *machine metaethics*, una riflessione di secondo grado che sottopone gli obiettivi e i modelli della *machine ethics* a uno scrutinio filosofico di più ampio respiro. Cosa significa aggiungere una dimensione etica alle macchine? È desiderabile farlo? L'etica è computabile? C'è una sola teoria etica corretta che dovremmo provare a implementare? S. L. Anderson, *Asimov's "three laws of robotics" and machine metaethics*, in "AI and Society", 2008, 22(4), pp. 477-493.

**Filosofia**, anno LXV, 2020, pp. 75-89 • Mimesis Edizioni, Milano-Udine
Web: https://www.ojs.unito.it/index.php/filosofia • ISSN (print): 0015-1823, ISSN (online): 2704-8195 • DOI: 10.13135/2704-8195/5234

dei principi etici; sarà necessario infondere i principi morali negli algoritmi stessi. Questo è un modo originale per preoccuparsi di meno che le macchine intelligenti in grado di svolgere un ruolo significativo nel migliorare la qualità della nostra vita non impattino negativamente sul nostro benessere.

Il campo della *machine ethics* è relativamente giovane e non ha ancora prodotto, a livello pratico, nulla che possa considerarsi minimamente all'altezza delle promesse fatte inizialmente. Il dibattito accademico sta ampiamente anticipando i tempi, tanto che la maggior parte degli articoli che si occupano di *machine ethics* non trattano dei tentativi concreti fatti per costruire 'macchine etiche', ma piuttosto di che cosa implichi un tale tentativo e se sia desiderabile intraprenderlo (*machine metaethics*).

Secondo molti autori ci sono buone ragioni per giustificare i tentativi di ingegneri, scienziati e accademici di costruire macchine che seguano uno o più principi etici per guidare il proprio comportamento[3]. I motivi solitamente addotti in favore della disciplina sono quattro.

Il primo è che i sistemi autonomi attualmente in via di sviluppo – come le auto a guida autonoma, i robot per l'assistenza sanitaria o i droni da guerra – dovranno prendere decisioni morali estremamente complesse. Dare a questi sistemi una capacità di ragionamento etico permetterà di garantire che le loro decisioni saranno moralmente accettabili[4].

Sebbene plausibile, questa opinione potrebbe derivare da un pregiudizio antropologico: quando osserviamo dei comportamenti lodevoli, questi sono spesso frutto di un ragionamento morale, o di un'assunzione più o meno razionale di alcuni principi etici. Ma non sembra esserci alcuna connessione necessaria tra il ragionamento morale e le decisioni etiche. Non solo nelle macchine, ma anche per gli esseri umani; la capacità di ragionamento morale non garantisce di per sé decisioni moralmente buone, e le decisioni moralmente desiderabili possono non essere generate dal ragionamento morale.

Solitamente si cerca di riformulare la questione in termini di complessità dei contesti di applicazione delle IA: ambienti più complessi e con più fattori generano, secondo gli autori, una complessità morale. La 'complessità morale' è intesa come difficoltà nella scelta dei principi per agire, come ad esempio la scelta tra due doveri *prima facie* validi in conflitto tra loro. Nel paragrafo 3 si vedrà un esempio di un sistema artificiale che gestisce il conflitto tra il dovere di proteggere la salute di un paziente e il dovere di rispettare la sua autonomia, attraverso principi codificati esplicitamente.

Il secondo motivo è che una macchina in grado di produrre ragionamenti etici espliciti sarebbe in grado di spiegare, e in un certo senso giustificare, i giudizi

3 Ivi, p. 3. In letteratura il termine 'macchina' viene utilizzato nella sua accezione più ampia, per includere sia le macchine fisiche – i robot autonomi – sia sistemi puramente algoritmici senza estensioni fisiche.

4 C. Allen, W. Wallach, I. Smit, *Why machine ethics?*, in "IEEE Intelligent Systems", 2006, 21(4), pp. 12-17.

che ha prodotto[5]. Dato che alcune macchine intelligenti andranno inserite in un tessuto sociale popolato da persone, il fatto che potranno esplicitare i propri ragionamenti, spiegare il motivo delle loro decisioni e indicare il valore alla luce del quale hanno preferito un corso d'azione rispetto a un altro, potrebbe renderle socialmente più adatte. È possibile che i rapporti futuri delle persone con le macchine diventino sempre più sfumati, fluidi e personalizzabili. Si creerebbe un sistema sociale di fiducia, in cui le parti potranno chiedere spiegazioni sul perché di determinate azioni e cercare di riparare il danno in caso di errore.

Una terza ragione è che le macchine con una capacità di ragionamento etico potrebbero anche migliorare il comportamento etico degli umani, e potrebbero farlo in due modi diversi: migliorando le singole decisioni umane grazie alla loro capacità di elaborare un maggior numero d'informazioni[6]; e/o migliorando la moralità umana nel suo insieme, aiutandola a formulare teorie etiche più chiare e coerenti e a ottenere un consenso maggiore sui dilemmi morali. Curiosamente, come si mostrerà nel paragrafo 4, molti autori hanno sostenuto l'opposto; ovvero che macchine in grado di prendere decisioni in campo etico eroderanno la morale sia degli individui sia delle comunità.

Infine, si sostiene che iniziare a comprendere la computabilità della morale servirà per prevenire i pericoli di una superintelligenza – ovvero di un intelletto le cui prestazioni cognitive superano di gran lunga quelle degli esseri umani o di altre entità artificiali con una intelligenza di livello umano[7] – non allineata[8].

Le motivazioni fornite potrebbero rivelarsi inconsistenti e i benefici auspicati incapaci di controbilanciare i pericoli. Tuttavia, l'impresa della *machine ethics* sembra avere dei motivi *prima facie* validi per essere perseguita.

## 2. Automatizzare l'assistenza sanitaria

Incorporare un comportamento etico nelle macchine – ovvero comprendere il processo decisionale morale e riuscire a formalizzarlo – è probabilmente uno dei compiti più impegnativi tra gli approcci computazionali alla cognizione di ordine superiore.

