Etica e intelligenza artificiale, cosa chiede l'Europa con le linee guida

Parliamo di etica e intelligenza artificiale. L'8 aprile 2019 è stata pubblicata la versione finale del documento "*Ethics guidelines for trustworthy AI*" la cui bozza era stata presentata il 18 dicembre 2018. Il documento è stato redatto da un gruppo di esperti (High-Level Expert Group on Artificial Intelligence) nominato dalla **Commissione Europea.**

Una volta emanata la bozza si è passati alla seconda fase dell'iter di pubblicazione della versione definitiva.

In questa fase la Commissione Europea ha coinvolto ed invitato una serie di soggetti accreditati, a livello europeo, ad inviare i propri contributi tecnici al fine di integrare e migliorare la bozza.

Sono stati così vagliati e controllati più di 500 commenti, suggerimenti e integrazioni che hanno arricchito la versione ufficiale. Quindi, l'attuale versione è il frutto di un impegnativo lavoro di analisi, approfondimento e integrazione di questi contributi.

Lo scopo delle **linee guida** è quello di promuovere l'implementazione di una Intelligenza Artificiale affidabile. Per essere considerata tale sono tre le componenti che deve possedere:

- deve essere conforme alla legislazione,
- deve seguire i principi e i valori etici,
- deve essere ben strutturata sia da un punto di vista tecnico che da quello sociale poiché, anche con le migliori intenzioni, un sistema di IA potrebbe causare effetti indesiderati.

Ciascun componente sopra citato è necessario ma non sufficiente per raggiungere l'affidabilità ricercata. Idealmente, i tre componenti dovrebbero lavorare in modo armonico e integrarsi nel loro funzionamento. Se a livello pratico dovessero sorgere tensioni fra questi elementi, sarà la stessa società ad adoperarsi e ad attivarsi per risolvere tale tensione.

Etica e intelligenza artificiale in tre capitoli

Analizzeremo in dettaglio il documento che si compone di tre capitoli.

Bisogna comunque tenere in considerazione che la guida in questione è basata su **tre livelli di astrazione**, dal più astratto, cioè il capitolo I in cui si parla di diritti, principi e valori fondamentali, al più concreto, ovvero il capitolo III dove viene presentata una lista di controllo per la valutazione. Nello specifico, il capitolo I si riferisce alle modalità per garantire la finalità etica dell'IA, definendo i diritti, i principi e i valori fondamentali che dovrebbe rispettare. Partendo da tali principi, nel capitolo II viene elaborata una guida relativa alla realizzazione di un'IA affidabile, tenendo conto sia della finalità etica che della robustezza tecnica.

A tal fine vengono elencati i requisiti per un'IA corretta e si fornisce una panoramica dei metodi tecnici e non tecnici che possono essere utilizzati per la sua implementazione. Il capitolo III, infine, rende operativi tali requisiti, fornendo una lista di controllo concreta ma non esaustiva per la valutazione dell'affidabilità dell'IA. Tale lista viene successivamente adattata a casi d'uso specifici.

A differenza di altri documenti che riguardano l'etica dell'AI, i presenti orientamenti non mirano a elencare ancora una volta i valori e i principi fondamentali per l'IA, ma hanno piuttosto lo scopo di **fornire una guida** su come attuare e rendere concretamente operativi tali principi nei sistemi di IA.

Per approfondire:

Il contributo di Aixia al codice etico europeo
https://bit.ly/2BimjKC

Cosa ha espresso il primo forum italiano sull'intelligenza artificiale
https://bit.ly/2Jbtdr1

Capitolo 1

Basandosi sui diritti fondamentali dell'uomo, il primo capitolo identifica i principi etici e i corrispondenti valori che devono essere rispettati nello sviluppare, fornire e usare i sistemi di IA.

Le indicazioni chiave che vengono descritte e fornite in questo capitolo riguardano:

1. il rispetto per l'autonomia umana, prevenzione del danno, equità e chiarezza. È necessario prestare particolare attenzione alle situazioni che coinvolgono i gruppi più vulnerabili come bambini, persone con disabilità e altri che storicamente sono stati svantaggiati o sono a rischio di esclusione. Anche altre relazioni sono da tenere in considerazione come quelle caratterizzate da asimmetrie di potere o di informazione, si pensi ai rapporti tra datori di lavoro e lavoratori o tra imprese e consumatori.
2. Il riconoscimento che, pur apportando sostanziali benefici agli individui e alla società, i sistemi di Intelligenza Artificiale possono comunque causare anche alcuni rischi e avere così impatti negativi, difficili da prevedere, identificare o misurare (ad esempio sulla democrazia, lo stato di diritto e la giustizia distributiva, o sulla stessa mente umana).
3. L'adozione di misure adeguate per attenuare questi rischi quando appropriato e proporzionalmente all'entità del rischio.

Capitolo 2

Basandosi sulle considerazioni emerse nel capitolo 1, il capitolo 2 fornisce una linea guida di come i sistemi di IA dovrebbero essere implementati aderendo ai seguenti 7 requisiti:

1. supervisione umana,
2. solidità tecnica e sicurezza,
3. privacy e governance dei dati,
4. trasparenza,
5. diversità, non discriminazione ed equità,
6. benessere ambientale e sociale,
7. responsabilità.

Tuttavia, potrebbero sorgere delle significative contrapposizioni e tensioni tra i diversi principi e requisiti. Per questo risulta fondamentale identificare, valutare, documentare e far comunicare continuamente questi trade-off e le loro soluzioni.

Capitolo 3

Tale articolo fornisce una lista concreta, anche se non esaustiva di punti da considerare per implementare i requisiti evidenziati nel capitolo precedente.

