POLICY BRIEF

InterAgency Institute
BEYOND INSTITUTIONAL BOUNDARIES

AI SUMMER: ALLERTE E PROPOSTE SULLA REGOLAMENTAZIONE DEGLI STRUMENTI INTELLIGENTI E DEI SISTEMI DI ARMI LETALI AUTONOME (LAWS)

Autore: Juliana Miranda Martins[I]

REPIOLOGO

Nell'immaginario collettivo delle civiltà, dalla mitologia greca alla storia delle religioni, è sempre esistito un rapporto dialettico di ammirazione e timore nei confronti dell'esistenza di un'intelligenza fisica o spirituale superiore a quella umana. All'inizio del '900, dai romanzi di Isaac Asimov e Aldous Huxley e con l'avvento della fantascienza nell'industria cinematografica, alcuni scenari a quel tempo intravisti solamente nelle pagine e negli schermi sono ora diventati reali per effetto dell'utilizzo delle tecnologie emergenti in azioni degli esseri umani ordinari. Negli ultimi anni, in nome di un mondo quasi privo di regole nel cyberspazio, le *Big Tech* hanno lavorato allo sviluppo di tecnologie con intelligenza artificiale, senza preoccuparsi degli effetti collaterali del *deep learning*. La mancanza di controlli esterni da parte degli Stati e la tardiva risposta della comunità scientifica nel mettere in discussione l'estrema libertà con cui operano tali società, ha determinato una lacuna legislativa e riflessiva sugli impatti i e le basi etiche e morali delle nuove tecnologie nelle società postmoderne. Lo scenario si è complicato a fine 2022 con il lancio del CHAT-GPT e, intensificando la corsa “geo-tecno-politica” all'IA da parte degli Stati e innescando, in contemporanea, diverse allerte mondiali, anche da parte delle stesse *Big Tech*, sui potenziali rischi dei tuoi strumenti intelligenti per il futuro.
Questo policy brief si propone una breve discussione su due documenti recenti nell’ambito del dibattito sui sistemi d'arma autonomi (LAWS): la lettera aperta *"Pause giant AI Experiments: An Open Letter"* (1) dell’istituto Future of life e la proposta presentata da alcuni Stati membri nell'ultima *Convenzione CCW sulla proibizione o la limitazione dell’uso di alcune armi convenzionali che possono essere considerate dannose o aventi effetti indiscriminati* (2)

SFONDO

Negli ultimi 5 mesi, una vera e propria "corsa geo-tecno-politica" tra le grandi aziende tecnologiche globali ha generato un intenso dibattito sugli usi dell'Intelligenza Artificiale e sui suoi impatti economici, sociali, morali ed etici per le società postmoderne. Allo stesso tempo, i suoi stessi creatori, i CEO e i ricercatori mettono in guardia sul potenziale rischio di questi sistemi a causa delle numerose incertezze derivanti dallo sviluppo di intelligenze Artificiali generative.
Uno degli allarmi più recenti è stato lanciato nel marzo 2023 dall’istituto *Future of life* con *"Pause giant AI Experiments: An Open Letter"*. La lettera è stata firmata da più di mille imprenditori e ricercatori informatici che rappresentano un'iniziativa del settore privato.

N.2 MAI 2023
ISSN 2789-8040

Il documento afferma che il progresso dell'IA, ai livelli e alla velocità con cui sta attualmente progredendo, potrebbe rappresentare un profondo cambiamento nella storia dell'umanità e dovrebbe essere pianificato e gestito con risorse proporzionate ai pericoli che rappresenta. Mette in guardia sui rischi di una "corsa incontrollata" verso lo sviluppo di menti digitali ogni volta più potenti, al punto che nemmeno i loro creatori possono comprenderle, prevederle o controllarle. Invita i laboratori di intelligenza artificiale a una sorta di *"AI Summer"*, ovvero una pausa di 6 mesi nello sviluppo di sistemi superiori a GPT-4 suggerendo che, se tale interruzione non può essere attuata rapidamente, spetta ai governi intervenire e istituire una moratoria. L'"AI Summer" sarebbe il tempo necessario ai laboratori per stabilire una serie di protocolli di sicurezza con criteri di accreditamento riconosciuti a livello internazionale e supervisionati da esperti indipendenti. Lo stesso documento aggiunge che tale pausa non si riferisce all'intelligenza artificiale in generale, ma alla pericolosa proliferazione di modelli *black-box*, sempre più grandi e imprevedibili. Auspica altresì il progresso di sistemi sicuri, interpretabili, trasparenti, ma anche allineati, affidabili e leali.