Attualmente gli sforzi compiuti in questa direzione si sono concentrati maggiormente sull'etica medica[9]. Ciò per due ragioni fondamentali: perché grazie alla

5 J. Bryson, A. Winfield, *Standardizing ethical design for artificial intelligence and autonomous systems*, in "Computer", 2017, 50(5), pp. 116-119.

6 C. Allen, G. Varner, J. Zinser, *Prolegomena to any future artificial moral agent*, in "Journal of Experimental and Theoretical Artificial Intelligence", 2000, 12, pp. 251-261.

7 N. Bostrom, E. Yudkowsky, *The ethics of artificial intelligence*, in *The Cambridge handbook of artificial intelligence*, a cura di W. Ramsey, K. Frankish, Cambridge, Cambridge University Press, 2014, pp. 1-22.

8 C. Shulman, H. Jonsson, N. Tarleton, *Machine ethics and superintelligence*, in "5th Asia-Pacific Computing and Philosophy Conference", 2013, pp. 1-5.

9 M. A. Pontier, J.F. Hoorn, *Toward machines that behave ethically better than humans do*, in "Proceedings of the 34th International Annual Conference of the Cognitive Science Society",

lunga tradizione della disciplina vi è un certo consenso su cosa sia considerabile come 'comportamento eticamente corretto'; e perché l'assistenza sanitaria sarà necessaria in tutti quei paesi dove l'invecchiamento della popolazione è veloce e il personale sanitario umano scarseggia. Gestire la domanda di nuovi assistenti sanitari mantenendo *standard* etici elevati richiederà l'utilizzo di robot per la cura, già oggi in uso non solo per aiutare gli assistenti nei loro molteplici compiti, ma anche per il trattamento terapeutico.

Il campo interdisciplinare di studio dell'interazione tra umani e agenti artificiali, lo *human-robot interaction* (HRI), cerca già da molti anni il modo di migliorare il rapporto tra i pazienti e i robot. Ad esempio, in uno studio del 2005, è stato utilizzato un robot per interagire con bambini con disturbi dello spettro autistico. Questi bambini hanno difficoltà nei comportamenti sociali e hanno interessi molto ristretti; tuttavia manifestano un curioso interesse a interagire socialmente con le macchine e in generale ne vengono stimolati in modo positivo. In questo studio specifico il robot è da intendersi come un catalizzatore per l'interazione sociale tra bambini e genitori o altri bambini, e non come interlocutore autonomo[10].

Il Giappone, anche a causa del rapido invecchiamento della sua popolazione[11], spicca tra i paesi del mondo per lo sforzo di automatizzare in maniera eticamente sostenibile l'assistenza sanitaria. Un istituto giapponese ha creato Paro, un robot con le sembianze di una foca, capace di interagire con i pazienti; ha dimostrato di essere in grado di migliorare gli stati d'animo di anziani con demenza senile o con l'Alzheimer, rendendoli anche più attivi e comunicativi tra di loro e con gli operatori sanitari[12].

Molti autori hanno espresso preoccupazioni etiche nell'affidare ai robot il dominio dell'assistenza di fasce della popolazione vulnerabili come i malati, i disabili e gli anziani, in quanto la cura sembra richiedere qualità umane, come l'empatia, totalmente assenti nelle macchine. Joseph Weizenbaum, uno dei padri dell'intelligenza artificiale, ha sottolineato la differenza tra decidere e scegliere. Mentre decidere è un'attività computazionale facilmente programmabile, la scelta è il prodotto del

2012, 12, pp. 2198-2203.

10 B. Robins, K. Dautenhahn, R.T. Boekhorst, A. Billard, *Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?*, in "Universal Access in the Information Society", 2005, 4(2), pp. 105-120. Le architetture di controllo dell'HRI non erano molto avanzate nel 2005, e non permettevano complesse interazioni. Tuttavia un semplice robot con sembianze umane, in grado di produrre bolle di sapone, qualche movimento e semplici parole di incoraggiamento, è riuscito a incentivare comportamenti sociali nei bambini autistici.

11 Secondo i dati del ministero giapponese degli affari interni nel 2016 il 27,3% della popolazione totale possedeva un'età pari o superiore a 65 anni. Secondo le proiezioni con l'attuale tasso di fertilità, gli over 65 rappresenteranno il 40% della popolazione entro il 2060. Statistics Bureau of Japan, *Statistical Handbook of Japan*, Tokyo, Statistical Bureau of Japan, 2017, p. 8.

12 Più in generale gli studi stanno mostrando che animali di compagna artificiali come Paro possono combattere la solitudine tanto quanto gli animali vivi, senza però richiedere le stesse cure. H. Robinson, B. Macdonald, N. Kerse, E. Broadbent, *The psychosocial effects of a companion robot: a randomized controlled trial*, in "Journal of the American Medical Director Association", 2013, 14(9), pp. 661-667.

giudizio cosciente, ed è ciò che secondo l'autore ci rende umani[13]. Questa distinzione l'ha portato a sostenere che le IA non dovrebbero essere usate per sostituire le persone in posizioni che richiedono rispetto, cura ed empatia. Se le macchine le sostituiranno, ci troveremo alienati, svalutati, frustrati e privati della nostra dignità[14].

Queste preoccupazioni sono legittime; e anche qualora si sostenesse la necessità di automatizzare i campi oggi considerati moralmente *sensibili* (come l'assistenza sanitaria), i problemi da risolvere o da analizzare sono ancora molti. I pazienti potrebbero affezionarsi ai robot, così che un allontanamento successivo potrebbe recare a loro maggiori disagi; gli stessi robot, almeno allo stato attuale, hanno capacità molto limitate, e potrebbero in breve tempo frustrare le persone con cui interagiscono; l'impiego di robot potrebbe ledere ulteriormente il legame con gli assistenti o con i parenti.

Opporsi a questa transizione, già ampiamente in atto, appare però irrealistico. Si può però lavorare perché la progettazione e l'introduzione dei robot sia mirata alla promozione dei valori e della dignità dei pazienti in un momento così vulnerabile e delicato della loro vita. Ad oggi, inoltre, si può pensare che, se progettati correttamente, i robot potrebbero migliorare la condizione di milioni di indigenti, combattendone la solitudine, aumentandone la mobilità e sottraendoli ai soprusi e alle violenze che potrebbero subire da altri esseri umani.

