EXChange AI · Modulo 3

IA E SOCIETÀ

NOTE DIDATTICHE PER EDUCATORI

Progetto **Expeditions for Change: Navigating the Skills & AI Landscape** · Numero progetto 2024-2-BG01-KA210-ADU-000296153 · exchangeai.eu

Co-funded by the European Union

Il progetto è finanziato dall'Unione Europea, ma le opinioni espresse appartengono agli autori e non riflettono necessariamente la posizione dell'UE o dell'autorità concedente.

Come Usare Queste Note

Scopo delle Note

Le note sono scritte in modo da poter essere copiate direttamente nel campo **Notes** di PowerPoint per la diapositiva corrispondente.

Il testo funziona come uno **script didattico**: cosa spiegare, come introdurre il tema e come attivare il pubblico.

Obiettivo del Modulo

Non solo *che cos'è* l'intelligenza artificiale, ma come essa stia già influenzando la società, il lavoro, l'istruzione, la fiducia e il modo in cui prendiamo decisioni.

L'obiettivo è uno sguardo **realistico ed equilibrato**: né una pubblicità dell'IA né un racconto basato sulla paura, ma una comprensione di benefici, limiti e rischi.

Invitate il pubblico a pensare in modo critico e a fare domande durante tutto il modulo.

La Regola di Riccioli d'Oro

Il dibattito pubblico sull'IA oscilla spesso tra due estremi opposti. La formazione responsabile richiede una terza posizione, equilibrata e critica.

1

Ottimismo Eccessivo

L'IA risolverà tutti i problemi del mondo

2

Posizione Equilibrata

L'IA è uno strumento potente, creato e controllato da persone

3

Pessimismo Eccessivo

L'IA porterà inevitabilmente a una catastrofe

Chiedete ai partecipanti: *"Quale di queste tre posizioni incontrate più spesso nei media o tra i colleghi?"*

Limiti dei Sistemi di IA

Ogni sistema di IA ha limiti, anche quando i risultati sembrano impressionanti. Un esempio emblematico viene dalla **diagnostica per immagini in medicina**: i sistemi possono essere addestrati a riconoscere pattern come il pneumotorace nelle radiografie, mostrando la forza dell'IA nel rilevare schemi, ma anche la necessità di un'interpretazione attenta.

⚠️ Prestazioni

Un sistema può funzionare bene in condizioni di test, ma commettere errori con dati diversi, pazienti diversi o un contesto diverso.

🔍 Spiegabilità

A volte è difficile capire *perché* un modello abbia prodotto proprio quel risultato. In ambiti sensibili non basta avere una risposta: è importante poter verificare la logica e l'affidabilità.

> L'IA deve **supportare**, non sostituire acriticamente, il giudizio professionale.

Discriminazione e Pregiudizi nell'IA

Origini del Problema

L'IA non nasce nel vuoto. Viene addestrata su dati creati da persone e società, e questi dati spesso contengono **disuguaglianze storiche**, stereotipi e rappresentazioni incomplete di diversi gruppi.

I pregiudizi non sono un semplice "errore del computer": sono un **problema sociale** che può essere amplificato dall'automazione.

> ⚠ Questo è particolarmente importante per gli educatori, perché l'istruzione dovrebbe promuovere equità, inclusione e pensiero critico.

Come l'IA Apprende Stereotipi

I sistemi possono apprendere associazioni come **"uomo — programmatore"** e **"donna — casalinga"**, non perché siano vere o desiderabili, ma perché tali associazioni erano frequenti nei testi di addestramento.

Il problema non è solo linguistico. Quando queste associazioni vengono integrate in sistemi reali — motori di ricerca, chatbot, strumenti di selezione del personale o piattaforme educative — possono influenzare decisioni e raccomandazioni.

Il modello non "comprende" l'equità: riconosce **pattern statistici**. Per questo sono fondamentali la revisione umana, la progettazione dei dati e il controllo etico.

Conseguenze Reali dei Pregiudizi

I pregiudizi contano perché i sistemi di IA sono sempre più utilizzati in processi ad alto impatto. Se il sistema è addestrato su dati distorti, può riprodurre o perfino **amplificare decisioni ingiuste**.

