Audizione nell'ambito dell'indagine conoscitiva sull'Intelligenza Artificiale: opportunità e rischi per il sistema produttivo italiano

Alessandro Nuara

ML cube S.r.l.

15/11/23

Egregio Presidente, Egregi Onorevoli,

Desideriamo iniziare questa relazione esprimendo la nostra più sentita gratitudine per averci concesso l'opportunità di contribuire a questa discussione fondamentale riguardante l'Intelligenza Artificiale e il suo impatto sui settori strategici del nostro Paese.
La nostra esperienza nella ricerca scientifica nell'ambito dell'Intelligenza Artificiale presso il laboratorio AI del Politecnico di Milano, e di sviluppo di sistemi AI in ML cube e AD cube ci ha permesso di esplorare a fondo il vasto potenziale dell'AI in diversi settori, nonché le sfide connesse all'introduzione dell'AI nei processi aziendali fondamentali. ML cube, spin-off del Politecnico di Milano, fornisce soluzioni di AI per il decision making in diversi settori, offrendo anche tool per la governance dei modelli di Machine Learning. AD cube si specializza nello sviluppo di sistemi automatici basati sull'Intelligenza Artificiale specifici per il settore del marketing online.
Gli spin-off universitari costituiscono, a nostro avviso, un veicolo imprescindibile per l'innovazione AI Made in Italy. Questi, da un lato sono in grado di portare sul mercato le soluzioni più innovative sviluppate nell'ambito della ricerca universitaria, e dall'altro, intercettano le esigenze e le traiettorie del mercato e per poi indirizzare la ricerca nella giusta direzione.
Questo ruolo chiave, ci offre un punto di vista privilegiato nella comprensione delle principali opportunità e sfide legate all'introduzione delle soluzioni innovative di Intelligenza Artificiale.

Premessa

L'Intelligenza Artificiale è attualmente la più rivoluzionaria innovazione tecnologica del nostro secolo. La sua pervasività in tutti i settori economici promette una trasformazione dell'economia, del mercato del lavoro e della società su scala globale. I vantaggi derivanti dall'implementazione dell'AI sono diversi e nelle aziende si traducono in un significativo aumento della produttività, nell'efficacia e nell'aumento di performance nell'esecuzione di molteplici task. Secondo un recente rapporto di McKinsey ("*The economic potential of generative AI: The next productivity frontier" - 2023*), l'adozione su larga scala dell'AI sbloccherà un valore economico che dai 2,2 ai 4,6 trilioni di dollari, distribuiti tra vari settori, e con un impatto trasversale su diverse aree delle aziende (in ordine Marketing e Sales, Customer Operations, Product R&D, Software Engineering, Supply Chain and Operations, Risk and legal, Strategy and Finance, Corporate IT, Talent organization).
Questi dati evidenziano chiaramente come tutti gli attori del mercato, tra cui le PMI, le grandi aziende e la Pubblica Amministrazione, avranno l'opportunità di sviluppare e consolidare la propria presenza sul mercato a livello nazionale e internazionale. Al contrario, coloro che non saranno in grado di intraprendere un percorso di trasformazione AI rischiano di ampliare in modo forse irreparabile il divario con coloro che sfrutteranno questa innovazione.
È essenziale, dunque, che il nostro Paese sviluppi strategie atte a garantire che tutti gli attori traggano vantaggio da questa innovazione, mitigando al contempo i potenziali rischi ad essa associati.

Problematiche legate all'adozione dell'AI

Fino a qualche anno fa, l'Intelligenza Artificiale era considerata principalmente una disciplina ristretta alle competenze tecniche all'interno delle aziende. Oggi, è chiaro che l'AI si estende trasversalmente a diversi processi e coinvolge una vasta gamma di figure professionali all'interno delle aziende.
L'AI non è più semplicemente un ambito riservato agli specialisti tecnici, ma piuttosto un elemento centrale che influisce su molteplici aspetti aziendali. Coinvolge attivamente professionisti in diversi settori, tra cui manager, analisti dei dati, esperti di sicurezza informatica, responsabili della privacy, dirigenti del settore legale e molti altri.
Oltre alle sfide tecniche legate alla progettazione e alla creazione dei modelli di Intelligenza Artificiale, infatti, è fondamentale considerare le questioni legate alla **trasparenza, alla fiducia, alla sicurezza e alla governance nell'implementazione dell'AI.**

