

<https://jacobinlat.com>
02.10.26

Le macchine assassine sono già qui

Intervista di Nick French a Robert Sparrow



Droni durante l'attivazione del primo battaglione di velivoli senza pilota presso la base militare di Tolemaida, in Colombia, il 10 ottobre 2025. (Raul Arboleda / AFP via Getty Images)

Nelle ultime settimane si è assistito a un'ondata di discussioni su possibili scenari apocalittici che coinvolgono l'intelligenza artificiale. Non dobbiamo dimenticare che l'IA sta già uccidendo.

persone, poiché viene sempre più utilizzato per gestire sistemi d'arma con una supervisione umana minima.

Il seguente articolo fa parte della serie "Economia politica dell'intelligenza artificiale", una collaborazione tra Jacobin Magazine e la Fondazione Rosa Luxemburg.

Nelle ultime settimane, il panico legato all'intelligenza artificiale ha raggiunto livelli altissimi. Informatori del settore dell'IA, così come politici come Bernie Sanders, lanciano l'allarme sulla possibilità di un'apocalisse causata dall'IA e sulla necessità di rallentare o sospendere lo sviluppo di questa tecnologia.

Ma non dimentichiamo che l'intelligenza artificiale sta già uccidendo esseri umani, a causa del crescente impiego di sistemi d'arma autonomi in ambito bellico. Un recente episodio di alto profilo, riportato dal New York Times il mese scorso, ha causato la morte di tre civili ucraini abbattuti da un drone guidato dall'IA. Un sistema di intelligenza artificiale è stato inoltre tristemente utilizzato per selezionare gli obiettivi negli attacchi genocidi di Israele contro Gaza.

Robert Sparrow è professore di filosofia alla Monash University, autore di numerose pubblicazioni sull'etica della guerra e sostenitore del divieto dell'uso di sistemi d'arma autonomi. Jacobin lo ha recentemente intervistato per discutere dell'attuale impiego di tali sistemi.

Discute delle armi basate sull'intelligenza artificiale, spiega perché considera il loro impiego particolarmente problematico dal punto di vista morale e cosa pensa dei recenti annunci riguardanti un'imminente catastrofe causata dall'IA.

NF

Il New York Times ha riportato il mese scorso che un drone completamente autonomo ha ucciso tre persone in Ucraina quest'estate. Iniziamo definendo cosa si intende per drone o sistema d'arma completamente autonomo. In che modo le morti in Ucraina differiscono dai precedenti casi documentati di utilizzo di sistemi d'arma semi-autonomi o senza pilota?

RS

La definizione standard di autonomia odierna è quella di un sistema d'arma che sceglie i propri bersagli o svolge una funzione significativa senza supervisione umana.

Alcuni classificheranno le munizioni "spara e dimentica" come autonome; altri richiederanno qualcosa di più simile al tipo di autonomia che stiamo osservando nell'intelligenza artificiale generativa e negli agenti di IA di cui si parla.

Riguardo a quella notizia, ciò che sembrava veramente nuovo non era tanto la capacità dell'arma. Il drone era stato progettato, essenzialmente, per colpire il gas. È stato sparato in direzione generale di una stazione di servizio, riconosce i serbatoi di propano e tenta di schiantarsi contro di essi.

Non c'è una grande quantità di

Lì era in gioco l'esito delle elezioni. La novità emersa è stata la notizia di vittime civili.

Ma questa capacità è presente nei sistemi da molto tempo. Esisteva una munizione a volo stazionario israeliana chiamata Harpy, che veniva lanciata verso una posizione in cui si temeva l'attivazione dei radar di difesa aerea. Volava autonomamente, poi qualcuno attivava un radar di difesa aerea e la munizione si dirigeva verso di esso, schiantandosi contro di esso. Gli americani hanno un missile da crociera chiamato Long-Range Anti-Ship Missile (L-RSS). Anche questo viene lanciato in direzione generale di un gruppo navale nemico. Quindi seleziona autonomamente il bersaglio di maggior valore. È dotato di radar o profili delle navi ed è progettato per volare in ondate. Una volta distrutto il bersaglio previsto, il missile riconosce il successivo bersaglio di maggior valore.