In letteratura sono già state compiute riflessioni tecniche su come poter valutare le possibili implicazioni nell'introdurre robot per la cura delle persone. Ad esempio, è stato proposto un *framework* con cui valutare i futuri robot in base a differenti fattori: il contesto in cui sono inseriti, i compiti che devono svolgere, i tipi di pazienti con cui devono interagire e altri fattori rilevanti. Un robot come Robear, in grado di sollevare, trasportare o aiutare a far camminare i pazienti, è estremamente diverso da Paro – che è essenzialmente un robot da compagnia con cui il paziente può instaurare un legame – e richiede quindi criteri etici differenti per essere valutato[15]. Nel progetto europeo ROBOT-ERA, invece, si è discusso il ruolo dei *personal care* robot direttamente con gli anziani e i loro parenti. L'opinione comune è che debbano essere utilizzati per valorizzare i rapporti tra la persona anziana, i familiari e i 'prestatori di cura' – i *caregiver*, sia professionali che non – e non per sostituirli completamente[16].

Wallach e Allen hanno una posizione più radicale: almeno in quelle comunità dove non ci saranno le risorse per rispondere ai bisogni e ai desideri dei pazienti

13 J. Weizenbaum, *Computer power and human reason: from judgment to calculation*, San Francisco, W.H. Freeman, 1976.

14 Curiosamente in ogni ruolo *sensibile* citato dall'autore, si sta cercando già da molti anni di realizzare IA per sostituire l'essere umano: come assistenti per gli anziani, soldati e anche giudici.

15 Il *framework* è stato proposto da A. van Wynsberghe, *Designing robots for care; care centered value-sensitive design*, in "Journal of Science and Engineering Ethics", 2013, 19(2), pp. 407-433.

16 A. Pirni, R. Esposito, A. Carnevale, F. Cavallo, *"Sostenibilità etica" dei personal care robot. Linee per un inquadramento preliminare*, in "Nuova Corrente", 2017, 159, pp. 133-151. È stata anche riscontrata una diversità di genere nell'apprezzamento di queste tecnologie: gli uomini sono più propensi a utilizzare e giudicare positivamente i supporti robotici rispetto alle donne; ivi, p. 146.

e a garantire loro un trattamento dignitoso, i robot saranno la migliore alternativa possibile[17]. Ed è proprio in questo quadro che si inserisce la *machine ethics*.

## 3. I tentativi compiuti in *machine ethics*

Oltre a migliorare il *software* e l'*hardware* dei futuri *care-robot*, gli autori del campo sostengono la necessità di fornire ai robot delle teorie o dei modelli morali attraverso cui ragionare. La ricerca è attiva dal 2005, ed è costellata da tentativi più o meno soddisfacenti; offrirne una revisione completa va aldilà dello scopo di questo paragrafo, data la diversità degli approcci, delle teorie utilizzate, degli scopi perseguiti e delle tecnologie impiegate[18]. Alcuni tentativi sono stati fatti con teorie normative quali l'utilitarismo, l'imperativo categorico di Kant, i doveri *prima facie* di Ross e l'etica delle virtù di stampo aristotelico; per costruire sistemi – per citare alcuni esempi – in grado di raggruppare dilemmi morali simili, prendere decisioni morali esplicite in domini ristretti o individuare principi morali con cui poi poter agire.

Semplificando il complesso mosaico dei tentativi compiuti, si possono osservare tre strategie con cui si sta tentando di dotare le macchine di un'etica: partendo dall'alto (*top-down*), dal basso (*bottom-up*) o adottando un ibrido tra le due strategie[19].

La strategia *top-down* prevede la formalizzazione di alcune regole e principi generali così da poterli fornire alla macchina. Dato che le tre leggi di Asimov – uno dei sistemi deontologici non religiosi più semplice a nostra disposizione – sono inadatte a questo compito[20], i ricercatori hanno provato ad adattare le teorie filosofiche classiche. Si è però notato che tutti gli approcci soffrono di almeno due problemi: uno specifico della teoria, l'altro dovuto a difficoltà che si hanno nel formalizzarla[21].

L'approccio deontologico considera la moralità un sistema di diritti e doveri e le diverse azioni sono considerate non ammissibili, ammissibili o obbligatorie in base

17 W. Wallach, C. Allen, *Moral machines: teaching robots right from wrong*, Oxford/New York, Oxford University Press, 2009, p. 45.

18 Probabilmente è per questo motivo che ad oggi non esiste uno studio sistematico di tutti gli esperimenti condotti in *machine ethics*. Per un'indagine parziale si veda L. M. Pereira, A. Saptawijaya, *Programming machine ethics*, Cham, Springer Publishing Company, 2016, vol. 26, cap. 2; e H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser, Q. Yang, *Building ethics into artificial intelligence*, in "Proceedings of the 27th International Joint Conference on Artificial Intelligence", 2018, pp. 5527-5533.

19 C. Allen, W. Wallach, I. Smit, *Why machine ethics?*, cit.

20 S. L. Anderson, *Asimov's "three laws of robotics" and machine metaethics*, cit. L'ambiguità intrinseca rende le tre leggi inservibili per compiti complessi. Sono troppo imprecise per dare risposte univoche in quasi ogni situazione pratica e la vaghezza dei termini che le compongono potrebbe causare problemi; quando ad esempio si sostiene che "un robot non può recare danno a un essere umano", con danno si intende solo il dolore fisico? O anche quello psicologico o legato all'ingiustizia sociale?

21 J.D. Greene, F. Rossi, J. Tasioulas, B.K. Venable, B. Williams, *Embedding ethical principles in collective decision support systems*, in *Thirtieth AAAI Conference on Artificial Intelligence* (2013), pp. 4147-4151.

a un insieme di regole esplicite. Tuttavia, non è ancora chiaro come si possa formalizzare anche una semplicissima regola come 'non uccidere', dato che il comando deve tradursi per guidare l'azione pratica del robot; come, ad esempio, un divieto di compiere determinati movimenti quando si è in determinate situazioni. Regole così formalizzate sembrano però inadatte. La maggior parte delle regole morali che gli esseri umani utilizzano con semplicità nella vita quotidiana, necessitano in realtà di un senso comune per essere applicate, che si presta male all'interpretazione più letterale delle macchine[22].