Considerato che queste tecnologie saranno fondamentali nel prossimo futuro, la disciplina non può che essere un aiuto essenziale per supportare tutte le organizzazioni, siano essere pubbliche che private, a raggiungere meglio (e forse prima) proprio il domani in questione.
Considerato che Iniziative governative e private a livello mondiale vanno tutte in questa direzione, è ragionevole prevedere che entro il 2026 **le organizzazioni che renderanno operativi temi di trasparenza, fiducia e sicurezza verso l'AI vedranno, per i propri modelli di intelligenza artificiale, ottenere un miglioramento dei risultati del 50% in termini di adozione, obiettivi di business e accettazione da parte degli utenti.** Allo stesso modo, Gartner prevede che entro il 2028 le macchine basate sull'Intelligenza Artificiale rappresenteranno il 20% della forza lavoro globale e il 40% di tutta la produttività economica. Non è difficile immaginare, dunque, che implementare strategie per garantire trasparenza, fiducia e sicurezza sarà un requisito indispensabile per competere nel mercato. Le organizzazioni che non adotteranno misure più incisive per gestire rischi e problematiche, infatti, potrebbero andare incontro a gravi danni alla reputazione, a importanti perdite economiche e malfunzionamenti dei sistemi AI.

Framework AI TRiSM

Nel tentativo di fornire una migliore comprensione dell'emergente ecosistema e dei temi da affrontare, degno di nota è certamente il framework creato da Gartner denominato **AI TRiSM (Artificial Intelligence Trust, Risk, and Security Management),** progettato per garantire una governance robusta e la protezione della privacy nell'utilizzo dell'AI. Un framework da noi adottato nella sua totalità in quanto perfettamente appropriato alle sfide che già affrontiamo in qualità di fornitori di soluzioni AI e, soprattutto, che sappiamo avere una traiettoria futura sempre più rilevante e complessa. Questo framework mira a implementare politiche di salvaguardia e di governance volte a prevenire l'abuso e l'uso inappropriato dell'Intelligenza Artificiale.
L'adozione di approcci come AI TRiSM può essere cruciale per una gestione più efficace dei rischi, consentendo un'implementazione più sicura, trasparente ed etica dell'AI all'interno delle organizzazioni, mitigando al contempo le potenziali violazioni normative e di sicurezza.

Tre gli ambiti sui quali finalizzare il focus:

- **AI Trust**
- **AI Risk**
- **AI Security Management**

e quattro i pilastri necessari a disegnare, modellare e costruire efficaci soluzioni di intelligenza artificiale:

- **Explainability/model monitoring**
- **Privacy**
- **ModelOps**
- **AI application security**

Il framework si presta bene a promuovere e condividere una disciplina (a oggi in rapido sviluppo e in costante evoluzione) utile a supportare un'adozione corretta dell'AI, sempre più diffusa anche in settori critici.

Explainability/model monitoring
Nel contesto dell'Intelligenza Artificiale, due aspetti fondamentali sono la spiegabilità dei modelli e il monitoraggio dell'AI in produzione. Questi elementi sono strettamente legati e lavorano insieme per per garantire che i modelli di IA funzionino in modo attendibile e coerente con le aspettative degli utenti e degli sviluppatori.

- **Spiegabilità dell'AI:** Questo aspetto si concentra sulla creazione di capacità e strumenti che consentono di comprendere e chiarire il funzionamento interno di un modello di AI. Questo processo di spiegabilità mira a:
    - **Descrivere il modello:** Fornire una descrizione dettagliata delle caratteristiche e del funzionamento del modello.
    - **Evidenziare punti di forza e debolezza:** Identificare le aree in cui il modello si dimostra efficace e quelle in cui potrebbe avere delle limitazioni.
    - **Prevedere il comportamento del modello:** Analizzare e prevedere il comportamento futuro del modello.
    - **Rilevare potenziali bias:** Identificare eventuali distorsioni o parzialità presenti nel modello.
    - **Chiarire il funzionamento del modello:** Consentire una comprensione accurata, equa e trasparente del modello, facilitando decisioni algoritmiche affidabili e responsabili.

  L'obiettivo principale della spiegabilità dell'AI è fornire una visione chiara e comprensibile del funzionamento del modello, creando fiducia negli utenti e consentendo la sua corretta interpretazione e utilizzo.
- **Monitoraggio dell'AI:** Una volta che il modello di AI è implementato in un ambiente operativo, diventa cruciale monitorare continuamente le sue prestazioni. Questo processo di monitoraggio serve a:
    - **Verificare l'accuratezza del modello:** Controllare e valutare l'accuratezza delle previsioni del modello.
    - **Rilevare anomalie o cambiamenti nei dati di produzione:** Individuare possibili deviazioni o variazioni nei dati di ingresso o nell'output del modello.
    - **Prevenire distorsioni o perdite di dati:** Identificare e correggere eventuali distorsioni o perdite di dati che potrebbero compromettere l'affidabilità del modello.
    - **Proteggere da potenziali attacchi:** Monitorare il modello per individuare possibili minacce o tentativi di attacco, proteggendo così l'AI da utilizzi malevoli o manipolazioni.

