volerelaluna LA POLITICA PUNTOACAPO

Fermiamo i “robot-killer”

Autore: Norberto Patrignani

L’intelligenza artificiale incombe. Tra entusiasmi, illusioni, paure. Spesso con una parificazione alla magia. Per lo più senza una consapevolezza, anzitutto politica, del suo potenziale ruolo sociale, della sua portata, delle modalità di un suo governo razionale. Per contribuire a colmare questa lacuna, abbiamo deciso di lanciare qualche punto, qualche goccia nel mare. Il primo contributo, di Davide Lovisolo, è stato pubblicato il 20 febbraio e può leggersi in https://vll.staging.19.coop/societa/2024/02/20/la-saga-di-sam-altman-lintelligenza-artificiale-e-la-vittoria-del-profitto/. (la redazione).

«*Non ci aspettavamo che nel mondo fosse necessaria una campagna per fermare i robot killer, ma lo è*».
Campagna internazionale *Stop Killer Robot*, 2024

Il 22 dicembre 2023 l’Assemblea Generale dell’ONU ha approvato (152 paesi favorevoli, 4 contrari e 11 astenuti) una **risoluzione contro i sistemi di armi letali autonome *(Lethal Autonomous Weapons Systems*, LAWS) noti come “robot-killer”**. Perché questo importante pronunciamento? Negli ultimi mesi in tutti i teatri di guerra vengono impiegati i robot-killer, comunemente chiamati “droni”. La diffusione di queste tecnologie, soprattutto quelle basate sulla cosiddetta “intelligenza artificiale”, pone profondi interrogativi tecnici, etici e legali.

Aspetti tecnici

Cosa si intende per robot-killer? Vediamo la definizione fornita dal Dipartimento della Difesa USA: «armi che, una volta attivate, sono in grado di selezionare e attaccare un obiettivo senza ulteriori interventi da parte dell’uomo» (DOD, 2023). Dunque sono macchine in grado di uccidere che basano questa decisione su algoritmi definiti da chi le ha progettate (hardware) e configurate (software). Nella generazione precedente di queste armi, l’umano era comunque coinvolto in quanto manovrava a distanza la macchina (con joystick, tastiera, comandi vocali etc.). Quando invece si parla di robot-killer si intendono macchine che agiscono senza più il coinvolgimento dell’umano. Vengono programmate per riconoscere una posizione (es. latitudine, longitudine con GPS) e un’immagine (di un edificio, di un mezzo militare, di una persona etc.) per poi colpire l’obiettivo riconosciuto.

Decisioni automatizzate sono sempre più diffuse nella società ma, essendo queste basate su “algoritmi che si calibrano con tanti dati” (spesso denominati “intelligenze artificiali”), vi sono enormi problemi legati agli inevitabili bachi (cioè guasti che causano dei malfunzionamenti) del software, ai pregiudizi incorporati nei dati stessi, soprattutto quando questi algoritmi comandano armi autonome. Infatti queste macchine sono tecnicamente delle “macchine a stati finiti” (in grado di svolgere un’attività senza l’intervento

dell'uomo) completamente diverse da quelle tradizionali basate su sistemi lineari. Il funzionamento di un sistema lineare può essere garantito purché le sollecitazioni a cui viene sottoposto siano entro una certa soglia. Nelle macchine "a stati finiti" questa soglia non esiste e, dato che bastano poche linee di software per arrivare a un numero di stati dell'ordine dei miliardi, è impossibile testare in modo esaustivo il corretto funzionamento del sistema in tutti i suoi stati. Come insegnava un grande scienziato dei computer, Edsger Dijkstra (1930-2002): «il test di un programma può essere usato per mostrare la presenza di bug ma mai per dimostrare la loro assenza!» (Dijkstra, 1972). La scienza dei computer suggerisce infatti di progettare il software in modo tale che, se dovesse succedere qualcosa di indesiderato, deve essere sempre possibile limitare le conseguenze, compensarle o, ancora meglio, "tornare indietro" (*undo*). Ma se una macchina uccide un umano per errore non è possibile tornare indietro. Questo è il motivo fondamentale che ha portato nel 1985 un altro grande scienziato dei computer, David Parnas (fondatore dell'*object-oriented programming, dell'information-hiding*), alle dimissioni dal comitato scientifico che doveva supervisionare il progetto USA *Strategic Defense Initiative*, SDI (noto come *Star Wars).* Il progetto prevedeva il lancio automatico di missili intercontinentali in risposta a un attacco e Parnas si rifiutò di sostenere il progetto motivando che «il software viene rilasciato per l'uso non quando si sa che è corretto, ma quando il tasso di scoperta di nuovi errori rallenta fino a raggiungere un valore considerato accettabile. […] A causa della nostra incapacità di testarlo, non possiamo, come scienziati, […] approvare questa applicazione» (Parnas, 1985). Nel nostro tempo, le persone esperte di informatica devono assumersi la responsabilità di spiegare tutto questo al pubblico, di dire cosa può essere e cosa non può essere automatizzato.

