lug 30, 2026

Appunti - AI ACT Modulo 1

Riepilogo

Il corso ha definito i rischi operativi dell'intelligenza artificiale e le necessarie politiche aziendali per l'uso consapevole.

Integrazione normativa dell'Intelligenza Artificiale
È stato avviato il percorso formativo obbligatorio sull'AI Act per evitare sanzioni fino a 35 milioni di euro. L'intelligenza artificiale non sostituisce l'uomo ma richiede supervisione esperta.

Rischi tecnici e allucinazioni
I modelli operano su basi probabilistiche e generano risposte convincenti ma spesso errate. L'uso di strumenti esterni comporta gravi pericoli per la sicurezza dei dati aziendali riservati.

Obbligo di supervisione umana
È stato stabilito l'obbligo inderogabile per ogni operatore di verificare personalmente l'accuratezza di qualsiasi output generato dall'intelligenza artificiale prima dell'utilizzo. La responsabilità legale resta esclusivamente umana.

Passaggi successivi

- Completare formazione: Seguire l'intero percorso formativo obbligatorio sull'uso dell'intelligenza artificiale.
- Condividere slide: Caricare le slide presentate durante la sessione all'interno del sistema e-learning per la consultazione.
- Segnalare strumenti: Comunicare l'utilizzo di eventuali strumenti AI esterni o dubbi al team IT per procedere con la regolarizzazione prevista dal regolamento.
- Eseguire censimento: Svolgere il censimento degli strumenti aziendali che suggeriscono, prevedono o generano contenuti come compito finale del corso.
- Verificare output e fonti: Verificare sempre le fonti e i dati forniti dall'intelligenza artificiale prima di considerarli validi. Leggere e validare ogni documento prodotto dal sistema.

Rispettare policy dati: Evitare l'inserimento di dati personali, riservati o operativi di sicurezza in sistemi di terze parti senza autorizzazione. Utilizzare esclusivamente dati anonimi per esperimenti con strumenti non autorizzati.

Interrogare Chat GPT: Interrogare il sistema di intelligenza artificiale per identificare Domenico Corchiola al fine di verificarne la veridicità delle risposte.

Verificare fonti: Risalire costantemente alla fonte originale dei riferimenti normativi o delle informazioni fornite per convalidare quanto prodotto dai modelli.

Valutare sistemi AI: Valutare le prestazioni dei sistemi basandosi sul numero di eventi reali intercettati e sul tasso di falsi allarmi, evitando di basarsi esclusivamente sull'accuratezza complessiva.

Dettagli

- **Obiettivo del corso e contesto normativo**: È stato avviato un percorso formativo obbligatorio in conformità con l'articolo 4 del regolamento dell'Unione Europea 2024/1689, noto come AI Act. L'obiettivo non è la memorizzazione di norme, ma la comprensione degli strumenti per evitare sanzioni che possono raggiungere i 35 milioni di euro.
- **Presenza pervasiva dell'Intelligenza Artificiale**: L'intelligenza artificiale è già integrata in numerosi strumenti aziendali quotidiani, come i filtri antispam, i correttori ortografici e diverse funzionalità dei software in uso. L'utilizzo di tali tecnologie è consentito e non è oggetto di rimprovero, purché avvenga in modo consapevole.
- **Natura probabilistica dell'Intelligenza Artificiale**: I modelli di intelligenza artificiale sono macchine che calcolano probabilità e non dispongono di una reale comprensione dei fatti. La plausibilità delle risposte non garantisce la verità, e il sistema può presentare informazioni errate con estrema sicurezza, citando persino regolamenti inesistenti.
- **Ruolo dell'Intelligenza Artificiale nel lavoro**: La tecnologia non è destinata a sostituire le persone, ma a migliorare l'efficienza dei professionisti che possiedono una specifica competenza di dominio. Lo strumento risulta utile solo se utilizzato da esperti in grado di valutarne criticamente l'output.
- **Struttura del percorso formativo**: Il corso è suddiviso in tre parti distinte: una panoramica sul funzionamento tecnico dei modelli, l'analisi del quadro normativo e legale, e infine una sezione pratica riguardante l'applicazione operativa, la formulazione delle richieste (prompting) e la verifica dei risultati.

