Quaderni della ReD OPEN FACTORY

Valentina Cavosi

GOVERNARE L'INTELLIGENZA ARTIFICIALE

SPUNTI PER LA PROGETTAZIONE DI SISTEMI DI INTELLIGENZA ARTIFICIALE LEGALI, ETICI E ROBUSTI

Introduzione di Federico Cabitza

Postfazione di Donatella Paschina

Ledizioni

2022 Ledizioni LediPublishing
Via Boselli 10, 20136 Milano
http://www.ledizioni.it
e-mail: info@ledizioni.it

Prima edizione Ledizioni: febbraio 2022

Valentina Cavosi, *Governare l'intelligenza artificiale. Spunti per la progettazione di sistemi di intelligenza artificiale legali, etici e robusti*

ISBN cartaceo 978-88-5526-622-2

In copertina: photo by Sander Weeteling on unsplash.com

Informazioni sul catalogo e sulle ristampe: www.ledizioni.it

Indice

Introduzione

Prof. Federico Cabitza
Università di Milano Bicocca
Comitato Scientifico ReD OPEN Factory

Assistiamo ad una crescente diffusione di sistemi digitali in grado di elaborare grandi quantità di dati, anche in formato non strutturato, e di applicare ad essi regole che nessun essere umano ha scritto direttamente, ma che sono state prodotte mediante tecniche e metodiche di "apprendimento automatico" (*machine learning*), per generare "contenuti, previsioni, raccomandazioni o decisioni che influenzano l'ambiente con cui tali sistemi interagiscono"[1], inclusi i suoi utenti. Per queste loro caratteristiche, sempre più frequentemente si assegna a queste tecnologie digitali l'etichetta di sistemi in grado di esibire o esprimere una certa "Intelligenza Artificiale", cioè la capacità di eseguire autonomamente compiti complessi che si pensa richiedano un certo grado (e un certo tipo) di intelligenza agli esseri umani per eseguirli correttamente.

Non è un caso che abbiamo introdotto questa locuzione – Intelligenza Artificiale – con cautela e in

[1] Proposta di regolamento del parlamento europeo e del consiglio che stabilisce regole armonizzate sull'intelligenza artificiale (legge sull'intelligenza artificiale) 2021/0106(COD). https://eur-lex.europa.eu/legal-content/IT/TXT/HTML/?uri=CELEX:52021PC0206&from=EN

maniera così perifrastica; infatti, temiamo che l'espressione stessa, in virtù della sua popolarità e forza mediatica (forse eccessive di questi tempi), possa distogliere l'attenzione da una semplice considerazione.

Quando si parla della applicazione dell'Intelligenza Artificiale in contesti organizzativi, non si fa riferimento ad altro che non sia l'automazione di operazioni che possono avere un impatto rilevante in processi aziendali dove siano elaborati molti dati, personali e non (i cosiddetti "big data" se vogliamo usare un'altra espressione altrettanto vaga e parimenti abusata), e in cui alcuni esseri umani sono chiamati a prendere decisioni che in molti casi riguardano altri esseri umani (ad esempio fornitori, clienti, impiegati, dirigenti), oppure che riguardano indirettamente un più ampio novero di possibili portatori di interesse.

I sistemi che chiamiamo Intelligenza Artificiale non riguardano perciò solo funzionalità particolarmente avanzate, sorprendenti o più tipiche di scenari fantascientifici, ma anche contesti aziendali e sociali che conosciamo e frequentiamo tutti i giorni.

In questo volume parleremo di come valutare questi sistemi: lo faremo partendo dall'estensione e diffusione di un lavoro che è stato oggetto di una tesi di laurea per il corso di laurea magistrale in Teoria e Tecnologia della Comunicazione dell'Università degli Studi di Milano-Bicocca, che ho personalmente seguito e che rappresenta un punto d'avvio per approfondimenti e implementazioni sul campo.

Riteniamo quindi opportuno riflettere sull'importanza dei processi di valutazione di qualsiasi tecnologia

che abbia un potenziale impatto sugli esseri umani, e di conseguenza anche – e soprattutto – sull'importanza di valutare i rischi connessi alle tecnologie informatiche che abilitano o potenziano processi amministrativi, analitici o decisionali al centro della catena del valore delle aziende.