Aspetti etici

Joseph Weizenbaum (1923-2008), docente al MIT negli anni '60, ha dato un contributo fondamentale alla riflessione etica sui limiti e sui rischi dell'automazione. Nel 1976 pubblica uno dei testi fondamentali dell'etica digitale: *Il potere del computer e la ragione umana*. In esso propone di distinguere tra "decisioni" (che si possono affrontare con il pensiero computazionale, con un algoritmo e che quindi possono essere programmate su un computer) e "scelte" (che comportano un giudizio, non sono il risultato di un calcolo; è proprio la capacità di scegliere che ci rende umani). Le "scelte" sono funzioni umane che «non dovrebbero essere delegate a macchine» (Weienbaum, 1976).

L'umanità rischia un impercettibile "slittamento epistemologico" dalla *previsione*, alla *predizione*, fino alla *prescrizione*. Infatti l'illusione di creare "intelligenze" sovrumane, solo perché in grado di elaborare quantità di dati inaccessibili all'umano (es. dell'ordine degli Exa-byte, miliardi di Gigabyte), sta sostituendo progressivamente l'uso delle macchine per fare *previsioni* (*prae-videre*, utili per stimare possibili eventi futuri), all'uso delle macchine per fare *predizioni* (*prae-dicere*, utili per formulare ipotesi future), fino ad arrivare a delegare completamente alle macchine: la macchina fa *prescrizioni* (*prae-scribere*, ordina e l'umano esegue, o la macchina esegue direttamente senza coinvolgere più l'umano).

La delega alle macchine di *scelte* importanti irriducibili a un algoritmo (*decisioni*), a un calcolo, rischia di portare al "sonno della ragione" (Goya, 1746-1828). Come raccomandava il fondatore della cibernetica, Norbert Wiener (1894-1964): «prima di delegare alla macchina scelte sarà meglio assicurarsi che lo scopo che immettiamo nella macchina sia lo scopo che realmente desideriamo» (Wiener, 1960). Forse aveva ragione Goethe che nell'*Apprendista stregone* (1797) metteva in guardia dal rinunciare alle nostre responsabilità delegandole a una macchina che rischiamo di non controllare, provocando danni irreversibili.

L'automazione della guerra con i robot-killer rischia di innescare una "corsa agli armamenti automatizzata": abbassando la soglia della guerra. Si rischia il *rebound effect*, la guerra diventerà ancora più diffusa, con una destabilizzazione a catena. Da tenere in considerazione anche i diversi tempi di risposta dell'umano, incomparabilmente più lenti rispetto alle macchine. Un aspetto inquietante è l'implicita impossibilità per i robot-killer di rispettare le leggi umanitarie internazionali (come la distinzione tra militari e civili e la proporzionalità della risposta armata). Come pure l'impossibilità di trovare colpevoli in caso di crimini di guerra (il classico *many hands problem* tipico dei sistemi basati su software). Infine forse il punto più grave, la dignità dell'umano non viene rispettata: non è possibile implorare e ottenere gesti di pietà umana da un robot-killer (Amoroso e Tamburrini, 2017; 2020).

Un aspetto cruciale riguarda il ruolo delle persone esperte di informatica che progettano questi sistemi. Esiste un codice etico per l'informatica? Il mondo della ricerca nel digitale ha visto crescere il numero di *computer professional* che mettono in discussione il coinvolgimento delle loro imprese in progetti militari (Shane, 2018). Il dibattito è aperto fin dagli albori dell'era dei computer. Norbert Wiener, nel 1947 dichiara: «Non penso di pubblicare nessun mio lavoro futuro che potrebbe fare danni nelle mani di militaristi irresponsabili» (Wiener, 1947). Fino ad arrivare al Codice etico per *computer professional* (*Code of Ethics and Professional Conduct*) definito da un comitato internazionale (ACM, 2018) che impone tra i principi fondamentali "non fare del male", "contribuire al bene comune" e, in caso di conflitto tra "interessi privati" e "bene collettivo", dare priorità al bene collettivo. Anche la professione informatica si è dotata di una deontologia. L'etica digitale (*computer ethics*) è un'area in grande espansione per tutti questi motivi, data la pervasività del digitale e la progressiva "delega all'algoritmo". D'altra parte la comunità informatica ha avviato una discussione su un approccio *Slow Tech*, verso un'informatica *buona*, *pulita*, e *giusta* (Patrignani e Whitehouse, 2018). Le persone esperte di informatica, *computer professionals*, sanno come è fatto un sistema digitale e come funziona: è tempo di chiedersi *come e perché* progettarlo, *per chi*, per *quali scopi*? Fino alla domanda cruciale: *se* progettarlo! È tempo insomma di assumersi delle responsabilità.

