

DIPARTIMENTO DI INFORMATICA

- Tipologia dell'attività: INSEGNAMENTI E LABORATORI DEDICATI AI DOTTORANDI E AGLI SPECIALIZZANDI

- Titolo dell'attività per l'acquisizione delle competenze trasversali*:
Transparency and Explainability in Artificial Intelligence

- Sede dell'attività e modalità di erogazione: **Dipartimento di Informatica**

- Periodo di svolgimento delle attività formative dal: **15/01/2024 al 31/01/2024**

- Durata dell'attività (in ore): **30 ore di lezione frontale in modalità online su Teams (Codice: ndhmf83).**

- Data entro la quale è possibile effettuare l'iscrizione (precedente allo svolgimento del 30% delle attività formative previste nell'ambito di ciascun insegnamento o laboratorio) : **19/01/2024**

- Giorni e orari di svolgimento delle attività formative fino alla loro conclusione prevista:

Lunedì: 14:00-17:00

Mercoledì: 14:00-17:00

Venerdì: 14:00-17:00

- Numero di CFU attribuibili agli/alle studenti/esse che avranno frequentato almeno il 70% delle lezioni/sessioni di apprendimento/laboratori/seminari ed avranno superato la relativa prova finale (esame di profitto o verifica): **4**

- Breve descrizione delle attività proposte e delle metodologie didattiche da adottare**:

L'Intelligenza Artificiale (IA) è diventata onnipresente in molteplici contesti della vita sociale, economica e politica. Tuttavia, nonostante la sua diffusione, i potenziali utenti conservano ancora una certa incertezza riguardo alla sua utilità e credibilità. Al fine di guadagnare la fiducia di tali utenti, è essenziale rispondere a domande cruciali: "Possiamo fidarci dei modelli di IA?", "Il modello di IA è in grado di spiegare le motivazioni di una decisione suggerita?".

Il corso è stato progettato per fornire una panoramica completa ed accessibile dei concetti principali dell'IA e della sua applicazione in diversi domini (ad esempio, legale, sanitario, ambientale). Verranno affrontate e discusse le varie sfide legate all'utilizzo dell'IA nel mondo reale, attraverso la discussione di esempi di casi reali. Partendo dall'analisi delle vulnerabilità nell'utilizzo dell'IA, il corso fornirà una panoramica di metodi e strumenti per rendere i modelli di IA comprensibili e trasparenti agli utenti finali, al fine di accrescerne la credibilità. A tal scopo, saranno presentate applicazioni di algoritmi in grado di spiegare i modelli di IA, e illustrati diversi framework (Scikit-learn, Dalex, SHAP) in grado di spiegare visivamente le decisioni di un modello di IA.

L'insegnamento è finalizzato a rafforzare le competenze trasversali su Intelligenza Artificiale Trasparente e si inquadra nelle attività di formazione promosse dal Progetto FAIR – Future AI Research (PE00000013) - spoke 6 "Symbiotic AI".

- Cognome, nome e recapito di posta elettronica o telefonico del/delle/dei docente/i responsabile/i:
Giuseppina Andresini, giuseppina.andresini@uniba.it; telefono 0805442203

* si raccomanda di non effettuare variazioni rispetto al titolo del progetto approvato dal CdA.

** inserire un testo pari a circa 1000 caratteri.