Questi processi di valutazione della tecnologia informatica sono infatti necessari per comprendere diverse cose:

- se una tale tecnologia avrà un impatto sull'organizzazione che l'adotta per l'automazione, parziale o totale, di certi processi;
- se il suo impatto sarà positivo, cioè se i vantaggi e gli aspetti positivi saranno maggiori degli svantaggi che con maggiore o minore probabilità si concretizzeranno in seguito alla sua adozione; e, infine,
- se l'eventuale impatto positivo sarà valso lo sforzo, cioè se l'adozione di questa tecnologia in esame sarà 'costo-efficace' e il ritorno sull'investimento, non solo economico, sarà significativo.

Questi tre aspetti costituiscono il nucleo di qualsiasi attività di valutazione della tecnologia (*technology assessment*), a prescindere dai metodi e dalle tecniche messe in campo per produrla; infatti, il primo passo di qualsiasi attività consulenziale per valutare l'impatto di una tecnologia informatica è la convinzione consapevole che nessuno degli aspetti che abbiamo citato debba essere dato per scontato, neppure per le tecnologie più avanzate, come quelle a cui associamo l'etichetta di "Intelligenza Artificiale". Ad esempio, una tecnologia potrebbe sì portare benefici all'azienda che la adotta,

ma risultare insostenibile, sia economicamente che socialmente (e, perché no, anche dal punto di vista dei consumi energetici e dell'impronta di carbonio); oppure essa potrebbe essere addirittura dannosa, portando più svantaggi che vantaggi, o svantaggi più rilevanti; oppure, più semplicemente, una tecnologia digitale potrebbe essere inutile, nel senso che, a conti fatti, non ha alcun impatto sul lavoro di tutti i giorni. Quest'ultimo scenario, anche se sembra il più scandaloso e improbabile in molti circoli consulenziali, è invece molto più comune in ambito aziendale di quanto si creda: lo riteniamo infatti molto frequente anche nell'attuale panorama delle applicazioni di Intelligenza Artificiale progettate per supportare gli ambiti professionali in cui se ne parla di più, quali la medicina, la gestione delle risorse umane, il giornalismo e la pratica legale.

Il lavoro che segue tratta ed esplora lo spazio di idee e concetti in cui si possono produrre e comprendere i risultati di una valutazione dell'impatto di sistemi di Intelligenza Artificiale, nella più ampia e concreta accezione menzionata sopra. Per questo motivo esso si rivolge a qualunque decisore, amministratore, o responsabile di processo che operi in contesti organizzativi, aziendali e istituzionali allo scopo di aiutarlo a comprendere alcuni termini chiave (come trasparenza, robustezza, affidabilità, o utilità) dal significato solo apparentemente comune, per scoprire la complessità e, purtroppo ambiguità, di certe definizioni; e per fornirgli un riferimento agile e accessibile per esercitare il governo responsabile e informato di processi decisionali informatizzati complessi, tra cui ovviamente

anche quelli che riguardano l'acquisto e adozione di un sistema software commercializzato sotto il nome di Intelligenza Artificiale (o certificato come tale, come in ambito sanitario). Infatti ogni sistema di Intelligenza Artificiale concepito e progettato per il contesto aziendale, non è tanto o solo uno strumento informatico come tanti altri di "burotica" e automazione di ufficio, al contrario è anche, e soprattutto, il motore e l'occasione per un più ampio cambiamento organizzativo che riguarda le modalità e priorità di produzione e gestione del dato – che in larga parte è personale e, come tale, richiede l'applicazione del regolamento del GDPR –, e le modalità in cui alcune importanti decisioni sono prese in azienda sulla base di quel dato.

Il perché occuparsi di questi temi dovrebbe essere chiaro. Le motivazioni che spingono verso crescenti livelli di informatizzazione di processi decisionali strategici ed operativi sono molteplici: la promessa di incrementi di efficienza (a parità di risultati o di risorse impiegate); una minore variabilità di criterio o esito (anche come conseguenza di una minore dipendenza dalla esperienza, arbitrio o pregiudizio del decisore umano); una maggiore efficacia (in termini o di minore tasso di errore o di maggiore qualità degli effetti della decisione); e infine, per limitarsi alle motivazioni principali, una maggiore ripetibilità e standardizzazione del processo, per garantire tanto una maggiore imparzialità per i soggetti coinvolti quanto una maggiore resilienza dell'azienda nei confronti di fenomeni di rotazione o avvicendamento delle sue risorse umane. Sono cose note; ma a queste considerazioni è giunto il