La cosa più importante, ma anche la meno pubblicizzata dai media, è il fatto che la lettera evoca la necessità di un lavoro congiunto tra i laboratori di intelligenza artificiale e i politici. La *governance* dell'IA prevede misure che dovrebbero includere: autorità di regolamentazione specializzate in IA; supervisione e monitoraggio dei sistemi di intelligenza artificiale; sistemi di provenienza delle filigrane per poter distinguere il reale dal sintetico e tenere traccia delle fughe di modelli; un sistema di audit, certificazione e responsabilità per i danni causati dall'IA; finanziamenti pubblici per la ricerca di risorse, per la sicurezza e protezione dalle drastiche trasformazioni economiche e dai rischi per i modelli democratici che l'IA potrebbe causare in futuro.

Un altro avviso importante dato dai media internazionali è apparso con l'intervista di Geoffrey Hinton, ex dirigente del settore AI di Google, al The New York Times. Lo psicologo cognitivo e informatico ha messo in guardia sui rischi dell'IA generativa, tra i quali, la riduzione massiccia di posti di lavoro che potrebbe far crollare l'attuale modello socio-economico, nonché sui rischi per il modello democratico derivanti da "ondate di disinformazione mondiale" che causerebbero l'incapacità delle persone di distinguere le informazioni vere da quelle false. Ha citato il problema dell'autonomia nei sistemi di armi letali come un problema del futuro, utilizzando la satirica figura del "robot armato con una tecnologia fuori calibrazione". L'intervista è stata concessa per ampliare il supporto relativo all'"*AI Summer*", ma è ben lungi dall'entrare in questioni rilevanti, dal momento che Hinton è un precursore nella ricerca sulle reti neurali sviluppata da Google Brain dal 2016. Un esempio è l'esperimento effettuato con le tre intelligenze artificiali, chiamate Alice, Bob ed Eve che avevano il compito di raggiungere obiettivi prestabiliti: Alice doveva trasmettere un messaggio crittografato a Bob, Bob doveva decodificarlo e proteggerlo ed Eve doveva intercettarlo e decodificarlo. I risultati hanno mostrato che i primi due avevano sviluppato una propria crittografia, riuscendo a comunicare in modo completamente confidenziale, senza alcuna precedente base algoritmica crittografica programmata da un essere umano. In questo senso si può concludere che l'IA ha generato un nuovo linguaggio crittografico completamente autonomo rispetto all'uomo. Da allora, la ricerca sulle reti neurali artificiali ha fatto progressi vertiginosi.

N.2 MAI 2023
ISSN 2789-8040

RISULTATI

Da parte degli Stati membri delle Nazioni Unite, è stato presentato un protocollo per la regolamentazione dei sistemi di armi letali autonome con la *Convenzione CCW sulla proibizione o la limitazione dell'uso di alcune armi convenzionali che possono essere considerate dannose o aventi effetti indiscriminati* (2), presentata da un gruppo di Paesi (3) che hanno in comune una storia di conflitti interni o esterni ed esperienze con armi proibite come, ad esempio, le mine terrestri antipersona.

Il protocollo riguarda i sistemi di armi letali autonome (LAWS) che presuppongono l'utilizzo dell'IA per scopi bellici. In questo contesto, l'uso dell'intelligenza artificiale si materializza nel concetto di "autonomia", facendo a meno dell'azione umana, nell'esecuzione di una serie di azioni di guerra bellicose e strategiche che caratterizzano la selezione degli obiettivi fino all'applicazione della forza. Il protocollo affronta i "gravi problemi etici, legali, umanitari e di sicurezza" imposti da questi sistemi di sicurezza e descrive il concetto di "autonomia" basato su una serie di funzioni belliche. Esso spiega quali sono le azioni sul campo che richiedono un "significativo controllo umano" incluso il mantenimento della gestione umana e nel decidere l'intervento effettivo sull'uso della forza e in una serie di situazioni specifiche inclusa l'attribuzione di responsabilità e contabilità.