Sono contento che si stia richiamando l'attenzione su questi sistemi, ma in quel caso specifico, la novità sembrava risiedere nel fatto che si trattasse di un utilizzo russo, e non ucraino, di un drone di questo tipo; nel fatto che i russi abbiano utilizzato un chip Nvidia, piuttosto diffuso – il che dimostra che si tratta di una capacità che altri potrebbero sviluppare – e nel fatto che abbia ucciso dei civili.

L'articolo del Times fa una distinzione tra guida dell'ultimo miglio e armi completamente autonome. Non mi era chiaro in quel caso particolare [che il drone fosse completamente autonomo], perché un essere umano aveva scelto di prendere di mira le bombole di gas. Aveva scelto di sparare e aveva scelto

in generale, non era chiaro quale stazione di servizio fosse la destinazione. Sembrava che il suo funzionamento fosse simile a quello di alcuni droni armati ucraini: si poteva designare un bersaglio e il drone avrebbe inseguito quella persona o quel carro armato in situazioni in cui non era possibile mantenere un operatore umano sul posto a causa delle troppe interferenze.

Ma è chiaro che, a meno che le persone non si uniscano per impedirlo, questo è il futuro della guerra: sempre più decisioni sugli obiettivi vengono delegate alle macchine.

NF

Quali sono le principali problematiche sollevate dai sistemi d'arma autonomi, che si tratti di qualcosa di simile a quello utilizzato in Ucraina o di qualcosa di più completamente autonomo? Con ogni tipo di arma, si verificano conseguenze indesiderate. Una bomba può essere sganciata o un proiettile sparato e colpire bersagli non previsti, ad esempio. Cosa c'è nelle armi autonome o a guida autonoma che solleva particolari problemi morali?

RS

Molto dipende dal fatto che si stia pensando a qualcosa come una mina anticarro. Una mina anticarro moderna, in un certo senso, sceglie il suo bersaglio. È progettata per distinguere se ciò che le è passato sopra è un trattore nemico o un veicolo blindato. Alcune sono progettate per contare i bersagli, in modo che la colonna di carri armati nemici si trovi all'interno del campo minato quando la mina esplode.

Prima la mia. È difficile preoccuparsi troppo di questo. D'altra parte, la gente immagina Terminator o qualcuno come Claude di Anthropic, che sceglie i bersagli. In quel caso è più facile capire che qualcosa non va.

Per me, ci sono due tipi di problemi. Il primo è che queste cose aumentano prevedibilmente il rischio di una guerra accidentale. Immaginate un futuro con sottomarini autonomi (e la guerra sottomarina è un campo in cui queste tecnologie avranno un'influenza notevole, perché comunicare con i sottomarini è difficile e la gente non ama più stare a bordo di un sottomarino) e scontri come "Caccia a Ottobre Rosso" tra sommergibili autonomi al largo delle coste cinesi o al largo della costa pacifica degli Stati Uniti: non è qualcosa che vorremmo affidare a sistemi il cui comportamento nella pratica è piuttosto difficile da prevedere.

Poi c'è la questione di cosa ci dobbiamo reciprocamente, anche in tempo di guerra, e se possa esistere un diritto a non essere uccisi da un robot. Le persone hanno la ragionevole intuizione che, anche in guerra, ci siano dei limiti imposti dal rispetto per la vita umana. Non prendiamo di mira direttamente i civili. Non usiamo armi che abbiano effetti su una persona che vadano oltre quelli richiesti dalle necessità militari. Ci sono divieti, ad esempio, sull'uso di proiettili avvelenati. Ci sono alcuni modi di uccidere persone sui quali la comunità internazionale ha semplicemente detto: "No, non lo faremo".

Quindi la domanda è: se si decide di dichiarare guerra, cosa si deve al nemico? Una cosa che si deve è che un essere umano decida se vivere o morire. Il fatto che ci sia un essere umano che deve assumersi la responsabilità e prendere quella decisione: questo è un dovere che abbiamo gli uni verso gli altri.

