Fabio Fossa

Etica dell'Intelligenza Artificiale: una *futuromachía*?

Intelligenza artificiale, etica e futuro

In un'epoca tanto definita dalle conquiste della scienza tecnologica, la sfida che il futuro riserva al pensiero morale non può che presentarsi anche come riflessione critica sulla relazione che intercorre tra tecnologie digitali e vita umana. In particolare, l'etica dell'Intelligenza Artificiale (IA) ha conosciuto negli ultimi decenni una notevole affermazione,[1] tanto da essere divenuta ingrediente fondamentale di ogni discussione circa lo sviluppo futuro della tecnologia e la sua integrazione alla società.[2]

Per quanto l'IA stia già riconfigurando vari settori del mondo contemporaneo, è soprattutto la dimensione futura del suo sviluppo ed utilizzo a imporre l'urgenza di un suo vaglio in chiave morale. L'etica dell'IA – il cui compito, in sintesi, consiste sia nella critica dei suoi impatti morali sia nell'elaborazione di princìpi, procedure e raccomandazioni per la sua progettazione e sviluppo etico – è legata alla dimensione futura in almeno tre nodi.

[1] La crescita di interesse nell'etica dell'IA è registrata da D.J. Gunkel, *Robot rights*, Cambridge 2018, p. 187, il quale nota come sia ormai impossibile offrirne un quadro accurato e esaustivo, sebbene l'impresa potesse ancora essere provata una decina di anni fa.

[2] Che la riflessione etica rappresenti un elemento irrinunciabile del più generale sforzo votato alla regolazione dell'IA è comprovato dall'attenzione ad essa riservata nei numerosi documenti emanati a livello nazionale e internazionale. Si vedano, ad esempio, House of Lords, *AI in the U.K.: ready, willing and able?*, 2017, https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf; TaskForceIA, *Libro Bianco sull'Intelligenza Artificiale*, 2018, https://ia.italia.it/assets/librobianco.pdf; C. Villani, *For a meaningful Artificial Intelligence: Towards a French and European Strategy*, 2018, https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf; AI HLEG, *Ethics Guidelines for Trustworthy AI*, 2019, https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai; Beijing Academy of Artificial Intelligence, *Beijing AI Principles*, 2019, https://www.baai.ac.cn/blog/beijing-ai-principles.

Innanzitutto, l'etica dell'IA rappresenta una forma di etica *nel* futuro. La discussione critica del difficile rapporto tra IA e valori umani inaugura una linea di indagine che costituirà una delle future aree dell'etica, un ambito di grande rilevanza sia filosofica che sociale in cui il pensiero morale potrà esercitare la propria caratteristica metodologia.

In più, l'etica dell'IA è un'etica *per* il futuro. Lo scopo perseguito è la costruzione di un futuro in cui le potenzialità delle nuove tecnologie siano messe al servizio del benessere comune e di ciò che a vario livello viene considerato degno di promozione. La costruzione di un futuro in cui l'IA sia veicolo di un generale miglioramento dell'esistenza umana costituisce la ragion d'essere di ogni sua discussione morale[3].

Infine, l'etica dell'IA è un'etica *del* futuro nella misura in cui ha per oggetto il futuro della tecnologia. Non solo, e non tanto, perché il processo di diffusione dell'IA è ancora alle sue prime fasi, per cui il futuro rimane necessariamente la dimensione temporale di riferimento; ma anche perché il fulcro della riflessione, anche quando considera le tecnologie di oggi e il loro essere come sono, riguarda pur sempre il dover essere delle tecnologie di domani e del nostro rapporto con loro. L'etica dell'IA è un'etica del futuro sia perché studia tecnologie e situazioni che già si profilano, ma promettono di esplicarsi pienamente solo nel domani; sia perché, nel suo lato normativo, si propone di agire sulle tecnologie del domani muovendo da una critica delle tecnologie dell'oggi.

Pochi dubbi, dunque, possono essere nutriti sul fatto che l'etica dell'IA sia profondamente intrecciata al futuro; ma *a quale immagine* del futuro attenersi è invece motivo di contesa. Diverse concezioni del futuro dell'IA conducono infatti a diverse posizioni circa i problemi da discutere e le iniziative da intraprendere. Si può affermare, al proposito, che sia in corso una *futuromachía*, una battaglia tra futuri che motivano impostazioni di pensiero contrastanti.