- **Rischi legati ai sistemi esterni**: L'utilizzo di chatbot esterni come ChatGPT, Claude o Gemini comporta rischi significativi legati alla sicurezza dei dati, poiché i server di tali servizi risiedono spesso all'estero. È stato segnalato che dati inseriti in conversazioni con tali strumenti sono stati in passato indicizzati pubblicamente, rendendo necessaria estrema cautela nel trattamento di informazioni aziendali o sensibili.
- **Definizione di Intelligenza Artificiale**: Il regolamento definisce l'intelligenza artificiale come un sistema automatizzato capace di operare con livelli di autonomia variabili. Tali sistemi deducono output, quali previsioni, raccomandazioni o decisioni, a partire dagli input forniti, avendo la capacità di influenzare ambienti fisici o virtuali.
- **Differenza tra software tradizionale e sistemi di intelligenza artificiale**: A differenza del software tradizionale, che esegue regole fisse fornendo risultati prevedibili e replicabili, l'intelligenza artificiale opera su base probabilistica. Il software standard è deterministico, mentre l'intelligenza artificiale "deduce" le risposte basandosi sui parametri appresi.
- **Evoluzione tecnologica e accessibilità**: Sebbene i concetti base dell'intelligenza artificiale risalgano alle teorie di Alan Turing, la diffusione attuale è dovuta alla disponibilità di vasti dataset, alla potenza di calcolo moderna e all'accessibilità semplificata per l'utente finale.
- **Metodologia per il censimento degli strumenti aziendali**: Per identificare correttamente l'uso di intelligenza artificiale in azienda, è necessario chiedere se i sistemi in uso suggeriscono, prevedono, riconoscono, generano o classificano informazioni, superando la semplice domanda diretta sull'uso di chatbot.
- **Processo di addestramento dei modelli**: L'addestramento avviene attraverso cicli ripetitivi in cui il modello genera una risposta, viene misurato l'errore rispetto alla risposta corretta e vengono aggiustati i parametri interni. Il risultato finale è un grafo complesso di connessioni non interpretabile direttamente dall'essere umano.
- **Importanza della qualità dei dati**: I dati utilizzati per l'addestramento costituiscono la materia prima fondamentale. Presenze di squilibri o mancanze nei dati storici (bias) portano il modello ad apprendere prassi errate o a non saper gestire situazioni non presenti nel dataset originale.
- **Funzionamento dei modelli generativi**: I chatbot conversazionali basano il proprio funzionamento sulla previsione statistica del token (parola o frammento di parola) successivo. Il sistema calcola quale elemento sia più probabile in una sequenza, simulando un ragionamento che in realtà è un calcolo probabilistico.
- **Variabili che influenzano l'interazione**: La casualità impostata nel modello e la finestra di contesto (la memoria della conversazione) influenzano le risposte fornite.

Lo stesso quesito può generare risultati diversi in momenti differenti, rendendo la riproducibilità non garantita e sottolineando la necessità di verifica costante.

- **Funzionalità dell'intelligenza artificiale generativa**: Gli strumenti di intelligenza artificiale sono efficacemente applicabili a compiti di trasformazione testuale, come il riassunto di documenti, la riformulazione di testi in base al destinatario, la traduzione linguistica e la conversione tra diversi formati di file.
- **Utilizzo dell'intelligenza artificiale per la produzione di documenti**: L'intelligenza artificiale è uno strumento efficace per generare bozze di documenti, estratti di dati o schemi, ma i risultati devono essere considerati come bozze e non come prodotti finali. La responsabilità della verifica, della firma e dell'invio dei documenti ricade interamente sull'operatore umano.
- **Limiti nella verifica della verità**: Il sistema non possiede la capacità di distinguere tra verità e falsità. Tende a generare risposte convincenti anche quando le informazioni sono errate, operando come un calcolatore avanzato piuttosto che come un esperto consapevole della veridicità delle proprie affermazioni.
- **Rischi legati alle date di addestramento**: I modelli di intelligenza artificiale hanno una data di taglio dell'addestramento e non conoscono le informazioni o le normative emesse successivamente. È essenziale verificare sempre le fonti, i riferimenti normativi e i numeri, poiché il sistema può inventare dettagli plausibili ma inesistenti.
- **Necessità di fornire contesto e dati**: Il sistema funziona correttamente solo se viene alimentato con il materiale e le istruzioni pertinenti. In assenza di dati specifici forniti dall'utente, il modello tende a improvvisare mescolando informazioni generiche con dati inventati.
- **Inaffidabilità nei calcoli**: L'intelligenza artificiale non è una calcolatrice affidabile e non deve essere utilizzata per operazioni complesse o per la gestione di dati contabili come bilanci o calcoli di royalties. Ogni output numerico deve essere rigorosamente verificato.
- **Il concetto di "agenti" e automazione**: La terminologia relativa agli "agenti" viene spesso utilizzata in modo mediaticamente allarmistico. Si tratta di strumenti per l'automazione che non possono sostituire le persone esperte dotate di conoscenza tacita e sensibilità al rischio.
- **Rischi di sicurezza e privacy dei dati**: L'uso indiscriminato di strumenti di intelligenza artificiale comporta il rischio che i dati aziendali escano dall'infrastruttura locale verso server esterni, esponendo l'azienda a potenziali violazioni della privacy e sanzioni, incluse quelle previste dal GDPR.