La questione si complica. Se immaginiamo un futuro in cui io dico: "Ho ucciso qualcuno con un robot", allora sono io che lo uccido. E ho preso una decisione. Ma se immaginiamo un futuro in cui è il robot a decidere chi uccidere, si potrebbe dire: "No, questo è irrispettoso nei confronti dell'umanità del nostro nemico".

Molti credono erroneamente che l'etica della guerra si basi esclusivamente sulla vittoria e sulle sue conseguenze, ma non è affatto così. Altrimenti, si potrebbe prendere di mira l'asilo nido del nemico e sviluppare una politica difensiva basata sul principio "se mi attacchi, brucerò i tuoi figli". Questo permetterebbe di vincere rapidamente una guerra e di portare la pace. Ma questa è la logica del terrorismo.

Esiste una teoria chiamata "teoria della guerra giusta", o diritto internazionale bellico, o etica militare. In tale contesto, poniamo delle restrizioni sui tipi di armi che siamo disposti a vedere utilizzate. Spero che aggiungeremo i sistemi d'arma autonomi all'elenco delle cose assolutamente proibite.

NF

Come si traccia esattamente la distinzione tra una persona che usa un robot per uccidere qualcuno e un robot che uccide qualcuno autonomamente?

RS

La domanda ha due aspetti. Il primo è di natura pratica: quanto controllo ho sul bersaglio del robot, su ciò che distrugge? Ad esempio, con quel sistema russo, si sarebbero potuti lanciare molti droni e alcuni sarebbero atterrati sulle bombole di gas. Oppure si potrebbe costruire un sistema in grado di riconoscere le bombole di gas, e in tal caso si avrebbe maggiore sicurezza su ciò che viene attaccato.

Oggi esistono situazioni in cui le macchine superano nettamente le prestazioni umane. I sistemi di difesa aerea, ad esempio: abbattere i missili da crociera in arrivo è qualcosa che gli esseri umani semplicemente non possono fare. Se ci provassero, abbatterebbero aerei amici.

Abbatterebbero aerei di linea civili. Ma è possibile avere un cannone controllato da computer che sia piuttosto affidabile.

Si pone quindi la questione pratica di quanta fiducia si possa riporre nel fatto che un sistema d'arma colpisca un bersaglio designato e solo quello. Esistono alcuni casi d'uso in cui è possibile un elevato grado di controllo, come ad esempio nella guerra antisommergibile.

Semplicemente non ci sono molti sottomarini civili. Le probabilità di uccidere civili se utilizzo munizioni a guida subacquea sono molto basse. Posso assumermi la responsabilità.

So che ho autorizzato questa cosa che verrà lanciata e che distruggerà un sottomarino.

Riconoscere la resa è un problema complesso per questi sistemi. È difficile progettarli in modo che siano in grado di riconoscere quando qualcuno si sta arrendendo. Tuttavia, esistono tipi di conflitto in cui il riconoscimento della resa non è più una pratica comune. Fatta eccezione per il riconoscimento della resa, alcuni di questi sistemi possono aderire al cosiddetto "principio di distinzione", che impone di attaccare solo obiettivi militari legittimi.

Poi c'è il caso in cui si dice: "Vola sopra questa città e uccidi chiunque sia mio nemico" oppure "Vola sopra la città e uccidi chiunque abbia un fucile". In certi contesti, le persone portano fucili perché è quello che fanno. In un contesto del genere, non si ha molto controllo su chi si attacca.

Questo è l'aspetto pratico. Un essere umano può sapere con certezza cosa sta per attaccare? Se sì, allora si potrebbe dire: "Bene, questo è un esempio di un essere umano che uccide un altro essere umano con un robot".

Si pone inoltre la questione filosofica relativa alla natura dell'autonomia del sistema. Se si riflette sull'esperienza odierna nell'interagire con ChatGPT, è facile cadere nella trappola del linguaggio dell'agente: "Ha deciso di rispondere in questo modo" o "Mi ha prenotato la corsa". Più si usa questo linguaggio in riferimento a un sistema, meno esso appare come un essere vivente.

Un essere umano che uccide un altro essere umano con un robot, e sembra più che altro che io stia mandando un robot là fuori a uccidere persone, ma è il robot che uccide.