tempo di aggiungere la consapevolezza degli elementi di fragilità e rischio inerenti alle tecnologie più avanzate, quali la loro vulnerabilità ad attacchi dall'esterno a scopo estorsivo (cyberattacchi o *adversarial attacks*); la dipendenza da quantità di dati la cui qualità e rappresentatività, anche per la loro dimensione e le fonti di dati disponibili, è difficile da verificare e garantire; la loro capacità di esacerbare disuguaglianze o perpetrare ingiustizie nei confronti di minoranze etniche, di genere o, peggio, di accesso alle risorse (ad esempio, per età, istruzione, o capacità economiche). A questa consapevolezza si deve affiancare una nuova sensibilità per la valutazione dell'impatto di iniziative di informatizzazione della decisione nei termini della sostenibilità sociale (relativa alle trasformazioni dell'occupazione della forza lavoro) e umana (che riguarda il rischio che l'automazione impoverisca l'*expertise* delle risorse umane e i processi di acquisizione di nuove conoscenze, o ne causi una lenta ma progressiva perdita di competenze o deresponsabilizzazione).

A differenza dei proclami che si sentono spesso levare dall'industria e da alcuni operatori del mercato, l'era delle macchine intelligenti non è ancora arrivata; o almeno non si è ancora concretizzata l'età in cui le macchine sono in grado di sostituire o influenzare, con autorità e ingombrante autorevolezza, i processi decisionali più delicati che caratterizzano le nostre aziende e organizzazioni. Anziché essere una notizia deprimente o demotivante per il lettore che si accinge ad affrontare le prossime pagine, questa è un'ottima cosa: infatti siamo ancora in tempo per acquisire quella cul-

tura di innovazione, o sarebbe meglio dire, progresso, che ci permetterà di giudicare con equilibrio e competenza l'opportunità di digitalizzare alcuni processi e di capire, con tempestività e adeguatezza, quali siano le modalità migliori per farlo con trasparenza, equità, rispetto del principio di legalità, e responsabilità.

Ritengo che il trasferimento tecnologico di risultati selezionati che derivano da studi e ricerche condotte in ambito accademico sia un'operazione opportuna e necessaria, soprattutto in questa fase di grande rinnovamento degli strumenti digitali che si rendono a disposizione del lavoro di analisi e dei processi decisionali in ambito organizzativo: il volume si situa in questo filone di attività, che segue e che rafforza la collaborazione e sinergia tra il mio centro di ricerca e la ReD Open Factory.

Guida alla lettura

Lo spirito che anima il lavoro che segue è quello di rendere disponibili anche al lettore non esperto dei materiali per facilitare la conoscenza e l'approfondimento di un tema che sta tuttora alimentando discussioni e conversazioni nei contesti giuridici, governativi e industriali sia in Italia che all'estero: l'Intelligenza Artificiale e il suo uso.

Tale tema, in continua evoluzione e sviluppo, influenza anche l'iter normativo, tuttora in via di definizione; proprio per questo motivo il lavoro è utile per chi vuole comprendere i principi fondamentali e di conseguenza poter seguire e approfondire quanto sta accadendo a livello geopolitico.

L'introduzione del prof. Federico Cabitza, e la postfazione di Donatella Paschina, sono gli ulteriori contributi che introducono e allargano il contesto di riflessione sul tema di questo lavoro.

Infatti, al di là del percorso normativo e giuridico di una futura normativa sull'uso di tecniche di cosiddetta "Intelligenza Artificiale", è fondamentale che si abbia consapevolezza dell'impatto che queste evolute tecniche di automazione potranno avere sulle vite quotidiane di tutti noi.

Oggigiorno, inoltre, esse arricchiscono la trasformazione digitale in corso da anni, con la conseguenza che, in aggiunta ai principi di riservatezza e privacy dell'individuo, si aggiunge il tema del governo delle tecniche di decisioni che impattano sulla vita delle persone e

dell'ambiente circostante.

Quello che segue è perciò un contributo specialistico per orientarsi sul tema, il cui aggiornamento è in corso all'interno di ReD OPEN Factory, il "CENTER FOR RESPONSIBLE INNOVATION" per la governance della trasformazione digitale.

Buona lettura.

Per ulteriori approfondimenti: www.redopenfactory.com