L'approccio consequenzialista mira invece a produrre le migliori conseguenze per ogni individuo, secondo una certa funzione di utilità. È associato a Jeremy Bentham, e nella sua versione più classica mira a massimizzare la quantità aggregata di felicità[23]. Il problema in questo caso è che i calcoli potrebbero essere estremamente complessi e la definizione rigorosa di un metro per misurare i risultati attesi è ad oggi una questione irrisolta. Attualmente non è neanche del tutto chiaro come si possa far rappresentare alla macchina tutte le possibili conseguenze pratiche di una data situazione complessa, così da permetterle di calcolare quale sia la migliore.

Nel 2005 è stato elaborato un primo prototipo di nome *Jeremy*, su cui è stato implementato un utilitarismo edonistico dell'atto. Una volta inseriti manualmente i nomi di tutte le persone coinvolte in un'azione X, la quantità di piacere che ognuna di essa dovrebbe sperimentare (secondo una scala che va da 2, molto piacevole, a -2, molto spiacevole), la durata del piacere (in ore o in giorni) e la probabilità con cui lo sperimenterà (molto probabile, abbastanza probabile o improbabile), *Jeremy* è in grado di stimare se X sia meglio di altre azioni – ad esempio Y o Z – anch'esse compilate manualmente. Questo modello è un semplice programma – non ha tecniche di intelligenza artificiale in senso stretto – e sebbene sia scomodo e assolutamente non-autonomo, è in linea teorica più affidabile dell'essere umano nel prendere decisioni utilitaristiche[24]. Nello stesso studio è stato costruito anche un modello sul calco dell'etica dei doveri *prima facie* di William D. Ross[25] – il modello *W.D.* – che si basa invece sulla *programmazione logica induttiva*, una tecnica di IA classica. I doveri *prima facie* si contrappongono ai doveri assoluti in quanto possono sempre essere annullati da uno più urgente. Dato che non vi è una chiara classificazione dei doveri, serve una procedura decisionale per determinare l'azione eticamente corretta nei casi in cui i doveri confliggono tra loro. Il modello informatico ha utilizzato un algoritmo di apprendimento per adeguare i giudizi in base ai doveri *prima facie* forniti e alle intuizioni degli esperti su casi simili. Queste

22 N.J. Goodall, *Machine ethics and automated vehicles*, in *Road vehicle automation*, a cura di G. Meyer, S. Beiker, Cham, Springer International Publishing, 2014, pp. 93-102.

23 J. Bentham, *Introduzione ai principi della morale e della legislazione* (1789), tr. it. di E. Lecaldano, Torino, Utet, 1998.

24 M. Anderson, S.L. Anderson, C. Armen, *Toward machine ethics: implementing two action-based ethical theories*, in "Proceedings of the AAAI 2005 Fall Symposium on Machine Ethics", 2005, pp. 1-7. Tuttavia il sistema eredita tutti i limiti umani nella misurazione dei dati da fornire alla macchina.

25 W.D. Ross, *Il giusto e il bene* (1930), tr. it. di R. Mordacci, Milano, Bompiani, 2004.

intuizioni sono state raccolte a loro volta attraverso l'analisi da parte della macchina di molti esempi forniti in fase di addestramento. *W.D.* è stato così in grado di fornire un giudizio in accordo con gli esperti anche su dilemmi nuovi non incontrati in precedenza[26].

L'approccio basato sull'etica delle virtù di stampo aristotelico considera il comportamento etico come il prodotto di una serie di disposizioni interiorizzate che non possono essere ridotte a una serie di regole prestabilite da seguire pedissequamente, né a un impegno di massimizzare l'utilità aggregata. Questa prospettiva ha il problema di essere poco definita e difficilmente insegnabile attraverso semplici comandi. D'altra parte, per Aristotele la ricerca in etica non aveva come fine ultimo una conoscenza pura[27].

A questa strategia si contrappone quella *bottom-up*, nella quale, invece di immettere le linee guida della morale direttamente nel codice della macchina, si lascia che sia l'IA stessa a 'impararla' da alcuni esempi forniti. Goodall[28] ha mostrato che c'è un parallelismo tra questo approccio e altre applicazioni del passato. Ad esempio, la traduzione del linguaggio si è a lungo basata su regole sviluppate da esperti e poi immesse nella macchina. Con l'avvento degli algoritmi di apprendimento automatico, non c'è stato più bisogno di produrre regole formali; è la macchina stessa a estrarle dai dati forniti, anche con prestazioni migliori delle nostre. Come è successo per la linguistica, anche in etica le tecniche di intelligenza artificiale possono rivelarsi utili: in entrambi i casi non abbiamo a disposizione regole abbastanza articolate da comprendere ogni circostanza in cui si potrà trovare una macchina, ma le IA potrebbero impararle autonomamente attraverso l'osservazione di azioni umane, identificando in esse le caratteristiche che le rendono etiche. Così facendo si rivaluta anche il ruolo fondamentale dell'esperienza, dell'apprendimento e dell'interiorizzazione del codice morale, quasi del tutto assente nelle strategie *top-down*.

Ma se il rischio per la prima strategia era quello di fornire codici incoerenti, limitati o eticamente inaccettabili, il rischio di quest'ultima è di allenare le IA con esempi sbagliati. La macchina potrebbe imparare fin troppo bene dai nostri esempi – essere *overfitting* – ed essere così incapace di affrontarne di nuovi. Oppure *leggere* i nostri comportamenti alla lettera, e imparare come di fatto agiamo e non

26 *Ibidem*. Con la stessa tecnologia M. Anderson, S.L. Anderson, C. Armen, *An approach to computing ethics*, in "IEEE Intelligent Systems", 2006, 21, pp. 56-63, hanno creato *MedEthEx*, un sistema esperto che verrà presentato più avanti.