Quindi, parte di ciò che serve qui consiste nel prestare attenzione al linguaggio che le persone usano quando parlano di questi sistemi e a come li descrivono. Se li descrivono in un modo che suggerisce che un essere umano non abbia il controllo e non sia responsabile di ciò che accade, allora questo è irrispettoso nei confronti dell'umanità del nemico.

Le funzioni di comando saranno probabilmente affidate a sistemi di intelligenza artificiale. Una delle cose che mi interessano in questo momento è se i militari saranno disposti a permettere a questi sistemi di prendere decisioni che potrebbero comportare l'uccisione dei propri soldati.

I militari, per la maggior parte, concordano con l'idea di lasciare che sia la macchina a decidere chi tra i nemici vive e chi muore. Ma se venisse loro chiesto: "Siete disposti ad accettare ordini da un sistema di comando basato sull'intelligenza artificiale, in cui tale sistema può decidere che un gruppo di soldati debba posizionarsi su quella collina e diventare un bersaglio, in modo che altri possano aggirare il fianco sinistro e vincere?",

Sei d'accordo? Il mio sospetto è che i militari esiteranno davvero quando le loro vite saranno in pericolo. Questo la dice lunga.

NF

Parliamo di questa questione dell'affidamento del controllo a sistemi di intelligenza artificiale. Nel 2023 hai scritto un articolo con Adam Henschke sul

L'idea della "Guerra del Minotauro" è un'immagine molto suggestiva. Potresti approfondire questa metafora e spiegare perché hai previsto che avremmo assistito a un aumento dei casi in cui l'intelligenza artificiale avrebbe comandato truppe umane?

RS

Lo stiamo già vedendo. Il film NAZA, appena uscito, tratta dei sistemi israeliani che selezionavano i bersagli a Gaza. Ed è esattamente quello che io e Adam avevamo previsto. Ci sono esseri umani che lanciano armi contro bersagli selezionati da macchine.

Nessuno vende mai un sistema di intelligenza artificiale a professionisti qualificati dicendo loro che li priverà del lavoro. Quello che viene detto loro è che si occuperà del lavoro più gravoso, gestendo le attività considerate a basso valore aggiunto, liberandoli per quelle importanti.

Si avverte anche una forte influenza hollywoodiana. Quando i militari iniziarono a pianificare un futuro in cui i robot avrebbero avuto un ruolo nella guerra, optarono per una sorta di modello alla Iron Man: l'idea di un essere umano circondato da una nuvola di robot, di macchine. L'essere umano sarebbe stato al comando e i robot avrebbero eseguito la sua volontà.

Il modello che è stato sviluppato era l'idea della "guerra dei centauri". Questo è un concetto influente. L'idea è quella di avere una testa umana sul corpo di un animale.

Questo è il centauro. Si tratta di una squadra composta da umani e macchine, dove le macchine seguono gli ordini degli umani. L'idea era che questo sistema avrebbe effettivamente sconfitto i sistemi completamente autonomi, perché gli umani sono ancora capaci, in alcuni contesti specifici, di prendere decisioni superiori.

L'idea era che squadre composte da umani e robot avrebbero superato in prestazioni sia gli umani che i robot. Questo si è rivelato vero nella storia degli scacchi, almeno in un certo periodo. La gente era entusiasta quando i computer scacchistici riuscirono a battere gli umani. Ma poi si scoprì che gli umani, collaborando con i computer scacchistici, potevano battere i computer da soli. Per un breve periodo, queste squadre ibride uomo-macchina sconfissero sia gli umani che le macchine. Oggi, negli scacchi, questo non è più vero; semplicemente, vincono le macchine.

In ogni caso, questo è stato il modello. Permette ai militari di rassicurare gli ufficiali sul fatto che manterranno il loro posto di lavoro. Fa parte di una sorta di tradizione illuminista, secondo la quale la mente è superiore al corpo; secondo la quale, quando si tratta di definire chi costruisce il grattacielo, alla fine risulta essere l'architetto e non gli operai sulle impalcature.

