mosaico di pace
Mensile
Data 03-2018
Pagina 26/28
Foglio 1 / 3

L'epoca della machina sapiens

Gli interrogativi della teologia di fronte alle scoperte della ricerca digitale. Quale cooperazione tra intelligenza artificiale e intelligenza umana e quali sfide attende la morale di domani?

Paolo Benanti
Francescano del Terzo Ordine Regolare, docente presso la Pontificia Università Gregoriana e l'Istituto teologico di Assisi

NUOVI ARTEFATTI

L'avvento della ricerca digitale, dove tutto viene trasformato in dati numerici, porta alla capacità di studiare il mondo secondo nuovi paradigmi gnoseologici: quello che conta è solo la correlazione tra due quantità di dati e non più una teoria coerente che spieghi tale correlazione. Oggi la correlazione viene usata per predire con sufficiente accuratezza, pur non avendo alcuna teoria scientifica che lo supporti, il rischio di impatto di asteroidi anche sconosciuti in vari luoghi della Terra, i siti istituzionali oggetto di attacchi terroristici, il voto dei singoli cittadini alle elezioni presidenziali USA, l'andamento del mercato azionario nel breve termine.

L'evoluzione tecnologica dell'informazione e del mondo compreso come una serie di dati si concretizza nelle *Intelligenze Artificiali* e nei *robot*: siamo in grado di costruire macchine che possono prendere decisioni autonome e coesistere con l'uomo. Si pensi alle macchine a guida autonoma che Uber, il noto servizio di trasporto automobilistico privato, già utilizza in alcune città come Pittsburgh, o a sistemi di radio chirurgia come il *Cyberknife* o i *robot* destinati al lavoro affianco all'uomo nei processi produttivi in fabbrica. Le IA, queste nuove tecnologie, sono pervasive. Stanno insinuandosi in ogni ambito della nostra esistenza. Tanto nei sistemi di produzione, *incarnandosi* in *robot*, quanto nei sistemi di gestione sostituendo i *server* agli analisti. Ma anche nella vita quotidiana i sistemi di IA sono sempre più pervasivi. Gli *smartphone* di ultima generazione sono tutti venduti con un assistente dotato di intelligenza artificiale. *Cortana*, *Siri* o *Google Hello* – per citare solo i principali –, che trasforma il telefono da un *hub* di servizi e applicazioni

a un vero e proprio *partner* che interagisce in maniera cognitiva con l'utente. Sono in fase di sviluppo sistemi di intelligenza artificiale, i *bot*, che saranno disponibili come *partner* virtuali da interrogare via voce o in *chat* che sono in grado di fornire servizi e prestazioni che prima erano esclusivi di particolari professioni: avvocati, medici e psicologi sono sempre più efficientemente sostituibili da *bot* dotati intelligenza artificiale.

La società conosce oggi una nuova frontiera: le interazioni e la coesistenza tra uomini e intelligenze artificiali. Prima di addentrarci ulteriormente nel significato di questa trasformazione, dobbiamo considerare un implicito culturale che rischia di sviare la nostra comprensione del tema. Nello sviluppo delle intelligenze artificiali la divulgazione dei successi ottenuti da queste macchine è sempre stata presentata secondo un modello competitivo rispetto all'uomo. Per fare un esempio, IBM ha presentato *Deep Blue* come l'intelligenza artificiale che nel 1996 riuscì a sconfiggere a scacchi il campione del mondo in carica, Garry Kasparov, e sempre IBM nel 2011 ha realizzato *Watson*, che ha sconfitto i campioni di un noto gioco televisivo sulla cultura generale *Jeopardy!*. Queste comparse mediatiche delle IA potrebbero farci pensare che questi sono sistemi che competono con l'uomo e che tra *homo sapiens* e questa nuova *macchina sapiens* si sia instaurata una rivalità di natura evolutiva, che vedrà un solo vincitore e condannerà lo sconfitto a una inesorabile estinzione. In realtà, queste macchine non sono mai state costruite per competere con l'uomo, ma per realizzare una nuova simbiosi tra l'uomo e i suoi artefatti: (*homo+machina*) *sapiens* (cfr. J. E. Kelly – S. Hamm, *Macchine intelligenti. Watson e l'era del cognitive computing*, Egea, Milano 2016, 5-42).

Non sono le IA la minaccia di estinzione dell'uomo, anche se la tecnologia può essere pericolosa per la nostra sopravvivenza come specie: l'uomo ha già rischiato di estinguersi perché battuto da una macchina *molto stupida* come la bomba atomica. Tuttavia esistono sfide estremamente delicate nella società contemporanea in cui la variabile più importante non è l'intelligenza ma il poco tempo a disposizione per decidere e le macchine cognitive trovano qui grande interesse applicativo.

> La capacità dei robot di mutare il loro comportamento in base alle condizioni in cui operano, per analogia con l'essere umano, viene definita autonomia

QUALI PROBLEMI?