Un futuro inedito

L'immagine del futuro dell'IA più nota, anche a motivo della sua componente fantascientifica, è incentrata sulla convinzione per cui il domani abbia in serbo un avvenimento che rivoluzionerà profondamente non solo l'esistenza umana, ma l'intera configurazione del

[3] Cfr. M. Taddeo – L. Floridi, *How AI can be a force for good*, «Science», 361, 6404 (2018), pp. 751-752.

nostro pianeta (se non oltre).[4] Si tratta della cosiddetta Singolarità, una piega nel tempo futuro oltre cui è impossibile gettare lo sguardo, poiché viene a mancare qualunque appiglio analogico.[5]

La Singolarità, in sintesi, marca la prima occorrenza di un sistema artificiale la cui intelligenza sia pari o superiore a quella umana. Che un simile risultato sia custodito nello scrigno degli accadimenti futuri, e non solo nella fantasia di alcuni visionari, è come vedremo oggetto di vivaci discussioni. Ma se si assume una tesi continuista in relazione al problema della riproduzione tecnologica dell'intelligenza – se si sostiene, cioè, che non si diano differenze essenziali tra intelligenza umana e sua simulazione computazionale[6] – l'eventualità che il continuo avanzamento tecnologico colmi il divario tra copia e modello appare più come una meta annunciata che come uno degli infiniti possibili.

La rilevanza della Singolarità risiede nella concatenazione di eventi che si presume possa innescare. Prendendo in mano le sorti della ricerca scientifica – così prosegue la narrazione – l'IA guiderà sé stessa ad uno sviluppo esponenziale il cui prodotto saranno intelligenze sintetiche (macchine ultraintelligenti, superintelligenza, IA forte o generale)[7] tanto superiori alla controparte biologica che noi umani non saremo più in grado né di comprenderne i princìpi né di esercitare qualsivoglia forma di controllo su di esse.[8] Si passerà, così, da uno

[4] Su ciò si veda M. Tegmark, *Vita 3.0. Esseri umani nell'era dell'intelligenza artificiale*, Milano 2018.

[5] Cfr. V. Vinge, *The Coming Technological Singularity: How to Survive in the Post-Human Era*, in *Vision-21. Interdisciplinary Science and Engineering in the Era of Cyberspace*, NASA Scientific and Technical Information Program 1993.

[6] Su ciò mi sia permesso rimandare a F. Fossa, *Artificial moral agents: moral mentors or sensible tools?*, «Ethics and Information Technology», 20 2 (2018), pp. 115-126.

[7] Per "IA generale" s'intende un sistema sviluppato non per l'esecuzione di un compito specifico (si parla in questo caso di intelligenza artificiale debole – *weak AI* – o meglio ristretta – *narrow IA*) ma che esibisca la «capacità di svolgere qualsiasi compito cognitivo almeno tanto bene quanto un essere umano» (M. Tegmark, *Vita 3.0.*, cit., p. 61). La coppia di termini *strong/weak IA* è stata invece introdotta da J.R. Searle, *Minds, brains, and programs*, «The Behavioral and Brain Science», 3 (1980), pp. 417-457, sebbene con un significato leggermente diverso da come è utilizzata in molti dibattiti sulla Singolarità. A teorizzare macchine ultraintelligenti già negli anni Sessanta è I.J. Good, *Speculations concerning the first ultraintelligent machine*, «Advances in Computers», 6 (1966), pp. 31-88. Il termine "superintelligenza" è invece stato introdotto da N. Bostrom, *Superintelligence. Paths, Dangers, Strategies*, Oxford 2014.

[8] N. Bostrom, *Ethical Issues in Advanced Artificial Intelligence*, in I. Smith et al. (cur.), *Cognitive, Emotive and Ethical Aspects of Decision Making in Humans and in Artificial Intelligence Vol. 2*, Tecumseh 2003, pp. 12-17.

scenario di delegazione, dove si utilizza l'IA per svolgere compiti che servono interessi umani, a uno scenario di sostituzione, dove sarà l'IA stessa a definire obiettivi secondo interessi propri – con il rischio di marginalizzare o strumentalizzare l'essere umano, forzato in un battito di ciglia ad abdicare dal trono che si è conquistato in migliaia di anni di evoluzione.