Il monitoraggio dell'AI mira a garantire la continuità delle prestazioni ottimali, la sicurezza dei dati e la protezione del modello da possibili minacce esterne.
Integrare questi due aspetti, spiegabilità e monitoraggio, rappresenta una tappa fondamentale nell'implementazione e nell'ottimizzazione dei modelli di Intelligenza Artificiale. La spiegabilità offre una chiara comprensione del modello, mentre il monitoraggio assicura la sua integrità e il mantenimento delle prestazioni ottimali nel tempo.

Privacy
Nel contesto dei progetti legati all'Intelligenza Artificiale, l'accesso ai dati, specialmente quelli personali, rappresenta un primo ostacolo significativo. La limitazione o il divieto di utilizzo di tali dati, sia nell'addestramento sia nella produzione dei modelli, è spesso dettato dai rischi legati alla privacy e alle possibili preoccupazioni di conformità. Secondo l'indagine 2021 sull'AI nelle organizzazioni condotta da Gartner, ben due organizzazioni su cinque hanno subito violazioni della privacy o incidenti di sicurezza legati all'AI. Inoltre, l'ampio panorama legislativo attuale mira a minimizzare tali violazioni. Le leggi sulla protezione dei dati si stanno sviluppando a livello mondiale, accompagnate da ulteriori disposizioni sulla governance dei dati e sull'utilizzo dell'AI, come riportato nello "Stato della Privacy - Unione Europea".
Parallelamente, le autorità (come, ad esempio, la CNIL – autorità di vigilanza francese) iniziano a pubblicare linee guida sull'uso dei dati personali nell'ambito dell'AI sottolineando la necessità di ridurre al minimo l'uso di dati identificabili. In fasi cruciali come l'ideazione e la fase iniziale dell'addestramento, l'utilizzo di dati personali non solo è fortemente sconsigliato, ma può addirittura configurarsi come illegale. Pertanto, al fine di evitare problemi di conformità nella fase di produzione, è raccomandato di evitare l'uso di dati identificabili.
Un approccio comune per affrontare le problematiche legate alla privacy consiste **nell'utilizzo di dati sintetici anziché informazioni identificabili**, specialmente nelle fasi iniziali come l'addestramento preliminare del modello. Tuttavia, non tutti i tipi di dati sintetici offrono le stesse solide garanzie per evitare il rischio di riconoscimento, soprattutto nel contesto dell'addestramento dei modelli di AI.
Altre soluzioni innovative che proteggono i dati durante l'utilizzo includono (combinazioni di) cifratura omomorfica e secure multiparty computing (sMPC). Collettivamente, tali tecniche sono denominate PET. Tuttavia, non tutti i PET sono immediatamente applicabili al contesto dell'AI e non tutti i fornitori di tecnologie per migliorare la privacy si concentrano su approcci specialistici nell'ambito dell'AI.
La scelta delle misure di mitigazione dipende da vari fattori, come la propensione al rischio dell'organizzazione, la robustezza tecnologica percepita e le specifiche caratteristiche del caso d'uso. Ad esempio, se i data scientist o altri stakeholder interni intendono utilizzare informazioni già disponibili in forma identificabile all'interno dell'organizzazione, l'approccio crittografico diretto potrebbe non essere appropriato. Tuttavia, la gestione delle chiavi univoche rimane valida in scenari distribuiti, garantendo che le entità contribuenti possano decriptare solo i propri dati.

ModelOps
ModelOps rappresenta un approccio dedicato alla supervisione e alla gestione completa del ciclo di vita dei modelli analitici, basati sull'Intelligenza Artificiale e decisionali. Questa gamma di modelli comprende soluzioni analitiche derivanti dal Machine Learning (ML), dai grafi di conoscenza, dalle regole, dall'ottimizzazione, dalla linguistica e dagli agenti. Attualmente, la maggior parte dei fornitori di ModelOps offre funzionalità che partono dalla distribuzione dei modelli e si estendono fino alla loro implementazione nella produzione. Nel prossimo futuro, questi stessi fornitori supporteranno l'intero ciclo di vita dei modelli, incluso il design e lo sviluppo.
Storicamente, i team di data science si sono focalizzati principalmente nello sviluppo di risorse di Intelligenza Artificiale. L'approccio frammentario nell'affrontare l'operatività a lungo termine di tali risorse ha impedito alle organizzazioni di realizzare appieno il valore dei loro investimenti in Intelligenza Artificiale e le ha esposte a rischi. Questa problematica si acuisce specialmente in settori ad alta regolamentazione, che richiedono un controllo più stringente rispetto ai tradizionali metodi di monitoraggio.