Le indicazioni chiave che emergono dal capitolo 3 sono le seguenti:

1. È necessario adottare un elenco per la valutazione dei sistemi di AI, affinché questi ultimi risultino attendibili. Tale valutazione è da effettuarsi durante lo sviluppo, l'implementazione o l'utilizzo e deve essere adattata all'uso specifico che si sta facendo della tecnologia AI in esame
2. L'elenco di valutazione fornito non sarà mai esaustivo. Infatti, per poter garantire un AI affidabile è necessario un processo continuativo che prevede: l'identificazione e l'implementazione dei requisiti, la valutazione delle soluzioni, la garanzia di migliorare i risultati nell'intero ciclo di vita del sistema AI e il coinvolgimento di tutti i soggetti interessati.

Una sezione finale del documento mira a rendere ancora più concrete alcune delle questioni affrontate nel documento. Vengono quindi forniti esempi nei quali l'impiego di soluzioni AI potrebbero condurre a impatti positivi e a benefici, così come vengono anche illustrati criticità ed impatti negativi che i sistemi di AI, se implementati, potrebbero causare e che dovrebbero essere attentamente considerati.

Tra i **benefici** che si potrebbero trarre dall'applicazione di intelligenza artificiale vengono citati: la riduzione dell'inquinamento con conseguente attenuamento del riscaldamento globale, l'ottimizzazione dell'utilizzo delle risorse energetiche e delle risorse in generale (si pensi all'acqua). Vengono anche citati i possibili utilizzi dell'AI volti a migliorare la vita della popolazione mondiale con particolare riferimento agli anziani, ai quali si potrebbero fornire strumenti di supporto per un invecchiamento sano.

E ancora: sistemi in grado di migliorare l'istruzione e la cultura, supportare i cittadini nel processo di trasformazione in atto rendendoli soggetti attivi e non oggetti passivi.

Mentre tra le **criticità** evidenziate troviamo: l'identificazione di comportamenti di singoli soggetti con finalità discriminatorie, la possibilità di sfruttare sistemi di IA mascherandone l'identità digitale (ovvero facendole apparire come un essere umano), la misurazione delle facoltà delle persone – effettuata a loro insaputa – per ricavarne tendenze e comportamenti prevedibili a livello del singolo, la creazione e l'utilizzo di armi autonome, e molto altro ancora che, oggi, non siamo nemmeno in grado di immaginare.

Etica e opportunità

Quindi, il documento fornisce una **visione estesa delle opportunità** che potremmo cogliere e delle criticità che dovremo affrontare a seguito dell'implementazione di sistemi di Intelligenza Artificiale. Ciò nonostante deve essere considerato come un punto di inizio e non come uno strumento esaustivo, la cui adozione può risultare risolutiva per indirizzare l'IA verso il bene dell'umanità e del pianeta.

Associazione Italiana per l'Intelligenza Artificiale

La roadmap europea: un'interpretazione

Prendendo atto di questo fatto, la Commissione Europea indica una **roadmap** di adozione costituita da più passaggi:

- Le **aziende** e gli enti pubblici interessati a contribuire al processo possono registrarsi nel sito della comunità europea e proporsi come candidati per eseguire delle attività pilota.
- I **feedback** di queste attività costituiranno la base per far evolvere le linee guida ed in particolare gli strumenti che consentano di aderire alle raccomandazioni nel 2020.

In tema di etica, pur riconoscendo la validità e la buona impostazione e realizzazione di tale documento, **dobbiamo chiederci se un framework di questo tipo ha realmente la capacità di essere significativo** e di essere adottato su vasta scala così da ottenere una IA volta al benessere delle persone che riduca, contemporaneamente, i rischi.

Per quanto le indicazioni siano opportunamente coniugate, **dobbiamo chiederci quanto queste raccomandazioni siano compatibili con la società moderna** globalmente estesa e quanto, in realtà, alcune di queste non entrino in rotta di collisione con i sistemi economici, politici e di convivenza con l'ambiente rappresentativi di quell'epoca geologica che il biologo Eugene F. Stoermer ha denominato "Antropocene". Ovvero l'epoca in cui l'Homo Sapiens ha un forte impatto sull'equilibrio del pianeta.

Come ormai molti economisti e studiosi di varie discipline denunciano, se il nostro desiderio ultimo è l'aumento del profitto e la crescita del PIL, non è detto che il miglioramento di questi indicatori porterà al benessere.

Se continuiamo ad adottare modelli economici di due secoli fa, in cui capitale e lavoro sono le risorse scarse, mentre l'ambiente viene considerato immutabile non riusciremo a conseguire quegli aspetti di etica che il documento correttamente evidenzia.

I sistemi di Intelligenza artificiale costituiscono **strumenti potenti**, capaci di esaudire i nostri desideri, dobbiamo perciò **porre attenzione a come li esprimiamo** e soprattutto a non esprimere quelli sbagliati**.**

Chi è l'autore

***Piero Poccianti** è presidente dell'Associazione Italiana per l'Intelligenza Artificiale (AIxIA), associazione scientifica senza fini di lucro fondata nel 1988 con lo scopo di promuovere la ricerca e la diffusione delle tecniche proprie dell'intelligenza artificiale.*
Considerato l'organismo accademico più autorevole in Italia nel campo dell'intelligenza artificiale conta oggi più di 1.000 membri tra ricercatori, professori universitari e società.
Si focalizza nella divulgazione dell'intelligenza artificiale intesa come formazione, ricerca e sviluppo di nuovi paradigmi e tecnologia a favore delle imprese e del Sistema Paese.