27 Con le parole di Aristotele nell'*Etica a Nicomaco* (Libro II, 2, 1103b-1104a, 25-10): "poiché dunque il presente studio non ha per fine una conoscenza pura (infatti non intraprendiamo questa ricerca per conoscere che cos'è la virtù, ma per diventare buoni, giacché altrimenti nulla sarebbe la sua utilità), è necessario esaminare le azioni per sapere come bisogna compierle. […] Prima però ci si accordi su questo punto, […]. Nel campo delle azioni e di ciò che è utile non c'è nulla di stabile, come nel campo della salute. Non c'è infatti una legge generale per i casi particolari, perché essi non rientrano in nessuna conoscenza tecnica e in nessuna regola fissa, ma spetta sempre a chi agisce tener conto di ciò che è opportuno, come avviene nell'arte della medicina e in quella della navigazione".

28 N. J. Goodall, *Ethical decision making during automated vehicle crashes*, in "Transportation Research Record", 2014, 2424(1), pp. 58-65.

come le nostre migliori intenzioni vorrebbero. Infine, la natura della tecnologia stessa porta al problema della trasparenza: sulla base degli *input* forniti spesso è difficile comprendere in che modo le reti neurali artificiali siano arrivate a un determinato *output*. A questi sistemi opachi – difficilmente prevedibili e ispezionabili – è preferibile l'utilizzo di una struttura decisionale ad albero (*decision tree*, tra i modelli predittivi più semplici a nostra disposizione), dove i passaggi logici possono essere ripercorsi fino alla loro origine, in quella che è nota come ingegneria inversa (*reverse engineering*).

Nel 2006, senza che prima gli venissero forniti dei principi morali su cui basarsi, una rete neurale artificiale – con strati di nodi per trovare relazioni complesse tra i dati forniti in *input* – è stata utilizzata per classificare come moralmente accettabili o inaccettabili alcuni casi[29]. Il risultato è stato che, a parte quelli banali, la rete non è riuscita nel compito di valutare correttamente tutti i casi nuovi forniti successivamente; probabilmente l'estrapolazione di alcuni fattori moralmente salienti non è sufficiente a esaurire l'intero dominio della morale.

Per questa ragione la strategia intermedia che cerca di conciliare la strategia *top-down* e *bottom-up*, è considerata oggi quella più promettente dalla maggior parte degli autori nel campo. L'utilizzo di tecniche di IA per far apprendere alle macchine come produrre giudizi morali – partendo da casi specifici da cui poi estrarre i principi più generali – e di tecniche di correzione dall'*alto*, sarebbe anche più simile alla capacità di giudizio morale negli esseri umani; le macchine future avranno bisogno di una moralità dinamica in grado di imparare da nuovi casi e porli in equilibrio con i principi ideali a cui tendono. Una sensibilità morale così complessa è però oggi ancora lontana. Attualmente si hanno alcuni esempi di modelli ibridi in grado di prendere decisioni in domini molto ristretti e con pochi principi[30]. Ma non offrono ancora buone prestazioni, né possono dirsi autonomi nelle loro scelte: per il momento sono rimasti quindi solo prototipi.

Nel 2006 è stato però costruito *MedEthEx*, un sistema esperto ideato per essere un consulente etico specialistico capace di dare risposte etiche in linea con quelle dei medici[31]. Attraverso l'utilizzo di un algoritmo di apprendimento automatico, i creatori del sistema hanno estratto un principio, su cui poi basare le decisioni di *MedEthEx,* da una serie di scenari particolari. Il principio prevede che un robot

29 M. Guarini, *Particularism and the classification and reclassification of moral cases*, in "Intelligent Systems", 2006, 21(4), pp. 22-28.

30 Ad esempio M. A. Pontier, J. F. Hoorn, *Toward machines that behave ethically better than humans do*. Attraverso il prisma di tre valori – in ordine di importanza: autonomia, non maleficenza e beneficenza – e alla luce delle preferenze dei pazienti, il sistema valuta alcuni dilemmi in campo biomedico e sceglie l'azione più in accordo con il parere degli esperti. La necessità di fornirgli i dati manualmente e la semplicità dei casi che valuta non lo rendono però uno strumento efficiente o realmente vantaggioso.

31 Il modello è stato creato dagli stessi ricercatori che hanno creato *Jeremy* e *W.D.*, ovvero Michael Anderson (informatico) e Susan Anderson (filosofa), due degli autori più impegnati nel campo della *machine ethics*; per un resoconto delle loro ricerche e si veda M. Anderson, S. L. Anderson, *Machine ethics: creating an ethical intelligent agent*, in "AI Magazine", 2007, 28(4), pp. 15-26.

sanitario deve mettere in discussione le decisioni del paziente – violando quindi l'autonomia del paziente – solo se nel non farlo violerebbe sia il principio di non-maleficenza che quello di beneficenza.

Una volta formalizzato, è stato testato nel 2010 all'interno del robot umanoide *Nao*, un robot sviluppato da una società francese. *Nao* è in grado di dirigersi verso un paziente a cui deve essere ricordato di prendere un farmaco, portargli il farmaco, interagire tramite il linguaggio naturale e comunicare telematicamente con il medico se necessario. Tramite i dati forniti in *input* (ora a cui somministrare il farmaco, quantità massima di danno che potrebbe verificarsi nel caso il farmaco non venisse assunto, tempo richiesto perché il danno si verifichi, quantità massima di bene attesa dall'assunzione del farmaco e tempo perché il beneficio si verifichi) *Nao* calcola il livello di accordo o violazione dei tre doveri e intraprende azioni diverse a seconda di come questi livelli cambiano nel tempo. Ad esempio, il robot notifica al medico che il paziente non prende il farmaco solo quando arriva il momento in cui il paziente potrebbe essere danneggiato, o potrebbe perdere considerevoli benefici per non aver preso il farmaco; non prima, per non violarne l'autonomia. Questo è stato il primo prototipo realizzato, basato su un solo principio e incapace di affrontare situazioni più complesse di quella descritta sopra.