Se questo è, per così dire, il destino dell'IA, l'unica dimensione temporale di cui sembra davvero urgente occuparsi è il *futuro inedito* che si dispiegherà al sorgere della superintelligenza. Un futuro che, così si spera, possiamo forse rendere per noi più ospitale – o, per lo meno, gestibile – prendendo sin da ora alcune precauzioni che ci permettano di mantenere l'IA bendisposta (*friendly*) nei nostri confronti.[9]

Optare per lo scenario del futuro inedito, incentrato sulla Singolarità, ha quindi effetti ben precisi sul profilo odierno dell'etica dell'IA, determinandone oggetti preferenziali di discussione, modalità di indagine, aree di intervento e carattere delle raccomandazioni. Innanzitutto ciò significa pensare l'IA non solo e non tanto come strumento di cui potersi liberamente servire, ma anche come forma *in nuce* di soggettività morale in grado di sviluppare un proprio sistema di interessi e valori (nonché di attività volte a soddisfarli). Un soggetto, in più, con il quale ogni relazione sarà complessa e imprevedibile, data la sua superiorità intellettiva e, dunque, alterità pratica.

Così concepita, l'etica dell'IA si rivolge alle tecnologie odierne solo in senso indiretto. Suo oggetto primario è una tecnologia futura di cui le IA moderne sarebbero predecessori. L'interesse per lo stato dell'arte non nasce da ciò che esso rappresenta per l'umanità contemporanea e del domani più prossimo, ma per ciò a cui prelude. Infine, l'etica dell'IA non è solo un'etica del buon design e dell'allineamento valoriale – il cui nocciolo è l'implementazione non sovrascrivibile del sistema di valori umano – ma anche e in particolare un'etica della convivenza con una specie aliena, uscita dalle nostre stesse mani ma che, come ogni grande opera, afferma la propria indipendenza e muove per strade non sempre parallele a quelle dei loro autori.

[9] Cfr. L. Muehlhauser – N. Bostrom, *Why we need friendly AI*, «Think», 13, 36 (2014), pp. 41-47. Alcuni studiosi contestano questa tesi, vedendo invece (con maggior coerenza) nella Singolarità l'opportunità di lasciare il pianeta in eredità ad esseri migliori di noi. Cfr. E. Dietrich, *After the humans are gone*, «Journal of Experimental and Theoretical Artificial Intelligence», 19, 1 (2007), pp. 55-67.

Un futuro analogo

Il futuro inedito incentrato sulla Singolarità non è l'unico scenario che spinge ad interrogarsi sul significato etico dell'IA. Al contrario, l'accesa controversia relativa al concetto di superintelligenza – alla sua validità logica prima che etica – segnala la presenza di un'alternativa che, invece di focalizzare l'obiettivo su un futuro eterogeneo ma indeterminato, si attiene ad una immagine più concreta. Si può scegliere, cioè, di ancorare la riflessione morale ad un futuro *analogo* al presente, in cui nuovi prodotti renderanno possibili nuovi corsi di azione e causeranno nuovi effetti da criticare e valutare pur rimanendo riconoscibile l'assetto generale della realtà[10].

Prescindere dalla contemplazione della Singolarità implica, allo stesso tempo, prescindere dalla superintelligenza sia come oggetto significativo di analisi sia come fenomeno alla cui luce leggere le tecnologie odierne. Quanto rimane è un'interpretazione, forse meno entusiasmante ma anche più allineata allo stato dell'arte,[11] dell'IA in termini strumentali. In breve, attenendosi ai caratteri dell'IA non solo così come esperibili ai nostri giorni, ma anche tenendo in considerazione i contesti sociali nel cui alveo l'IA si manifesta[12], le tecnologie digitali in genere sono interpretate secondo due concetti fondamentali: il concetto di prodotto[13], per cui esse sarebbero la risultante di un fare umano intelligente[14] e costruite per svolgere funzioni i cui scopi sono dati; e il concetto di strumento[15], per cui l'IA sarebbe innanzi-

[10] Si veda, a titolo esemplificativo, L. Floridi, *What the Near Future of Artificial Intelligence Could Be*, «Philosophy & Technology», 32, 1 (2019), pp. 1-15.

[11] Cfr., ad esempio, J. Kaplan, *Artificial Intelligence: What Everyone Needs to Know*, New York 2016.

[12] Cfr. D.G. Johnson, *Computer Systems*, in M. Anderson, S.L. Anderson (cur.), *Machine Ethics*, Cambridge 2011, pp. 168-183.