- **Pericoli legati ai contenuti generati (Immagini, Voce, Video)**: La capacità di generare contenuti realistici comporta rischi elevati, inclusa la creazione di deepfake. Sistemi basati su intelligenza artificiale locale, progettati con criteri di "privacy by design", sono preferibili per compiti specifici, mentre la clonazione vocale e la manipolazione video rappresentano rischi significativi per la sicurezza e la frode.
- **Codice e vulnerabilità di sicurezza**: Lo sviluppo di software tramite intelligenza artificiale è rischioso se l'utente non possiede competenze di programmazione. Il codice generato può presentare falle di sicurezza non evidenti che mettono a rischio l'intera infrastruttura aziendale.
- **Regole per la produzione di codice**: È severamente vietato inserire codice generato dall'IA in produzione senza una rigorosa verifica di sicurezza. Le infrastrutture critiche sono soggette a vigilanza nazionale, e compromettere la sicurezza aziendale per un utilizzo non autorizzato di software è inaccettabile.
- **Insidie nelle tabelle e dati strutturati**: Le tabelle generate dall'IA possono contenere errori nascosti in mezzo a dati corretti. Se il volume di dati è elevato, identificare il singolo errore diventa difficile, rendendo necessaria una supervisione costante da parte di un esperto di dominio.
- **Modelli interni versus esterni**: Esistono compromessi tra l'uso di modelli esterni (spesso più potenti ma con rischi di riservatezza e costi elevati per implementazioni aziendali) e modelli interni (più controllabili ma potenzialmente meno avanzati). La scelta dipende dalla natura dei dati trattati.
- **Politiche di gestione dei dati**: È vietato immettere dati riservati, personali o operativi di sicurezza in sistemi di terze parti senza autorizzazione. Gli strumenti devono essere utilizzati solo in conformità con l'elenco autorizzato, preferendo dati anonimizzati o pubblici.
- **Natura delle allucinazioni**: Le allucinazioni non sono bug temporanei, ma caratteristiche strutturali intrinseche dell'intelligenza artificiale. I segnali tipici includono precisione eccessiva in ambiti di nicchia, link non funzionanti o coerenza sospetta su temi controversi.
- **Non ripetibilità e vulnerabilità alle istruzioni**: Le risposte del sistema possono variare anche per la stessa domanda e il comportamento può essere alterato da istruzioni nascoste nei documenti forniti. Questo conferma l'importanza di mantenere il controllo umano durante l'utilizzo.
- **Proprietà intellettuale e licenze**: La titolarità dei contenuti generati non è sempre garantita. Inoltre, il codice generato può incorporare frammenti soggetti a licenze incompatibili con l'uso commerciale o aziendale, richiedendo verifiche attente.