Probabilmente con l'utilizzo di tecniche di IA di ultima generazione si potranno ottenere migliori risultati e a un minor costo. Nel 2018, ad esempio, un gruppo di ricercatori ha dato nuova linfa al campo utilizzando una particolare tecnica di apprendimento automatico – l'*apprendimento per rinforzo* (RL) – che, attraverso il comando di massimizzare le ricompense cumulative, spinge l'agente RL a compiere azioni sempre più accurate[32]. Hanno prima creato una funzione di ricompensa che portasse il sistema a emulare il comportamento umano, assumendo che questo fosse etico. Il sistema riceve un segnale di rinforzo (ricompensa) non appena compie un'azione positiva e viene invece punito per le decisioni etiche negative. I tre scenari in cui è stata applicata la tecnica sono molto semplici. Nel primo, ad esempio, un robot costruito per aiutarci a far la spesa deve andare verso uno scaffale e prendere una bottiglia di latte nel minor tempo possibile, evitando di ferire dei bambini o consolandoli nel caso piangano. Con l'esperienza, l'agente, che ha il solo obiettivo di trovare il percorso più breve per prendere il latte, costruisce una funzione in grado di massimizzare il rinforzo; a sua volta il rinforzo è stabilito osservando il comportamento umano, che in generale evita di ferire i bambini e li consola quando piangono; così anche la macchina ha *imparato*, per rinforzo, a fare lo stesso. In generale, la tecnica di apprendimento per rinforzo è oggi vista come una delle più promettenti per allenare le macchine[33]. Questa tecnica ha il vantaggio

32 Y.-H. Wu e S.-D. Lin, *A low-cost ethics shaping approach for designing reinforcement learning agents*, in "AAAI", 2018, pp. 1687-1694.

33 È stata utilizzata anche per *AlphaZero* e *AlphaGo Zero*, rispettivamente i sistemi più forti nel gioco degli scacchi e del Go. *AlphaZero* è una rete neurale che è stata in grado, partendo dalle sole regole degli scacchi, di raggiungere in 24 ore prestazioni sovraumane attraverso un algoritmo di RL; D. Silver, T. Hubert, J. Schrittwieser, I. Antonoglou, M. Lai, A. Guez, M. Lanctot, L. Sifre, D. Kumaran, T. Graepel, T. Lillicrap, K. Simonyan, D. Hassabis, *A general*

di sgravare i programmatori di questi agenti RL dall'onere di dover codificare l'etica, ed è anche poco costosa. Ma ha anche due svantaggi: il primo è che basarsi su pochi dati aumenta il rischio di *bias* nel sistema, e aumentare la mole dei dati è molto costoso. Il secondo è che il comportamento delle persone su cui si basa il sistema potrebbe essere subottimale e quindi non essere una base su cui fare affidamento.

È probabile che serviranno molti altri tentativi prima di trovare il modo migliore per conferire abilità morali alle macchine, e che i primi modelli che usciranno, implementati su armi automatiche o robot per la cura, saranno estremamente limitati.

## 4. I pericoli della *machine ethics*

Va infine indagato se il tentativo della *machine ethics* sia necessario rispetto ad altre opzioni disponibili, come quella di lasciare la macchina eticamente analfabeta e concentrarsi nel progettarla e crearla nei limiti della morale. Che sia in veste di consulente o di decisore morale, è veramente desiderabile creare un tipo di macchina atta ad automatizzare il processo di *decision making* etico? Anche alla luce del fatto che i benefici potenziali delineati sopra sono per lo più promesse incerte, i probabili pericoli dovrebbero farci desistere dal percorrere questa strada? Analizzare tutti i possibili rischi va aldilà dello scopo di questo paragrafo, che si limiterà a discutere quelli più salienti[34].

Innanzitutto la capacità di ragionamento morale non garantisce un processo decisionale eticamente corretto. La macchina può prendere decisioni singole in modo scorretto: ad esempio, se le venissero forniti dei dati sbagliati, delle premesse fuorvianti o delle informazioni parziali, le conclusioni a cui giungerebbe potrebbero essere immorali. Ma si dà anche il caso in cui la macchina può prendere decisioni sistematicamente sbagliate: ad esempio se venisse addestrata male, ovvero fornendole dilemmi con risposte inadeguate da cui inferire principi morali. Per il corretto comportamento della macchina è quindi essenziale la presenza di buoni comportamenti da mimare: sia intesi come l'insieme di dilemmi e risposte esatte che le vengono fornite come *training data*, sia, se la macchina ne è in grado, quelli appresi attraverso l'interazione costante con l'ambiente esterno, in particolare con altre persone o robot. Questa seconda opzione è problematica, in quanto l'influenza negativa di individui malevoli sul comportamento dei sistemi di IA è già stata osservata. Un caso particolarmente significativo riguarda il *chatbot* di *Twitter* di nome Tay, lanciato da *Microsoft* nel 2016. Il sistema di IA è stato progettato per imitare i modelli linguistici di una ragazza americana di 19 anni e per imparare dall'interazione con gli utenti

*reinforcement learning algorithm that masters chess, shogi, and Go through self-play*, in "Science", 2018, 362(6419), pp. 1140-1144.

34 Per una parziale analisi dei rischi si veda M. Brundage, *Limitations and risks of machine ethics*, in "Journal of Experimental & Theoretical Artificial Intelligence", 2014, 26(3), pp. 355-372; S. Cave, R. Nyrup, K. Vold, A. Weller, *Motivations and risks of machine ethics*, in "proceeding of IEEE", 2018, pp. 1-13.

umani. Ma, a sole sedici ore dal suo lancio, il progetto è stato chiuso definitivamente perché alcuni utenti umani avevano insegnato a Tay, attraverso semplici commenti, a comportarsi in maniera razzista: a scrivere commenti negazionisti verso l'olocausto e di supporto ad Adolf Hitler[35]. *Microsoft* non aveva fornito al *chatbot* una comprensione degli atteggiamenti inappropriati, e perciò il sistema ha imitato qualsiasi tipo di comportamento, anche quello deliberatamente offensivo di alcuni utenti. Questo caso è interessante perché il danno arrecato non deriva da un malfunzionamento o da qualche altro *bug* isolato del sistema, ma dal corretto funzionamento in casi non previsti. *Microsoft* non ha dato nessun limite alla capacità di Tay di processare il linguaggio naturale, perché non ha pensato alla possibilità di un attacco da parte di un gruppo di utenti malevoli ben organizzati.