Adam ed io abbiamo fatto notare che è piuttosto improbabile. Si è sempre pensato che i compiti mentali e cognitivi sarebbero stati davvero difficili per l'intelligenza artificiale. Pianificazione, matematica, scacchi: questi erano tutti i test. In realtà, le macchine sono diventate molto

Sono bravissimi in quello, imparano molto velocemente. Ma, d'altra parte, non sanno camminare. Non si può ancora costruire un robot che entri in ufficio e svuoti il cestino. Le cose che i bambini piccoli sanno fare – correre, riconoscere gli oggetti e raccogliarli – si sono rivelate davvero difficili. E tutte le cose cognitive si sono rivelate fattibili. Ma non riescono a fare le cose fisiche.

Pensiamo a quei contesti in cui coesistono team composti da esseri umani e macchine. Pensiamo a Uber. Per molto tempo abbiamo avuto l'idea di poter lasciare un segnaposto su una mappa e che un robotaxi sarebbe venuto a prenderci. Ma, nella maggior parte dei casi, ciò che abbiamo è manodopera immigrata, esseri umani che guidano taxi a cui un algoritmo indica la destinazione. Oppure pensiamo a cosa succede nei grandi centri di distribuzione di Walmart o Amazon, dove ci sono persone che sono essenzialmente le mani dei computer. Il lavoro umano consiste nello stare in piedi, raccogliere gli oggetti e metterli in un sacchetto di carta. Nel frattempo, l'intelligenza artificiale decide la destinazione, stampa l'etichetta e ordina i prodotti: in altre parole, si occupa di tutto il lavoro cognitivo.

Osservando la collaborazione effettiva tra esseri umani e macchine, si scopre che ciò che la fanteria può fare è molto più difficile. Ma impartire ordini è già possibile.

Pensiamo quindi che, in sostanza, ufficiali e persino generali saranno sostituiti dai computer. Ma la fanteria – le persone che dovranno strisciare nel fango e scavalcare il filo spinato – sarà composta da esseri umani. Il modello di riferimento è una testa mostruosa su un corpo umano, ovvero il Minotauro. L'intelligenza artificiale impartisce ordini agli umani, e gli umani obbediscono ai suoi ordini.

Pensiamo che queste cose potrebbero effettivamente avere successo e che le persone potrebbero preferire essere comandate da una macchina. Sono giunto a questa conclusione esaminando un progetto in cui l'esercito americano voleva dotare tutti di una sorta di Microsoft HoloLens. Oggi tutti hanno telecamere montate sulla testa; tutti avranno i Google Glass. Si poteva intuire cosa la gente immaginasse che questi sistemi avrebbero fatto: valutazione delle minacce. Gli occhiali avrebbero detto loro: "Hai 30 secondi per arrivare lì e sparare a quella persona". Era fin troppo chiaro che le persone sarebbero diventate marionette di questi algoritmi di valutazione delle minacce.

Si parte dal presupposto che le macchine possano dare consigli. Ma si pensa anche che siano capaci di prestazioni sovrumane. Se il mio consigliere è migliore di me, devo semplicemente fare quello che mi dice. Quanto al modello di mettere in discussione ciò che dicono le macchine, è valido solo per un periodo di tempo molto limitato. Se sono davvero capaci di prestazioni sovrumane, come tutti sostengono, allora gli esseri umani non dovrebbero più prendere queste decisioni. Se la macchina mi dice di abbassarmi, mi abbasso. Se non lo faccio, muoio.

Preferisco che sia una macchina a dirmelo, piuttosto che un essere umano, che è più lento e meno affidabile.

Di nuovo, questo è ciò che stiamo vedendo ora; è il ruolo che l'IA sta giocando in Ucraina o a Gaza, dove questi algoritmi assegnano essenzialmente i bersagli ai sistemi di artiglieria umana. C'è un'integrazione dei dati di ogni drone, di ogni satellite, sulla base della quale il

L'algoritmo, in sostanza, trasmette gli obiettivi alle persone.
Questo è un esempio precoce della Guerra del Minotauro.

NF

Se a prendere le decisioni sono i robot, sorgono forse diverse preoccupazioni di natura morale? Talvolta, in relazione a questo tipo di sistemi, si solleva la preoccupazione che non ci sia nessuno a cui chiedere conto delle loro azioni.