[13] Si veda ad esempio A. Bertolini, *Robots as products: The case for a realistic analysis of robotic applications and liability rules*, «Law, Innovation and Technology», 5, 2 (2013), pp. 214-247.

[14] J.C. Pitt, *It's Not About Technology*, «Knowledge, Technology & Policy», 23 (2010), pp. 445-454.

[15] Cfr. H. Jonas, *A Critique of Cybernetics*, «Social Research», 20, 2 (1953), pp. 172-192; J.J. Bryson – P.P. Kime, *Just an artifact: Why machines are perceived as moral agents*, in *IJCAI International Joint Conference on Artificial Intelligence*, 2011, pp. 1641-1646.

tutto un mezzo a cui gli essere umani delegano lo svolgimento di determinati compiti pur rimanendo in controllo, per quanto lo si possa essere, della generale progettualità della propria esistenza.

Mettere a fuoco l'IA come prodotto e strumento implica certamente una tendenza a proiettare la riflessione nella dimensione futura, essendo ogni prodotto migliorabile o adeguabile a ulteriori esigenze etiche. Tuttavia, il futuro qui in questione mantiene una somiglianza significativa con i caratteri del presente: proprio tale continuità motiva l'impegno volto all'affinamento progressivo non solo dell'uso umano delle tecnologie, ma anche del profilo morale *delle tecnologie stesse*.

Per quanto, infatti, si critichi l'approccio strumentale sulla base della constatazione che il significato morale degli artefatti non sia interamente riducibile al loro uso – come una dottrina del mero strumento potrebbe indicare[16] – è ormai chiaro che sia necessario concepire gli artefatti tecnologici come mediatori di valori etici: ovvero come enti che, svolgendo la funzione a loro preposta, contribuiscono all'affermazione di valori (o disvalori) di carattere etico e dunque presentano un profilo morale indipendente dal modo in cui vengono usati.[17] Invece di ignorare tale aspetto, come a volte si lascia intendere in sede polemica, l'approccio strumentale predispone la concettualità necessaria a pensare il fenomeno dell'implementazione di valori negli artefatti (ivi compresa l'IA) senza da una parte procedere a imprudenti estensioni terminologiche – si pensi, ad esempio, all'uso del termine *agente morale artificiale* – né tantomeno ridurre la questione etica alla sola componente dell'uso.

Incentrare l'etica dell'IA su un futuro analogo al tempo presente, e su tecnologie pensate in modo continuo rispetto a quanto osservabile oggi, determina un'immagine della disciplina assai diversa da quanto visto in precedenza. In questo caso, l'oggetto principale dell'analisi non

[16] Cfr. A.H. Kiran – P.-P. Verbeek, *Trusting Our Selves to Technology*, «Knowledge, Technology & Policy», 23, 3-4 (2010), pp. 409-427; D.J. Gunkel, *Robot rights*, cit., pp. 53-55.

[17] Cfr. L. Winner, *Do Artifacts Have Politics?*, «Daedalus», 109, 1 (1980), pp. 121-136; H. Nissenbaum, *How Computer Systems Embody Values*, «Computer», 34, 3 (2001), pp. 119-120; B. Friedman – P.H. Kahn, *Human Values, Ethics, and Design*, in A. Sears – J.A. Jacko (cur.), *Human Computer Interaction Handbook: Fundamentals, Evolving Technologies, and Emerging Applications*, New York-London 2008, pp. 1241-1266; P. Kroes – P.-P. Verbeek (cur.), *The Moral Status of Technical Artefacts*, Dordrecht 2014.

è l'IA generale, ma l'IA ristretta, cioè applicazioni pensate per risolvere compiti specifici, sebbene variamente determinati[18]. In più, per quanto lo sguardo di una simile interrogazione sia comunque rivolto alle tecnologie future, il suo esercizio si applica senz'altro allo stato attuale dell'avanzamento tecnologico. Non prevedendo alcuna vera e propria frattura tra i sistemi di oggi e di domani, offrire una critica o partecipare al design delle tecnologie presenti significa già, infatti, influenzare gli analoghi sistemi futuri. L'etica dell'IA è dunque concepita come un'etica del buon uso e del buon design dello strumento tecnologico – cioè dell'uso e del design riflessivo, competente, responsabile e informato dai valori morali pertinenti. L'IA, qui, non prelude all'emergenza di un nuovo soggetto morale, fonte di una inedita visione del mondo, ma richiede discussione critica e intervento proattivo in quanto nuovo mezzo di affermazione e negazione di valori etici.