Sebbene il caso possa sembrare una svista facilmente evitabile, in realtà mostra chiaramente il pericolo di utilizzare sistemi di IA capaci di apprendere automaticamente in ambienti aperti, ovvero in ambienti dove è possibile l'interazione con altri esseri umani e con i loro comportamenti complessi e difficilmente prevedibili[36]. Dovremmo assicurarci che una macchina capace di imparare ragionamenti etici abbia buoni maestri ed esempi da seguire – il che non può essere garantito con assoluta certezza – o strumenti abbastanza raffinati per comprendere quali comportamenti vadano rifiutati e quali accettati come buoni esempi.

Un altro problema è che le macchine possono essere facilmente ingannate, come un *team* del MIT ha recentemente dimostrato. Un'IA di *Google* utilizzata in contesto pubblicitario per classificare le immagini, può infatti essere confusa utilizzando delle 'immagini contradditorie'; immagini di soggetti che qualsiasi occhio umano saprebbe riconoscere, ma che anche l'IA più potente non riesce. Sono riusciti a far riconoscere un'immagine 2D di un gatto, per una di una ciotola di guacamole con una certezza del 99%; e in maniera più significativa hanno creato una tartaruga tridimensionale che l'IA riconosce, a ogni angolazione e illuminazione, come un fucile con una certezza del 90%. Un essere umano non nota alcuna differenza, ma è sufficiente aggiungere del rumore all'immagine o alla superficie di un oggetto per ingannare completamente un'IA[37].

Esempi come le immagini contraddittorie e gli oggetti contradditori mostrano come i sistemi artificiali possono avere *performance* migliori di quelle umane in svariati contesti, ma non riuscire dove le aspettative che abbiamo per gli esseri umani sono molto alte. Ciò potrebbe ingannarci e portarci a non prevedere gli errori dell'IA. Quando le macchine falliscono, lo fanno in modi diversi da quelli

35 F. Santoni de Sio, *Ethics and self-driving cars: A white paper on responsible innovation in automated driving systems*, The Dutch Automated Vehicle Initiative (DAVI), 2016, online, pp. 1-33.

36 In un secondo tentativo sono stati forniti a Tay dei filtri per far sì che non apprendesse dai commenti offensivi. Ciononostante il programma è stato nuovamente chiuso perché gli utenti sono riusciti a farle scrivere dei commenti inerenti all'utilizzo di droghe. Il problema quindi non è così facilmente risolvibile come può apparire in prima battuta.

37 A. Athalye, I. Sutskever, *Synthesizing robust adversarial examples*, in "Proceedings of the 35th International Conference on Machine Learning", 2017, pp. 1-19.

dell'essere umano su cui calibriamo le nostre aspettative. E in un campo come la morale, dove il 'senso comune' gioca un ruolo essenziale, non è detto che le macchine siano sempre buoni sostituti.

In secondo luogo, c'è il rischio di silenziare le prospettive morali diverse da quella prescelta. Dal momento che potremmo delegare i ruoli decisionali ai sistemi di IA – per le automobili a guida autonoma o i robot che si prendono cura di anziani e disabili – le considerazioni su quale etica insegnare loro sono inevitabili. Una difficoltà della *machine ethics* è generata proprio dal problema di non poter far riferimento a un'unica teoria di che cosa sia la morale e di quale sia la teoria normativa migliore, e secondo alcuni questa mancanza di accordo potrebbe scoraggiare gli specialisti a intraprendere ricerche nel *machine ethics*[38]. Ad esempio un sondaggio del 2009 ha interrogato 931 filosofi professionisti su quale teoria normativa in etica fosse quella corretta, e i risultati sono stati: *deontologia* 25,9%, *consequenzialismo* 23,6%, *etica della virtù* 18,2%[39].

Nessuna teoria etica sembra godere del sostegno maggioritario, perciò fare totale affidamento su una di esse appare quantomeno problematico. Inoltre, si potrebbe sostenere che una pluralità di opinioni, *framework* e teorie normative sia più desiderabile di una completa polarizzazione verso una singola prospettiva, soprattutto se non si è sicuri che si tratti di quella giusta. È stata anche suggerita la possibilità di distribuire le teorie su agenti artificiali diversi, da fare poi confrontare per trarre un giudizio ponderato sull'aggregazione delle singole prospettive degli agenti[40]. Viceversa la *machine ethics* potrebbe portare a qualcosa di simile a un 'imperialismo etico', ovvero a una universalizzazione degli interessi e dei valori di un singolo gruppo a discapito di quelli di altri. Il gruppo può essere un'azienda, un *team* di programmatori, uno stato illiberale o qualunque altra organizzazione che, in maniera più o meno consapevole, polarizza la sfera della morale a propria scelta.

Infine, il rischio più grande e importante è quello di minare la responsabilità e l'*agency* morale delle persone che cooperano con i sistemi di IA in grado di compiere ragionamenti morali. Non solo la nostra *capacità* di esprimere giudizi morali potrebbe venir erosa nel tempo, ma anche, e soprattutto, la nostra *volon-*

38 Ad esempio J. H. Moor, *The nature, importance, and difficulty of machine ethics*, in "IEEE Intelligent Systems", 2006, 21(4), pp. 18-21, p. 21. Anche M. Klincewicz, L. E. Frank, *Making metaethics work for AI: realism and anti-realism*, in *Envisioning robots in society – power, politics, and public space*, a cura di M. Coeckelbergh, M. Loh, J. Funk, M. Seibt, J. Nørskov, Amsterdam, IOS Press, 2018, pp. 311-318, hanno sostenuto che lo scetticismo può portare all'antirealismo in metaetica, e quindi all'erosione delle motivazioni degli specialisti nel costruire IA etiche. Tuttavia, secondo gli autori, l'antirealismo avrebbe i suoi vantaggi: se non scoraggia gli specialisti in principio, porterebbe convincerli a costruire sistemi che non impongono scelte etiche, ma si limitano a dare consigli, evitando così il problema di un 'imperialismo etico'.