RS

La questione della responsabilità è la stessa sia per i sistemi d'arma autonomi sia per le funzioni di comando dell'intelligenza artificiale. La cosa interessante è che le persone reagiscono in modo diverso.

Se mando uno sciame di droni a distruggere carri armati nemici e mi chiedono "Chi è il responsabile?", rispondo che sono io, perché ho progettato il sistema, gli ho assegnato il compito di distruggere i carri armati e ne ho verificato il funzionamento. Lo stesso dovrebbe valere per chi progetta il sistema di comando basato sull'intelligenza artificiale. Tuttavia, sospetto che le persone saranno molto meno a loro agio all'idea di essere essenzialmente mandate a morte da una macchina, e che si aspetteranno il coinvolgimento di un essere umano.

C'è anche il rischio di deviazione dalla missione. In qualcosa come la logistica, le macchine sono davvero brave a gestire grandi operazioni, con molti

Camion e container. Se sono le macchine a decidere quali truppe ricevono le armi, allora stanno già determinando chi vive e chi muore. Se la macchina dice: "Sei terzo nella nostra lista di priorità e oggi non riceverai munizioni. Ma qualcuno deve mantenere questa posizione mentre gli altri si ritirano, e quella persona morirà", la macchina sta dicendo che sarai tu a morire. Penso che le persone si sentiranno profondamente a disagio con questa idea, anche se non hanno problemi a inviare questi elicotteri d'attacco autonomi per distruggere carri armati o individuare i movimenti delle truppe nemiche.

Per quanto riguarda il sistema israeliano Lavender, utilizzato per selezionare gli obiettivi a Gaza: qualcuno lo ha creato e, a quanto ho capito, il cursore per le vittime civili poteva essere alzato o abbassato, regolando il numero accettabile di vittime civili. C'erano persone che verificavano gli obiettivi, ma impiegavano venti secondi per tale verifica. E qualcuno decideva il livello di prova necessario per stabilire che una persona fosse un membro di Hamas. In altre parole, c'erano degli esseri umani coinvolti, ma a un certo livello è abbastanza chiaro che questo algoritmo selezionava gli obiettivi e che gli esseri umani si limitavano a fare ciò che il sistema diceva loro di fare.

NF

A questo proposito, lei sostiene il divieto dei sistemi d'arma autonomi e afferma che un divieto categorico sarebbe più realistico rispetto al tentativo di imporre opportune restrizioni morali e legali. Può spiegarci questo ragionamento?

RS

Quasi nessuno pensa che queste cose siano completamente innocue. Persino l'esercito è molto preoccupato: qual è l'impatto sulla coesione delle unità?

Come faccio a sapere che questo aggeggio non ucciderà accidentalmente le mie truppe?

Tutti pensano di metterli a capo, ad esempio, di un secondo attacco nucleare; solo il vero dott.

Strangelove è disposto a percorrere quella strada. Esiste un precedente storico per quello che viene definito "lancio di allarme precoce". Sembra che per un certo periodo l'arsenale nucleare russo sia stato progettato in questo modo. I radar di allarme precoce sono collegati al computer, al sistema di controllo del tiro, perché si teme che qualcuno possa effettuare un attacco mirato a decapitare il nemico, e questo fornisce una capacità affidabile per un secondo attacco. Ciò è incredibilmente pericoloso, considerando i limiti che conosciamo del nostro sistema.

La maggior parte delle persone, quando pensa a queste cose, afferma che è necessaria una regolamentazione. Ma tutti riconoscono anche che si tratta di una tecnologia molto difficile da regolamentare. Si può avere un sistema che funziona da remoto, con un essere umano in collegamento, ma poi qualcuno collega una chiavetta USB e, improvvisamente, il sistema si autoalimenta.