Futuromachía

L'interesse per il futuro del genere umano è un tratto distintivo dell'etica dell'IA. Sia che si focalizzi l'attenzione sul futuro inedito dell'umanità al cospetto della superintelligenza, sia che si concentri l'analisi sul futuro analogo che ci attende, la significatività morale delle tecnologie odierne è intesa non solo in relazione agli effetti che già esercitano, ma soprattutto in vista di come il loro impiego determinerà gli anni, i decenni, i secoli a venire – un futuro che, su ciò non si dubita, conoscerà una diffusione capillare delle tecnologie digitali e una loro intima integrazione all'esistenza umana. Ma se il ragionamento, che ha come base l'idea sempre più centrale di responsabilità

[18] Distinguere tra specificità e determinazione di un compito automatizzato può essere utile quando si discute di IA generale e ristretta. Un compito è specifico quando l'obiettivo prefissato e le aspettative connesse siano esprimibili con una certa precisione, mentre è determinato quando la sua esecuzione è strutturata. Un sistema di IA che fornisca ad un robot la capacità di muoversi in un ambiente sconosciuto può essere considerato come un sistema che tende a uno scopo specifico (in quanto si possono elaborare aspettative circa la sua efficienza che ne guidino la valutazione) ma indeterminato (dal momento che, per essere eseguito, il compito richiede flessibilità e un ventaglio di abilità generico). Un robot in grado di navigare con successo un ambiente noto svolge un compito specifico e altamente determinato; un robot in grado di navigare con successo qualsiasi ambiente svolge anch'esso un compito specifico, ma meno determinato rispetto al precedente. In nessuno dei due casi, comunque, avrebbe senso parlare di IA generale: si tratta di diversi gradi di IA ristretta.

per il futuro, è ampiamente condiviso, gli scenari tratteggiati nei paragrafi precedenti competono in un dibattito alimentato da una profonda dissonanza di carattere etico – relativa, cioè, ad un'ideale di ricerca responsabile. Infuria una *futuromachía*, una contesa sull'immagine del futuro che debba presiedere non solo agli sforzi di analisi, ma anche all'impegno positivo di elaborare raccomandazioni e cornici regolative circa l'uso, il design e lo statuto sociale delle nuove tecnologie.

Un rapido confronto alle reciproche accuse, per quanto solo esemplificativo, può rendere l'idea delle energie coinvolte nel dibattito. Da una parte, chi opta per l'immagine di un futuro inedito giustifica la propria scelta a partire dalla possibilità che la Singolarità si verifichi effettivamente. Se anche la probabilità di un simile evento fosse irrisoria, argomentano ad esempio Bostrom[19] e Chalmers[20], le sue conseguenze sarebbero di tale portata che sarebbe avventato, se non irresponsabile, trascurare un'indagine etica che si occupi sin da ora non solo della futura sopravvivenza e felicità del genere umano, ma anche dello statuto morale delle nuove tecnologie e della convivenza delle due specie.

Tuttavia, è proprio partendo dalla medesima premessa (la bassa probabilità che il futuro abbia in serbo superintelligenze artificiali) e mobilitando la medesima logica (relativa alla responsabilità etica della ricerca) che altri autori polemizzano contro l'idea di un futuro inedito. Così, ad esempio, conduce il suo attacco Floridi[21], da una parte insistendo sulla risibile implausibilità delle ipotesi relative alla Singolarità, dall'altra sottolineando l'irresponsabilità insista nel distrarre l'opinione pubblica e l'indagine scientifica dai veri mali e dai reali pericoli che attanagliano il mondo contemporaneo, inquinando la discussione circa il ruolo che l'IA gioca e può giocare in relazione ad essi.

Davanti a un simile crocevia si è tentati di sospendere il giudizio e archiviare la diatriba nella nutrita classe dei diverbi concettuali relativi all'IA. Si tratterebbe, però, di un errore. La *futuromachía* non è riducibile ad uno scontro di opinioni puramente teoretico, ma ha immediate ripercussioni sul piano pratico. Essa concerne, infatti, la

[19] N. Bostrom, *Ethical Issues in Advanced Artificial Intelligence*, cit.

[20] D.J. Chalmers, *The singularity: A philosophical analysis*, «Journal of Consciousness Studies», 17, 9-10 (2010), pp. 7-65.