Si apre a questo livello tutta una serie di problematiche etiche su come validare la cognizione della macchina alla luce proprio della velocità della risposta che si cerca di implementare e ottenere. Tuttavia, il pericolo maggiore non viene dalle IA in se stesse, ma dal non conoscere queste tecnologie e dal lasciare decidere sul loro impiego a una classe dirigente assolutamente non preparata a gestire il tema. Se l'orizzonte del prossimo futuro – in realtà già del nostro presente – è quello di una cooperazione tra intelligenza umana e intelligenza artificiale e tra agenti umani e agenti robotici autonomi diviene urgente cercare di capire in che maniera questa realtà mista, composta da fattori autonomi umani e fattori autonomi robotici, possa coesistere.

PRIMUM NON NOCERE

Il primo e più urgente punto che le intelligenze artificiali pongono nell'agenda dell'innovazione del lavoro è quello di adattare le nostre strutture sociali a questa nuova e inedita società fatta di agenti autonomi misti.

Una primissima sfida è di natura filosofica e antropologica. Queste frontiere delle innovazioni, la realizzazione di queste macchine "*sapiens*", per utilizzare un termine molto evocativo, ci interroga in profondità sulla **specificità dell'*homo sapiens*** e in particolare su quale sia la specifica componente e qualità umana del lavoro rispetto a quella macchinica: le rivoluzioni industriali hanno dimostrato che non è l'energia, non è la velocità e, ora, che anche la cognizione e l'adattabilità alla situazione non sono specifiche solamente umane.

QUALE MORALE?

Un secondo, e altrettanto urgente tema, è quello di definire come e in che maniera si può garantire la coesistenza tra uomo e IA, tra uomo e *robot*. Per rispondere a questa domanda procederemo nel seguente modo. In primo luogo, cercheremo di formulare una direttiva fondamentale che deve essere garantita dalle IA e dai *robot* e poi cercheremo di definire cosa questi sistemi cognitivi autonomi *devono imparare* per poter convivere e lavorare cooperativamente con l'uomo.

La prima e fondamentale direttiva da implementare può essere racchiusa nell'adagio latino *primum non nocere*. La realizzazione di tecnologie controllate da sistemi di IA porta con sé una serie di problemi legati alla gestione dell'autonomia decisionale di cui questi apparati godono. La capacità dei *robot* di mutare il loro comportamento in base alle condizioni in cui operano, per analogia con l'essere umano, viene definita *autonomia*. Per indicare tutte le complessità che derivano da questo tipo di libertà decisionale di queste macchine, si è introdotto il termine *Artificial Moral Agent* (AMA): parlando di AMA, si indica quel settore che studia come definire dei criteri informatici per creare una sorta di *moralità artificiale* nei sistemi IA portando alcuni studiosi a coniare l'espressione *macchine morali* per questi sistemi (*cfr.* W. Wallach – C. Allen, *Moral Machines: Teaching Robots Right from Wrong*, Oxford University Press, New York 2008, 55-79). Quando si usa il termine *autonomia* legato al mondo della robotica si vuole intendere il funzionamento di sistemi di IA la cui programmazione li rende in grado di adattare il loro comportamento in base alle circostanze in cui si trovano a operare (*cfr.* E. Yudkowsky, "*Levels of Organization in General Intelligence*", *in Artificial General Intelligence – Cognitive Technologies, a cura di Goertzel, B., Pennachin, C., Springer, Berlin 2007, 389-498*).

Tuttavia, racchiudere tutta la questione degli agenti morali autonomi, dell'utilizzo di *robot* cognitivi in un ambiente di misto umano-robotico non può esaurirsi in questa direttiva primaria. Sfruttando un linguaggio evocativo, potremmo dire che le macchine "*sapiens*", per poter coesistere con i lavoratori umani, devono *imparare* almeno quattro cose.

INTUIZIONE

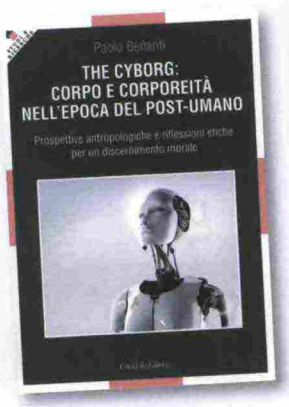
Quando due esseri umani cooperano, normalmente l'uno riesce ad anticipare e assecondare le intenzioni dell'altro perché riesce a intuire cosa sta facendo o cosa vuole fare. In un ambiente misto uomo-*robot*, le IA devono essere in grado di *intuire* cosa gli uomini vogliono fare e adattarsi alle loro intenzioni cooperando. Solo in un ambiente di lavoro in cui le macchine sapranno capire l'uomo e assecondare il suo agire, potremo veder rispettato l'ingegno e la duttilità umana. La macchina si deve adattare all'uomo e alla sua unicità e non viceversa.