Selezione del Personale

Candidati qualificati esclusi automaticamente per caratteristiche demografiche

Riconoscimento Facciale

Tassi di errore più elevati per determinati gruppi etnici

Concessione di Prestiti

Valutazioni del rischio distorte che penalizzano interi quartieri

Personalizzazione Servizi

Raccomandazioni che rinforzano disuguaglianze esistenti

Nel contesto educativo, incoraggiate i partecipanti a chiedere: *Chi è rappresentato nei dati? Chi manca? Chi controlla i risultati? Quali sono le conseguenze di un errore?*

Combattere i Pregiudizi: Soluzioni

La tecnologia da sola non basta. Servono **trasparenza**, procedure chiare di responsabilità e team diversificati, capaci di riconoscere problemi che potrebbero restare invisibili a un gruppo di sviluppo omogeneo.

Per gli educatori: quando usiamo strumenti di IA in un corso, dobbiamo verificare i risultati, confrontare prospettive e non presentare la risposta automatica come una verità neutrale.

Attacchi Adversarial e Sicurezza dell'IA

Cosa Sono gli Attacchi Adversarial

Non parliamo di errori ordinari, ma di **piccole modifiche intenzionali** nei dati o nell'ambiente che possono ingannare un modello. Per una persona il cambiamento può sembrare quasi invisibile, ma per il sistema può portare a una classificazione completamente diversa.

Attacchi nel Mondo Digitale

Si aggiunge una piccola perturbazione a un'immagine e il modello cambia drasticamente risposta — per esempio da *"colibrì"* a *"martello"*, o da *"coniglio"* a *"scrivania"*. Per l'occhio umano l'immagine può rimanere quasi identica, ma la rappresentazione matematica per il modello cambia abbastanza da confonderlo.

⚠ Un'alta accuratezza nei test standard non significa sempre robustezza contro input malevoli o insoliti.

Attacchi nel Mondo Fisico

Gli attacchi adversarial possono avere una forma fisica: occhiali, adesivi, cerotti, modifiche a cartelli o altri oggetti che confondono i sistemi di riconoscimento.

- Un modello può riconoscere male una persona
- Non riconoscere un segnale di stop
- Essere ingannato da un cerotto progettato appositamente

Se un sistema del genere viene usato per il controllo degli accessi, la sicurezza stradale o la sorveglianza, l'errore può avere **effetti reali**.

Chiedete: *"Dove saremmo più preoccupati da una vulnerabilità simile — nei social media, nelle automobili, nelle banche o nelle scuole?"*

Difese e Limiti: La Corsa agli Armamenti

Esistono difese contro gli attacchi adversarial, ma ogni difesa ha un costo. La sicurezza dell'IA non è un'impostazione una tantum, ma un **processo continuo**.

Addestramento Adversarial

Includere esempi di attacchi nel processo di addestramento per rendere il modello più robusto

Test Più Rigorosi

Valutare il sistema su scenari estremi, input insoliti e casi limite prima del rilascio

Monitoraggio Continuo

Sorvegliare il comportamento del sistema in produzione e aggiornare le difese regolarmente

Filtraggio degli Input

Rilevare e bloccare input sospetti prima che raggiungano il modello principale

> Come spam e anti-spam: quando le difese migliorano, anche gli attaccanti si adattano. La sicurezza dell'IA è un processo di **test, monitoraggio e aggiornamento** continui.

Usi Impropri dell'IA: Frodi, Manipolazione e Disinformazione

Distinguiamo tra **danno non intenzionale** e **uso deliberatamente scorretto**. Qui il tema è come le persone possano usare intenzionalmente l'IA per frodi, manipolazione, sorveglianza o disinformazione. Questo non significa che l'IA sia "cattiva": significa che gli strumenti potenti possono essere usati per scopi diversi.

Deepfake

Video o audio in cui una persona sembra dire o fare qualcosa che non ha mai detto o fatto, minando la fiducia nei media e nelle istituzioni

Commenti Falsi di Massa

Generazione automatica di opinioni false che crea un'illusione di consenso pubblico

Truffe Automatizzate

Messaggi di phishing più convincenti e personalizzati grazie all'IA generativa

Sorveglianza IA

Riconoscimento facciale e analisi del comportamento che in contesti oppressivi possono limitare libertà e diritti

L'alfabetizzazione all'IA include non solo la capacità di creare contenuti, ma anche quella di **riconoscerli, verificarli e valutarli criticamente**. Integrate pratiche di alfabetizzazione mediatica: controllare le fonti, cercare il contesto, riconoscere segnali manipolativi.