Mi sembra che, se vogliamo avere un impatto sul futuro della guerra, tracciare delle linee rosse che rendano politicamente difficile per le persone impiegare questa tecnologia sia più efficace di qualcosa che, sulla carta, potrebbe essere più giustificato. C'è una sorta di

Il dibattito sul divieto delle mine antiuomo verte su questo punto: il divieto dovrebbe applicarsi solo alle mine antiuomo o anche a quelle che non si autodistruggono? Esistono diverse soluzioni tecnologiche per le vittime civili delle mine antiuomo. Tuttavia, affermare semplicemente che non sono ammesse crea una chiara aspettativa morale e rende più facile, ad esempio, per le persone rifiutarsi di lavorare con esse.

Sarà davvero necessario che la professione ingegneristica dica: "Non costruiremo macchine che scelgono i propri obiettivi". I tentativi più sofisticati di regolamentare questi sistemi sono più difficili da verificare e lasciano più margine di manovra. È un po' come dire: "Ok, il mio sistema non è completamente autonomo...".

Dobbiamo prestare attenzione a quella che viene definita la funzione espressiva della regolamentazione, che consiste nel rendere chiaro che siamo tutti d'accordo su un divieto.

Se si guarda alla storia, ad esempio, del divieto dei gas velenosi, si trovano molte persone che affermano: "Questo è un metodo di guerra molto più umano"; "Vinceremo le guerre più velocemente"; "Che differenza c'è tra morire per gas o per armi leggere? Morire per armi leggere è orribile".

Ma il divieto categorico stronca tutto questo sul nascere.

Si tratta di una discussione su come avvengono i cambiamenti sociali e politici in questo ambito. Gestì audaci e di grande impatto hanno molte più probabilità di essere efficaci rispetto a un tentativo burocratico e frammentario di perfezionare i dettagli. Questo non farebbe altro che prolungare il dibattito all'infinito e...

Dà alle persone un margine di manovra, in modo che non si trovino di fronte al vero orrore di ciò che stanno facendo.

NF

Vorrei conoscere la tua opinione sul recente ciclo di notizie allarmistiche riguardanti l'intelligenza artificiale, l'idea che essa rappresenti una minaccia esistenziale per l'umanità e che sia necessaria una sorta di pausa o rallentamento del suo sviluppo.

RS

Innanzitutto, va detto che è incredibilmente difficile per chiunque al di fuori della cerchia ristretta di queste aziende comprendere appieno di cosa siano capaci le modelle. Molti aspetti dell'incidente del "volto che abbraccia" non sono ancora stati rivelati a causa della forte censura interna esercitata dalle organizzazioni.

Gran parte del nostro modo di pensare al futuro è stato plasmato dalla fantascienza. Gli ingegneri si sono prefissati di costruire oggetti futuristici. Che aspetto avrà il futuro? Sarà come Star Trek degli anni '70. C'è una chiara influenza fantascientifica, inclusa l'idea di un'apocalisse robotica. C'è anche un'interessante dinamica politica di razza e classe che ruota attorno ai robot, perché l'idea stessa di robot era originariamente, in sostanza, quella di una sorta di lavoratore. I timori di ribellioni robotiche erano timori di socialismo.

Tutto ciò rende difficile capire quanto siano giustificate queste ondate di panico. Detto questo, se dicessi che c'è una probabilità del 10% che al-Qaeda uccida tutti nel mondo – se immaginaste una minaccia terroristica in cui le persone dell'industria della difesa dicessero: "Crediamo che ci sia una probabilità del 10% che questa organizzazione terroristica uccida tutti" – nessuno direbbe: "La la la, vediamo cosa succede".

Quindi, sono propenso a credergli sulla parola. Figure importanti nella storia dell'IA si sono espresse apertamente affermando che essa rappresenta una minaccia esistenziale. Da un lato, non c'è nulla di così assurdo da non poter trovare un professore che ci creda. Il fatto di essere un fisico o un informatico brillante non significa essere qualificati per parlare del futuro, perché il futuro non è solo una questione tecnica. È politico: è qualcosa che possiamo cambiare.

Ma visto il numero di persone che affermano che questo rappresenta un pericolo e che ci ucciderà tutti, penso che dovremmo chiudere tutti i laboratori. Semplicemente chiuderli tutti. Trovo davvero strana l'idea di procedere con cautela o di fare una pausa, considerando ciò che hanno detto sulla natura della minaccia.