Si sta assistendo ad un inizio di adozione del ModelOps, che prende avvio con la distribuzione dei modelli. Questa pratica fornisce alle organizzazioni una fonte centrale di verità durante la distribuzione su larga scala all'interno dell'azienda. ModelOps si presenta come un approccio agnostico dal punto di vista tecnico e della piattaforma, fornendo una visione olistica del processo di Intelligenza Artificiale e agevolando la collaborazione tra diversi team aziendali. Da diverse indagini emerge che il ModelOps è considerato più come una capacità aziendale che come una competenza specifica dei team di data science, trovando applicazioni su larga scala e trasversali all'azienda.

AI application security

Rilevare e fermare gli attacchi all'Intelligenza Artificiale richiede nuovi framework e tecniche per testare, convalidare e migliorare la robustezza dei flussi di lavoro dell'AI. Gli attacchi cyber all'AI (sia interni che presenti in modelli di terze parti) possono provocare diversi tipi di danni e perdite organizzative, ad esempio di natura finanziaria, reputazionale o legati alla proprietà intellettuale, alle informazioni personali delle persone o a dati proprietari. Poiché i rischi e le conseguenze possono variare, i responsabili della sicurezza devono aggiungere controlli e pratiche specializzate al portfolio di misure di protezione dati e di sicurezza delle applicazioni che implementano per altri tipi di applicazioni.

La sicurezza delle applicazioni di Intelligenza Artificiale coinvolge pratiche specializzate, politiche, strumenti di rilevamento e prevenzione delle minacce progettati per proteggere dalle nuove minacce e dagli attacchi specializzati, come gli attacchi ai media digitali nelle applicazioni di computer vision.

La resistenza agli attacchi avversari è un elemento fondamentale della sicurezza delle applicazioni di Intelligenza Artificiale che comprende l'addestramento dei modelli affinché siano in grado di ignorare o rispondere in modo diverso agli input avversari durante lo sviluppo, il testing e continuamente una volta in produzione. In tal senso, si applicano tipicamente specifiche tecniche in grado di rafforzare i modelli affinché possano tollerare un certo livello di rumore e potenziali dati avversari senza che le prestazioni vengano eccessivamente compromesse.

Gli attacchi avversari modificano gli output degli algoritmi di Machine Learning a vantaggio dell'attaccante. Solitamente ciò avviene fornendo input maliziosi ("input avversari") o perturbazioni dei dati una volta che il modello è in produzione. Alcuni strumenti di resistenza agli attacchi avversari possono evidenziare la vulnerabilità dei modelli agli input e alle perturbazioni dannose e fornire input di addestramento per mitigarli.

Conclusioni

Le implementazioni di strategie di Intelligenza Artificiale comportano un notevole potenziale di crescita economica e progresso tecnologico per il nostro Paese. Tuttavia, affrontano anche sfide significative legate alla sicurezza, alla privacy e alla governance dei dati. È cruciale che il governo e le istituzioni supportino attivamente l'adozione di queste strategie attraverso iniziative di sensibilizzazione, educazione e sostegno finanziario. Queste misure devono mirare a garantire una corretta comprensione dei rischi e delle opportunità legate all'AI, nonché a fornire risorse e strumenti per sviluppare capacità specializzate e garantire la conformità alle normative vigenti.

Inoltre, è fondamentale che le organizzazioni collaborino con enti normativi e autorità preposte, integrando efficacemente standard e best practice nelle proprie operazioni. Questa collaborazione favorirà un ambiente di lavoro più sicuro e trasparente nell'utilizzo

dell'AI, mitigando rischi di non conformità, violazioni della privacy e vulnerabilità della sicurezza.
In sintesi, per garantire una transizione armoniosa verso un futuro guidato dall'AI, è essenziale un impegno comune tra settore pubblico e privato, lavorando insieme per creare un ecosistema in cui l'innovazione tecnologica sia guidata da normative chiare, pratiche sicure e valori etici. È solo attraverso questo impegno congiunto che possiamo massimizzare i benefici dell'AI mentre proteggiamo la società dalle possibili minacce e conseguenze negative.