[21] L. Floridi, *Singularitarians, AItheists, and Why the Problem with Artificial Intelligence is H.A.L. (Humanity At Large), not HAL*, «APA Newsletter Philosophy and Computing», 14, 2 (2015), pp. 8-11.

possibilità di porre le giuste domande relative all'IA; e dalla qualità delle domande deriva anche la bontà delle risposte che ad esse si danno quando si stilano linee guida e raccomandazioni per il futuro sviluppo etico dell'IA. Qui le due prospettive non possono più convivere: la *futuromachía* deve risolversi a favore di uno dei contendenti.

A riprova di ciò rimane una nota apposta dai membri dello High Level Expert Group on Artificial Intelligence al capitolo 5.5 del documento *Draft Ethics Guidelines for Trustworthy Technology*[22]. Il capitolo, che riassume concisamente le preoccupazioni etiche relative ad un futuro inedito, è introdotto da poche righe in cui il gruppo di esperti nominato dalla Commissione Europea informa il lettore di come le tesi esposte più sotto siano state accompagnate, in sede di dibattito, da accese discussioni prive di un risultato condiviso. Nella versione finale del documento[23], rilasciata quattro mesi più tardi, rimane solo un accenno estremamente cauto e sospettoso ad un futuro inedito come orizzonte temporale dell'etica dell'IA, mentre l'intero documento sposa una prospettiva più strumentale. In definitiva, la contesa intorno ai futuri dell'IA non è un elemento marginale del dibattito etico, ma anzi determina prese di posizione che ne definiscono l'assetto generale anche a livello istituzionale e normativo.

Epilogo

Alla luce di quanto detto, sembra che la *futuromachía* obblighi a misurarsi nell'agone delle argomentazioni. Senza la presunzione di risolvere in così breve spazio una questione tanto impegnativa, si può però notare come le concezioni del futuro in competizione non esibiscano la medesima rilevanza da un punto di vista etico. Ciò avviene non tanto per il caso specifico della superintelligenza, sulla cui possibilità epistemologica (prima che tecnologica) è forse lecito dubitare[24], quanto per l'indeterminatezza insita nel costruire un pensiero morale sulle fondamenta di un evento rivoluzionario. Se, come suggerisce

[22] AI HLEG, *Draft Ethics Guidelines for Trustworthy AI*, 2018, https://ec.europa.eu/digital-single-market/en/news/draft-ethics-guidelines-trustworthy-ai.

[23] AI HLEG, *Ethics Guidelines for Trustworthy AI*, cit.

[24] Ho espresso i miei dubbi circa l'interpretazione dell'imitazione tecnologia (in questo caso, l'intelligenza artificiale) come duplicazione del modello (l'intelligenza umana) in F. Fossa, *Fare e funzionare. Sull'analogia di robot e organismo*, «InCircolo», 6 (2018), pp. 73-88.

MacIntyre, è vano tentare di prevedere la configurazione della scienza a seguito di una scoperta rivoluzionaria, in quanto se ne ridurrebbe l'alterità a quanto già noto (ma paradossalmente superato dalla rivoluzione che si vuole pensare)[25], è forse più prudente optare per un'immagine del futuro basata sulle istanze etiche e sociali già emerse in relazione alle tecnologie IA, così da impostare una prassi riflessiva volta al loro progressivo allineamento ai valori umani.

È lungo questa linea di pensiero che il rapporto tra etica dell'IA e temporalità si struttura nel modo più convincente. Concependo un futuro analogo, l'etica dell'IA rimane un'etica *nel* futuro e *per* il futuro, ma si fa un'etica *del* futuro solo nella misura in cui accetta di essere prima di tutto un'etica *del presente* – relativa a stati di cose per quanto possibile osservabili e verificabili, su cui sia possibile esercitare una critica razionale e suggerire proposte tangibili. Integrare la dimensione futura nella riflessione morale è un processo delicato: ipotecare la connessione con il presente in virtù di controversi sviluppi a venire ne indebolisce ulteriormente la struttura epistemologica, rischiando di minarne irreparabilmente la tenuta. È invece in un futuro capace di mostrare i propri tratti di analogia con l'esperienza vissuta nel tempo presente che l'etica dell'IA può trovare il proprio concreto ambito d'azione.

[25] Cfr. A. MacIntyre, *After Virtue. A Study in Moral Theory*, London-New York 2007, pp. 109-110.