Il documento propone la proibizione dell'intera catena produttiva e logistica dei LAWS che non possono garantire un significativo controllo umano, comprese "quelle che operano in modo che non può essere previsto, spiegato, anticipato, compreso o tracciato" Le parti contraenti si impegnano a garantire la supervisione umana dei LAWS al fine di consentire l'intervento e la disattivazione in tutte le fasi di un'azione; garantire la capacità di azione e informazione degli esseri umani per controllare l'intero processo, specialmente nelle azioni restrittive e di ritiro; e infine impegnarsi a "istituire misure e meccanismi per prevenire pregiudizi di automazione nelle operazioni di sistema e per escludere tali pregiudizi algoritmici, inclusi i pregiudizi di genere e razza, nelle capacità di intelligenza artificiale invocate in relazione all'uso di un sistema di armi".

La proposta innova chiedendo agli Stati membri di impegnarsi alla trasparenza riguardo agli effetti derivanti dall'autoapprendimento dei LAWS, oltre a incoraggiare gli altri membri all'identificazione di tali effetti e all'elaborazione di buone pratiche nella revisione sistematica dei LAWS. Nel protocollo emerge la preoccupazione di mitigare i rischi delle armi derivanti dalle nuove tecnologie, invitando gli Stati membri a salvaguardare la sicurezza fisica e non fisica da eventuali attacchi informatici, dalla contraffazione, acquisizione e pirateria relativamente a tali tecnologie da parte di attori non statali, di gruppi terroristici di varia natura, nonché a provvedere alla formazione del personale che possa impedire la diffusione illegale di informazioni riservate e la violazione del protocollo. Ed infine, si invita a cooperare, anche con l'uso degli strumenti giuridici delle Nazioni Unite nel caso di eventuali controversie o per divergenze di interpretazione del protocollo, riunendosi una volta all'anno e fornendo relazioni annuali sulla sua applicabilità e aggiornamento.

N.2 MAI 2023
ISSN 2789-8040

CONCLUSIONI

I documenti presentati in questo policy brief, rappresentano, seppur in modo asimmetrico, un'evoluzione rispetto alle riflessioni sull'IA, soprattutto per quanto riguarda la regolamentazione del settore a livello internazionale. Secondo gli esperti l'idea dell'*"AI Summer"* è idealistica e non basta a risolvere il problema. La lettera non solleva la questione fondamentale dell'uso dell'IA nelle armi, quindi, non menziona i rischi che i LAWS pongono in termini di aggressione ai Diritti Umani. Tuttavia, si presuppone che tali tecnologie abbiano dimensioni ancora sconosciute agli stessi creatori, pur restando inteso che l'IA generativa può acquisire "autonomia" rispetto all'intelligenza umana. Si è anche convinti che l'uso di tali tecnologie abbia un profondo impatto sull'attuale modello economico e sociale e che sia necessaria una *governance* internazionale per lo sviluppo e l'applicazione dell'IA che preveda la regolamentazione, la responsabilità e *l'accountability* da parte dei produttori.

La comunità scientifica internazionale ritiene che la lettera avrebbe potuto essere più specifica. Il documento suonava come un bizzarro mea culpa o un monito che induce più al dubbio che all'azione, da parte di un settore che ha usato e abusato delle libertà del mondo virtuale per espandere le imprese su scala globale, infrangere le regole dei mercati nazionali e ridefinire le regole del mercato del lavoro, per non parlare della mancanza di impegno fiscale del mercato virtuale, della mancanza di trasparenza nell'uso e nella vendita dei dati privati degli utenti, e così via.

E ciò che è più grave, secondo la nostra analisi, è il fatto che la lettera ignori i LAWS come la dimensione più aggressiva dell'"Era AI", data la sua stessa natura di dominio, lasciando il lettore alla deriva per l'assenza di principi etici e parametri morali necessari per la creazione, lo sviluppo e la programmazione degli algoritmi di questi strumenti. Seguendo la logica del liberalismo globale, la lettera assegna agli Stati la responsabilità di vigilare e finanziare la *governance* dell'IA. Si intuisce che i danni generati dall'estrema libertà di cui ha sempre goduto il settore informativo debbano essere mitigati da un'azione governativa globale, suggerendo altresì che anche le Big Tech debbano essere ritenute responsabili. Le *Big Tech* sono state protagoniste nella generazione di un ambiente deregolamentazione delle istituzioni sociali dal cyberspazio.

