Piano di Attività

Progetto di ricerca "Sistemi automatizzati per la moderazione delle fake news: analisi informatico-giuridica".

La ricerca proposta si svolgerà nell'ambito delle attività del progetto PRIN 2022 PNRR DAFNE (Democratic governance of Automated systems for Fake News, Prot. P2022R7RS9). La ricerca, organizzata in coerenza con l'avanzamento del progetto DAFNE e in collaborazione con il relativo team di ricerca, sarà sviluppata in tre principali attività: (1) raccolta dati e analisi socio-tecnica (tecnologie, attori, processi) dei sistemi automatizzati di moderazione delle fake news utilizzati dalle piattaforme online; (2) analisi dei metodi tecnico-giuridici e delle linee guida per lo sviluppo e l'uso di sistemi automatizzati di moderazione delle fake news nel rispetto dei diritti fondamentali e dei principi democratici. La ricerca si articolerà nelle seguenti tre fasi.

Prima fase: Marzo 2025 - Giugno 2025 (M1-M4)
La prima fase della ricerca mira a ottenere conoscenze empiriche sullo sviluppo e l'uso di sistemi automatizzati per rilevare e moderare le fake news. Consisterà nella raccolta di dati relativi ai sistemi direttamente dai siti web delle piattaforme e di altre organizzazioni coinvolte (es. termini e condizioni, politiche sulla privacy, report di trasparenza, valutazioni di rischio, discussioni, ecc.). La raccolta dei dati riguarderà i settori dei due casi studio inclusi nel progetto, ovvero i social media per la comunicazione politica (Facebook, Twitter, TikTok, Instagram, ecc.) e i siti di pubblicità commerciale e di e-commerce (es. Google, Amazon, Airbnb, TripAdvisor, ecc.).

Seconda fase: Luglio 2025 - Ottobre 2025 (M5-M8)
La seconda fase della ricerca sarà dedicata alle pratiche di moderazione dei contenuti attraverso un metodo socio-tecnico che analizzi come i sistemi automatizzati vengono sviluppati, introdotti e integrati nelle interazioni sociali delle piattaforme online, specificando le attività loro affidate e il loro ruolo nei vari processi coinvolti. Verranno esaminati, ad esempio, gli attori (piattaforme, utenti, segnalatori affidabili, servizi commerciali terzi), le tecnologie impiegate (big data, machine learning, NLP), i compiti delegati (rappresentazione delle fake news, rilevamento delle fake news/account), i processi rilevanti (fact-checking, misure di rimozione) e l'interazione tra ciascun componente.

Terza fase: Novembre 2025 - Febbraio 2026 (M9-M12)
La terza fase sarà incentrata sullo studio e sull'analisi di approcci normativi e by design per integrare il rispetto dei diritti fondamentali e dei valori democratici nei sistemi socio-tecnici di moderazione delle fake news. Saranno sviluppate linee guida per sviluppatori, utenti e altri stakeholder. Gli aspetti by design riguarderanno lo sviluppo e l'uso dei sistemi per garantire che essi rispettino proattivamente la normativa, includendo tecniche e standard per la progettazione di sistemi di moderazione delle fake news, rispetto della privacy e protezione dei dati, non discriminazione, trasparenza, gestione del rischio e valutazione dell'impatto.