- **Obiettivo delle norme aziendali**: Le restrizioni sull'uso dell'IA non hanno scopo punitivo ma di salvaguardia. L'obiettivo è garantire la sicurezza aziendale e personale attraverso processi tracciabili e verificabili.
- **Metriche di valutazione del sistema**: L'accuratezza dichiarata (es. 99%) è spesso una metrica fuorviante. È necessario valutare il sistema in base alla capacità di individuare eventi reali (mancate rilevazioni) e al numero di falsi allarmi prodotti, gestendo il compromesso tra questi due errori.
- **Domande chiave per i fornitori**: In fase di valutazione di un sistema, è necessario chiedere: su quali dati è stato addestrato, in quali casi fallisce, cosa accade quando incontra situazioni inedite e come influiscono gli aggiornamenti del modello.
- **Fattore umano e competenza di dominio**: L'intelligenza artificiale aumenta la produttività solo se utilizzata da esperti. Affidarsi ciecamente all'automazione comporta il rischio di atrofia delle competenze. Il ruolo umano deve passare da creatore a controllore attento e responsabile.
- **Responsabilità legale e professionale**: Chi firma un atto o una decisione rimane il responsabile, indipendentemente dal fatto che sia stato supportato da un'intelligenza artificiale. Non deve avvenire una diluizione della responsabilità o un'attribuzione di intenzioni alla macchina.
- **Gestione quotidiana e workflow**: L'uso dell'IA riduce il tempo iniziale di scrittura, ma aumenta il tempo necessario per la verifica. Il valore umano risiede nella capacità di validare i contenuti prodotti dalla macchina.
- **Concetti tecnici fondamentali**: Il modello opera tramite calcoli probabilistici (inferenza) e associazioni statistiche, non tramite una reale comprensione o causalità. Concetti come token, parametri e finestre di contesto definiscono il funzionamento tecnico, ma non sono leggibili o ispezionabili come il codice tradizionale.
- **Miti da sfatare**: Non è vero che il modello "impara" dai dati dell'utente durante l'uso o che sia "perfetto" e incapace di sbagliare. Queste credenze sono pericolose e possono portare a un uso superficiale dello strumento.
- **Criteri per l'uso sicuro**: Si deve "premere il freno" (evitare l'uso dell'IA) quando il compito richiede fatti precisi che il sistema non possiede, quando il contesto differisce da quello di addestramento o quando nessuno è in grado di verificare l'accuratezza del risultato.
- **Gestione degli errori nei sistemi aeroportuali**: È stato evidenziato che gli eventi rari rappresentano le sfide più critiche, poiché la gestione dell'ordinario non presenta complicazioni. Gli errori si annidano prevalentemente nella fase di lettura delle

informazioni, motivo per cui è fondamentale non riporre una fiducia eccessiva nei sistemi automatizzati.

- **Affidabilità dei modelli linguistici su temi di nicchia**: È stato spiegato che i modelli linguistici (Large Language Models) non sono intrinsecamente affidabili quando trattano argomenti specifici. Se un modello fornisce un riferimento normativo preciso su un tema di nicchia, è altamente probabile che tale informazione sia frutto di un'invenzione, poiché le allucinazioni tendono a concentrarsi proprio dove sono richiesti dati verificabili.
- **Verifica dell'output dei modelli linguistici**: È stato sottolineato che la ripetizione di una risposta da parte di un modello non costituisce una forma di verifica. L'unica procedura valida per accertare l'accuratezza dell'output è risalire direttamente alla fonte originaria; in mancanza di una fonte accessibile, il dato prodotto va considerato inaffidabile.
- **Interpretazione critica delle metriche di accuratezza**: È emerso che un valore di accuratezza del 99% dichiarato per un sistema di rilevamento, specialmente in contesti di eventi rari, è un dato privo di significato. Tale percentuale può essere raggiunta anche da un sistema che non segnala mai alcun evento.
- **Parametri essenziali per la valutazione dei sistemi**: Per una valutazione corretta, è necessario richiedere ai fornitori dati specifici su quanti eventi reali vengano intercettati e quanti falsi allarmi vengano prodotti dal sistema. La scelta del compromesso operativo tra questi due fattori rappresenta una decisione basata sulla gestione del rischio e non una questione meramente tecnica.
- **Chiusura del modulo e gestione amministrativa**: Il modulo corrente è stato concluso, con l'indicazione di procedere al modulo successivo. Sono state inoltre brevemente discusse questioni amministrative riguardanti un partecipante della Regione Calabria e la gestione della registrazione della sessione.