Il protocollo presentato alla CCW, a sua volta, esemplifica una proposta di alcuni Stati membri in merito all'uso più letale dell'IA: da un lato, propone una regolamentazione efficace in relazione all'autonomia dei LAWS e, dall'altro, definisce categoricamente un concetto opposto, ovvero qual è il "controllo umano significativo" su tali armi, senza trascurare responsabilità, *accountability*, mitigazione degli effetti, la proprietà sconosciuta dei sistemi di autoapprendimento e, infine, il controllo nello sviluppo di nuove tecnologie belliche emergenti. Visto che i Paesi firmatari del protocollo sono Paesi del sud del mondo, che hanno conosciuto gli effetti disastrosi di alcune armi letali nei propri territori, che hanno subito (o subiscono) le azioni di governi autoritari e violazioni di Diritti umani, disuguaglianze sociali e conflitti etnici, religiosi e di genere - in molti casi causa di movimenti migratori e di popolazioni vulnerabili - il protocollo rappresenta per gli altri Stati membri un importante punto di partenza nella formulazione di proposte contro l'uso dei LAWS.

N.2 MAI 2023
ISSN 2789-8040

SUGGERIMENTI

Per concludere, suggeriamo alcune misure da prendere in merito:

1. L'intelligenza artificiale deve essere creata e sviluppata per una cultura di pace tra i paesi. I documenti concordano sulla necessità di stabilire parametri etici e morali affinché l'IA sia uno strumento per lo sviluppo umano. In nome di questo impegno, tutti gli attori sociali coinvolti, le *Big Tech* e i laboratori indipendenti, i governi, la comunità scientifica dei gruppi multidisciplinari, devono costruire protocolli e certificazioni internazionali il cui punto di partenza sia la proibizione dei sistemi di armi letali autonomi;
2. Abbandonare a tempo indeterminato la ricerca sull'Intelligenza Artificiale generativa con le reti neurali;
3. I documenti indicano la strada per soluzioni normative sull'uso delle IA, sia nel campo degli strumenti virtuali a disposizione della popolazione generale, sia in termini di armi militari. Entrambi puntano a soluzioni che riguardano il controllo dell'autonomia, l'autoapprendimento, gli usi previsti e imprevisti, il pericolo di uso improprio di informazioni e tecnologie e la programmazione algoritmica;
4. Rafforzare il ruolo degli esseri umani nella formazione dei programmatori e del personale dei laboratori di intelligenza artificiale e di qualsiasi tipo di operazione che crei, sviluppi e aggiorni strumenti nel settore dell'intelligenza artificiale sull'uso legale e responsabile delle tecnologie emergenti e sui limiti etici e morali del rispetto legislazione internazionale vigente e/o da istituire da organizzazioni internazionali;
5. Certificazione di strumenti virtuali preventivi e obbligatori prima del lancio del software in rete secondo parametri stabiliti, nonché normative internazionali ONU;
6. Rafforzare la responsabilità sociale delle *Big Tech* e dei laboratori informatici, poiché riconoscono l'impatto della tecnologia a livello socio-economico. Obbligo dei laboratori informatici di stabilire strategie per mitigare gli effetti collaterali dei loro strumenti, nonché obbligo di stabilire strategie di inclusione e sviluppo digitale nei paesi bisognosi di tecnologie.

N.2 MAI 2023
ISSN 2789-8040

RIFERIMENTI

(01)Pause Giant AI Experiments: An Open Letter, march 22, 2023. Disponibile: https://futureoflife.org/open-letter/pause-giant-ai-experiments/?ref=nucleo.jor.br

(02)CCW Convention on Prohibitions or Restrictions on the Use of Certain Conventional Weapons Which May Be Deemed to Be Excessively Injurious or to Have Indiscriminate Effects, Group of Governmental Experts on Emerging Technologies in the Area of Lethal Autonomous Weapons System Geneva, 6-10 March, and 15-19 May 2023. Disponivel: https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-Group_of_Governmental_Experts_on_Lethal_Autonomous_Weapons_Systems_(2023)/CCW_GGE1_2023_WP.3_REv.1_0.pdf

(03)Argentina, Ecuador, El Salvador, Colombia, Costa Rica, Guatemala, Kazakhstan, Nigeria, Palestine, Panama, Peru, Philippines, Sierra Leone e Uruguay.

(04)HABERMAS, J., O future do natureza humana: a caminho de uma Eugenia liberal?,Traduçao de Karina Jannini, Sao Paulo: Martins fontes, 2004.

---

[i] Dottora in Antropologia presso L'Università degli Studi di Padova (Italia) e dottora in Economia presso L' Università di Alicante (Spagna). Ricercatrice Associata InterAgency Institute. Mail: juliana.martins@interagency.institute

N.2 MAI 2023
ISSN 2789-8040