IA, Lavoro e Trasformazione Professionale

Un Cambiamento Strutturale

L'IA non è solo una questione tecnologica; ha **dimensioni economiche e sociali**. Da un lato può aumentare la produttività, automatizzare compiti ripetitivi e creare nuove professioni. Dall'altro può modificare la struttura del lavoro, richiedere nuove competenze e aumentare il consumo di energia e risorse.

I dati mostrano che entro il 2030 l'automazione può coinvolgere centinaia di milioni di posti di lavoro, mentre si prevede anche la creazione di nuovi posti. È importante non concentrarsi solo sulla "sostituzione", ma sulla **trasformazione**: molte professioni non scompariranno del tutto, ma cambieranno i compiti al loro interno.

Chiedete ai partecipanti: *"Quali compiti nel vostro lavoro possono essere supportati dall'IA e quali dovrebbero restare sotto controllo umano?"*

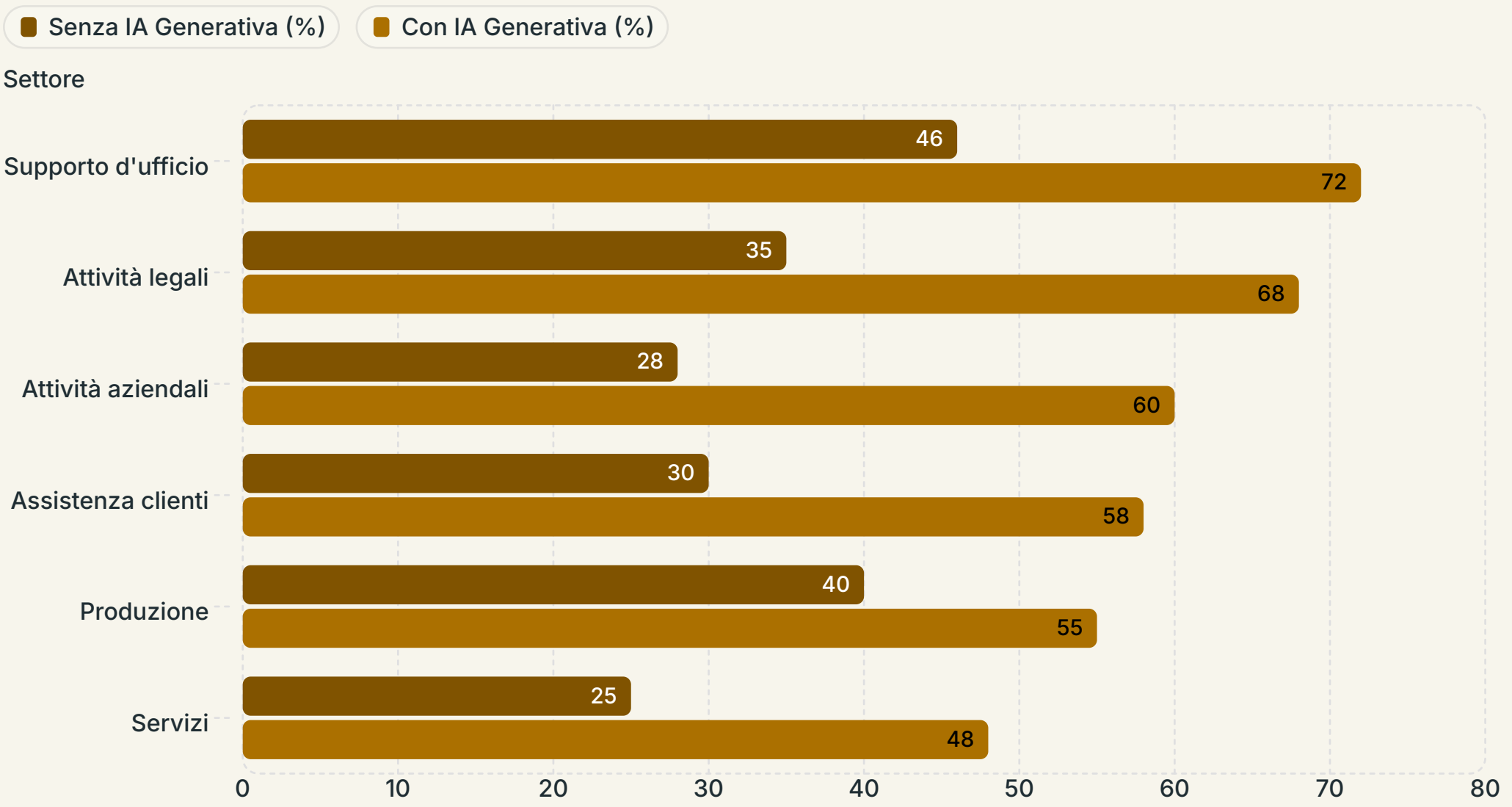
Competenze Umane Irrinunciabili

- Pensiero critico ed etico
- Comunicazione interpersonale
- Progettazione didattica
- Lavoro con le persone
- Giudizio contestuale
- Creatività e innovazione

Proprio gli adulti in apprendimento e i professionisti già attivi avranno bisogno di riqualificazione, competenze aggiuntive e supporto all'adattamento.