INTELLEGIBILITÀ

I *robot* in quanto macchine operatrici funzionano secondo algoritmi di ottimizzazione. I *software* ottimizzano l'uso energetico dei loro servomotori, le traiettorie cinematiche e le velocità operative. Dovremmo far sì che la persona che condivide con la macchina lo spazio di lavoro possa sempre essere in grado di intuire qual è l'azione che la macchina sta per compiere.

ADATTABILITÀ

Un *robot* dotato di IA si adatta all'ambiente e alle circostanze per compiere delle azioni autonome. Tuttavia, non si tratta di progettare e realizzare algoritmi di intelligenza artificiale che siano in grado di adattarsi solo all'imprevedibile condizione dell'ambiente donando alla macchina una sorta di consapevolezza sulla realtà che la circonda. In una situazione di cooperazione e lavoro mista tra uomo e macchina, il *robot* deve *adattarsi* anche alla personalità umana con cui coopera.
Per esemplificare questa caratteristica, proviamo a fare un esempio. Supponiamo di avere un'automobile a guida autonoma. La macchina dovrà adattarsi alle condizioni del traffico: in condizioni di intenso traffico, se la macchina non possiede degli efficienti algoritmi di adattabilità, rischia di rimanere sempre ferma perché gli altri veicoli a guida umana le passeranno sempre avanti cercando di evitare l'ingorgo. Oppure, se non fosse abbastanza adattabile, rischierebbe di causare degli incidenti non capendo l'intenzione furtiva di cambiare corsia del guidatore che ha davanti. Tuttavia, vi è un ulteriore e più importante adattamento che la macchina deve saper fare: quello alla sensibilità dei suoi passeggeri. Qualcuno potrebbe trovare la lentezza della macchina nel cambiare corsia esasperante o, al contrario, potrebbe trovare il suo stile di guida troppo aggressivo e vivere tutto il viaggio con l'insostenibile angoscia che un incidente sia imminente. La macchina deve *adattarsi* alla personalità con cui interagisce. L'uomo non è solo un essere razionale ma anche un essere emotivo e l'agire della macchina deve essere in grado di valutare e rispettare questa unica e peculiare caratteristica del suo *partner* di lavoro. La dignità della persona è espressa anche dalla sua unicità. Saper valorizzare e non mortificare questa unicità di natura razionale-emotiva è una caratteristica chiave per una coesistenza che non sia un detrimento della parte umana.



Questo dossier è pubblicato con il contributo economico dell'otto per mille della Tavola Valdese.

ADEGUATEZZA DEGLI OBIETTIVI

Un *robot* è governato da degli algoritmi che ne determinano delle linee di condotta. Se in un ambiente di sole macchine l'assolutezza dell'obiettivo è una *policy* adeguata, in un ambiente misto di lavoro uomo-*robot* questo paradigma non sembra essere del tutto valido. Se il *robot* vuole interagire con la persona in una maniera che sia conveniente e rispettosa della sua dignità, deve poter aggiustare i suoi fini guardando la persona e cercando di capire qual è l'obiettivo adeguato in quella situazione. In un ambiente misto è la persona e il suo valore unico ciò che stabilisce e gerarchizza le priorità: è il *robot* che coopera con l'uomo e non l'uomo che assiste la macchina.
Se queste quattro direttrici possono essere quattro dimensioni di tutela della dignità della persona nella nuova e inedita relazione tra uomo e macchina *sapiens*, bisogna poterle garantire in maniera certa e sicura. Si devono allora sviluppare degli algoritmi di verifica indipendenti che sappiano in qualche modo quantificare e certificare questa capacità di intuizione, intellegibilità, adattabilità e adeguatezza degli obiettivi del *robot*. Questi algoritmi valutativi devono essere indipendenti e affidati a enti terzi certificatori che si facciano garanti di questo. Serve implementare, da parte del governo, un *framework* operativo che, assumendo questa dimensione valoriale, la trasformi in strutture di standardizzazione, certificazione e controllo che tutelino la persona e il suo valore in questi ambienti misti uomo-*robot*.

perapprofondire

PAOLO BENANTI HA SCRITTO NUMEROSI TESTI E ARTICOLI SULL'ARGOMENTO DELLE NUOVE TECNOLOGIE. TRA GLI ALTRI, SEGNALIAMO:

The cyborg. Corpo e corporeità nell'epoca del post-umano, CITTADELLA ED, 2012
AAVV, *Un secolo di novità complesse. Tragitti panoramici sulla scienza, sulla filosofia e sulla teologia del XX secolo*, CITTADELLA ED, 2013
La condizione tecno-umana, EDB, 2016
Postumano. Troppo postumano, CASTELVECCHI ED., 2017
AAVV, *L'animale e la macchina. Come il post-umano interpella la pastorale*, EDB, 2017