39 D. Bourget e D. J. Chalmers, *What do philosophers believe?*, in "Philosophical Studies", 2014, 170, pp. 465-500.

40 Alcuni sistemi saranno consequenzialisti, mentre altri adotteranno la deontologia o l'etica delle virtù, formando così una rete multi-agente capace di decisioni collettive; J. D. Greene, F. Rossi, J. Tasioulas, B. K. Venable, B. Williams, *Embedding ethical principles in collective decision support systems*, cit.

*tà* di farlo, ovvero la volontà di assumersi responsabilità di compiere decisioni morali e sopportarne le conseguenze. Ogni macchina in grado di automatizzare un'azione che prima era compiuta dall'uomo ha – in linea teorica – il potenziale di erodere l'abilità umana nel compierla. Ormai da molto tempo le calcolatrici hanno reso le persone inabili a compiere semplici calcoli algebrici; da qualche decennio i navigatori dotati di GPS hanno reso possibile compiere interi percorsi senza né conoscere la strada né richiedere alla persona di memorizzarla; domani i traduttori simultanei potrebbero erodere la nostra volontà di imparare più di una lingua. Allo stesso modo, un sistema di IA potrebbe essere capace di aiutare un essere umano a prendere una decisione etica; e nei luoghi dove le macchine prenderanno le decisioni, è possibile che le persone non svilupperanno l'abilità o la volontà di farlo. Ad esempio, nel caso dei robot sanitari, il personale umano che lavora con quest'ultimi potrebbe non sviluppare la sensibilità necessaria per decidere quando intervenire paternalisticamente per convincere un paziente a prendere la sua medicina.

Si potrebbe pensare che, fintantoché i robot sono in funzione, così come le calcolatrici e i GPS, l'inettitudine umana potrebbe non essere un grande problema. Ma questa idea è pericolosa. Dato che i sistemi automatici tendono a fallire – sia in situazioni semplici, ma soprattutto in quelle insolite e complesse – l'intervento umano è necessario per colmare i *deficit* della macchina. Ma c'è il serio rischio che, a seguito dell'automatizzazione, la persona sia mal preparata, poco motivata, o peggio ancora, completamente deresponsabilizzata, per assumersi il compito di riallineare il comportamento della macchina, o di sostituirne temporaneamente la funzione.

Ma c'è un'ulteriore problema: se i sistemi di *machine ethics* più sofisticati venissero utilizzati per prendere decisioni in situazioni complesse, e gli esseri umani riducessero enormemente il loro esercizio delle facoltà morali, quest'ultimi non sarebbero nemmeno in grado di comprendere quando la macchina sta fallendo nella valutazione. Un assistente che non ha mai approfondito la nozione di autonomia del paziente, o che l'ha appresa ma mai applicata nella pratica, saprebbe riconoscere che una macchina sta violando l'autonomia del paziente? Avrebbe il senso di responsabilità necessario per intervenire?

All'interno delle società, l'esercitazione costante delle abilità morali da parte degli individui è una condizione necessaria per il mantenimento dei legami personali, di fiducia e per il rispetto reciproco. Il ragionamento morale non è, almeno in prima battuta, legato allo studio di nozioni o teorie slegate dalla pratica. È piuttosto una facoltà esercitata sin da bambini – inizialmente con l'educazione genitoriale e via via resa più profonda e complessa dall'interazione sociale e dalla riflessione – e portata avanti fino all'età adulta, nei contesti professionali e negli ambienti lavorativi. Automatizzare per intero il processo di *decision-making* morale avrebbe effetti non del tutto prevedibili sulla nostra comprensione degli altri e di noi stessi, sui rapporti interpersonali e, in generale, sulla morale per come la conosciamo oggi.

## 5. Considerazioni conclusive

In conclusione, progettare macchine con capacità di ragionamento morale potrebbe potenzialmente migliorare l'allineamento etico di uomini e macchine. Tuttavia, questa ragione, sebbene plausibile, non fornisce di per sé sufficienti motivi per perseguire gli obiettivi della *machine ethics*, a meno che i rischi evidenziati non vengano risolti: sia sviluppando soluzioni che possano mitigare i pericoli, come il continuo esercizio delle nostre facoltà morali; sia formulando regolamenti che limitino, almeno inizialmente, l'uso della *machine ethics* ai contesti a basso rischio e che regolino l'introduzione dei sistemi in modo che sia graduale e strettamente controllata.

È comunque probabile che nei prossimi decenni si sviluppi maggiormente il dibattito su questi temi. Già oggi può essere facilmente osservato che il campo della *machine metaethics*, che si occupa dei pericoli e dei fondamenti teoretici della *machine ethics*, attrae maggior interesse rispetto al suo soggetto d'indagine (la *machine ethics*). Ciò può essere considerato positivo: per evitare danni futuri o riflettere semplicemente sulla possibilità che l'impresa dia buoni risultati, è fin da subito necessario discutere le questioni fondamentali – metaetiche, appunto. Tuttavia, bisogna fare attenzione che non si verifichi, come invece a volte si percepisce, uno scollamento tra ciò che è il livello della ricerca attuale, e le speculazioni su un futuro a breve termine. Sviluppare un'ampia letteratura di *machine metaethics*, limitandosi a brevissimi accenni ai prototipi oggi disponibili – e in alcuni casi evitando completamente di farvi riferimento – potrebbe causare una distorsione nella percezione delle nostre capacità attuali e creare un allarmismo difficilmente giustificabile. Una riflessione che si sviluppa di pari passo con i guadagni tecnologici può essere più utile anche per la ricerca empirica stessa e per il suo sviluppo nei diversi domini.

Fino al 1800 quella di tessere era una delle abilità padroneggiate unicamente dagli esseri umani. Una volta introdotti i telai meccanici, però, l'attività è stata piano piano delegata alle macchine, con cui le persone hanno iniziato a cooperare. Oggi con i telai industriali pochi sanno realmente tessere, e l'abilità è ormai *in mano* alle macchine. Non è certo né scontato, ma in futuro molte delle scelte etiche che oggi caratterizzano il nostro modo di vivere potrebbero essere affidate alle macchine. Siamo disposti, nell'arco di qualche decennio, a delegare questa nostra abilità a sistemi artificiali? O una nuova forma di luddismo nascerà in campo etico?