Infine è da segnalare l'importante presa di posizione contro i robot-killer da parte della Croce Rossa Internazionale nel 2021: «I sistemi d'arma autonomi […] sollevano

preoccupazioni etiche fondamentali per l'umanità, sostituendo di fatto le decisioni umane sulla vita e sulla morte con processi basati su sensori, software e macchine» (ICRC, 2021).

Aspetti legali

Dal 2013 è partita una campagna internazionale per la messa al bando di queste "armi autonome": *Stop Killer Robots*. È supportata da 250 organizzazioni non governative di 70 paesi diversi (SKR, 2023). In Italia ha come partner l'Unione Scienziati per il Disarmo, la Rete italiana Pace e Disarmo e l'Archivio Disarmo e il supporto dello scienziato Giorgio Parisi, Premio Nobel per la Fisica 2021: «Ci sono anche dei risvolti pericolosi dell'intelligenza artificiale. […] Il sistema di armi letali non può essere lasciato alle macchine, non possono essere loro a decidere chi uccidere o meno» (Arachi, 2021). Un primo grande risultato della campagna si è ottenuto il 22 Dicembre 2023, con la votazione dell'Assemblea Generale delle Nazioni Unite che ha adottato la prima risoluzione sulla «necessità urgente per la comunità internazionale di affrontare le sfide e le preoccupazioni sollevate dai sistemi di armi autonome, consapevoli che le nuove applicazioni tecnologiche in ambito militare, comprese quelle legate all'intelligenza artificiale e all'autonomia dei sistemi d'arma, sollevano problemi anche dal punto di vista umanitario, legale, di sicurezza ed etico» (UN, 2023). Dopo tanti anni di discussioni internazionali, in un contesto di rapidi sviluppi tecnologici, questo voto rappresenta un passo avanti fondamentale. E apre la strada alla messa al bando internazionale delle armi autonome.

Conclusioni

Un ottimo strumento di sensibilizzazione, adatto soprattutto per le scuole, è il documentario *immoral-code* (IC, 2024); un documentario prodotto nell'ambito della campagna *Stop Killer Robots* che riflette sull'impatto di queste armi in un mondo sempre più automatizzato, dove le macchine autonome arrivano a prendere decisioni su chi uccidere. Nel documentario diverse persone vengono esposte a interrogativi etici sempre più complessi ai quali vengono fornite risposte spesso diverse e a volte poco convinte. Solo alla domanda finale («*should a machine be able to decide if someone lives or dies?*») tutte rispondono in modo convinto allo stesso modo. Il documentario sta circolando in tutte le scuole e rappresenta un'ottima occasione per discutere di questi temi e per conoscere la campagna *Stop Killer Robot*.

Riferimenti

– ACM (2018), *Association for Computing Machinery, Code of Ethics*, https://www.acm.org/code-of-ethics

– Amoroso, D. Tamburrini, G. (2017). *The Ethical and Legal Case Against Autonomy in*

Weapons Systems, “Global Jurist”, n. 17, v. 3

– Amoroso, D., Tamburrini, G. (2020), *Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issies*, Current Robotics Report (ed. G. Verruggio)

– Arachi A. (2021), Giorgio Parisi: “Il mio Nobel per la Fisica a sostegno dell’intelligenza artificiale e del pianeta”, *Corriere della Sera*, 5 ottobre 2021.

– Dijkstra, E. (1972), *Notes On Structured Programming*, Technical University of Eindhoven, The Netherlands

– DOD (2023), *Autonomy in Weapon Systems*, DoD Directive 3000.09, p. 21, Department of Defense, USA

– ICRC (2021), *ICRC position on autonomous weapon systems*, www.icrc.org, Geneva, 12 May 2021

– IC (2024), *Immoral Code*, https://www.stopkillerrobots.org/take-action/immoral-code/

– Parnas, D. (1985), *Why the SDI software system will be untrustworthy*, American Scientist, 73:5, Sept-Oct 1985, 432-440

– Patrignani, N., Whitehouse, D. (2018), *Slow Tech and ICT: A Responsible, Sustainable and Ethical Approach*, ? Palgrave Macmillan

– Shane, Scott, Wakabayashi, Daisuke (2018, 4 Aprile). “*The Business of War”: Google Employees Protest Work for the Pentagon*, *New York Times*

– SKR (2023), *Stop Killer Robots* – about us, www.stopkillerrobots.org

– UN (2023), *United Nations, General Assembly, Agenda item 99, Lethal autonomous weapons systems*, A/C.1/78/L.56

– Weizenbaum, J. (1976), *Computer Power and Human Reason: From Judgment To Calculation*, Freeman (trad. it. *Il potere del computer e la ragione umana*, Edizioni Gruppo Abele)

– Wiener, Norbert (1947, January), A Scientist Rebels, Atlantic Monthly

– Wiener, N. (1960), *Some Moral and Technical Consequences of Automation*, Science,

May 6.