Potenziale di Automazione per Settore

Il grafico seguente illustra il confronto tra il potenziale di automazione **con e senza IA generativa** per diversi gruppi professionali. L'IA generativa aumenta questo potenziale perché può lavorare con testi, immagini, codice e compiti comunicativi che prima erano considerati più difficili da automatizzare.



"Potenziale di automazione" non significa automaticamente perdita di lavoro. Significa che alcuni compiti possono essere modificati, accelerati o parzialmente svolti da strumenti. I curricula devono adattarsi: includere alfabetizzazione all'IA, competenze d'uso degli strumenti, etica e capacità di apprendimento permanente.

Soluzioni: Apprendimento Permanente e Politiche

Le soluzioni richiedono una combinazione di misure **individuali, organizzative e politiche**. Non esiste una soluzione universale unica, ma un ecosistema di risposte complementari.

Reti di Sicurezza Sociale

Il reddito di base condizionato è un esempio di idea che offre una rete di sicurezza, collegata a incentivi per l'apprendimento e la riqualificazione

Supporto Organizzativo

Le organizzazioni devono sostenere l'aggiornamento continuo delle competenze dei propri dipendenti

Apprendimento Permanente

Le persone non potranno contare solo sulle conoscenze acquisite nella formazione iniziale: l'aggiornamento continuo è essenziale

Ruolo degli Educatori

Gli educatori sono mediatori fondamentali tra il cambiamento tecnologico e i bisogni reali delle persone

Impronta Ambientale dell'IA

Il Ciclo di Vita dei Modelli IA

I grandi modelli hanno un ciclo di vita con impatti ambientali in ogni fase. I servizi digitali non sono "senza peso": usano elettricità, acqua, materiali e infrastrutture.

01

Progettazione e Sviluppo

Ricerca, sperimentazione, iterazioni

02

Costruzione Data Center

Materiali, energia, acqua per il raffreddamento

03

Produzione Hardware

GPU, chip, componenti elettronici

04

Addestramento del Modello

Fase ad alto consumo energetico

05

Uso e Inferenza

Ogni richiesta ha un'impronta reale

Fattori che Determinano l'Impatto

Anche una singola richiesta ha un'impronta piccola ma reale — emissioni, acqua e materiali. Contano:

- Localizzazione dei data center
- Mix energetico (rinnovabile vs fossile)
- Dimensione del modello
- Frequenza e complessità delle richieste

Consiglio pratico per educatori: insegnare ai partecipanti a lavorare in modo efficiente — prompt chiari, raggruppamento delle domande, evitare ripetizioni inutili, scegliere lo strumento adatto. Così uniamo produttività e responsabilità.

> Questo non significa smettere di usare l'IA, ma usarla in modo **consapevole e proporzionato** al compito.

Conclusione: Alfabetizzazione all'IA come Competenza Civica

L'alfabetizzazione all'IA non è solo una competenza tecnica, ma anche **civica, professionale ed etica**. I partecipanti devono saper usare l'IA, ma anche comprenderne limiti, rischi, possibili usi impropri e conseguenze sociali.

🎯 Pregiudizi e Discriminazione

Dati storici distorti amplificati dall'automazione in decisioni ad alto impatto

🛡️ Attacchi Adversarial

Vulnerabilità intenzionali nei sistemi reali: sicurezza come processo continuo

⚠️ Usi Impropri

Deepfake, disinformazione, sorveglianza: alfabetizzazione mediatica essenziale

🌍 Lavoro e Pianeta

Trasformazione professionale e impronta ambientale: uso consapevole e apprendimento permanente

ⓘ Incoraggiate gli educatori a introdurre l'alfabetizzazione all'IA nei loro corsi attraverso discussioni, attività pratiche, analisi di casi e riflessione critica. Visitate **exchangeai.eu** per gli altri moduli e le risorse del progetto.

Visita exchangeai.eu