Дезинформация 2.0: Разследващи техники за дигитални измами

Преподавател: Лора Петкова

Анотация:

В ерата на все по-бързо развиващи се интелигенти и автоматизирани технологии традиционните форми на дезинформацията постоянно еволюират, достигайки ниво на сложна самоподдържаща и саморазвиваща се мултиплатформена система, в която ботове, тролове и генеративни модели са част от нова проява на информационна агресия, изведена в научните среди като дезинформация 2.0. Появата на изкуствения интелект (ИИ) не само усилва миграцията и влиянието на манипулациите и заблудите чрез анализ на потребителски реакции, персонализиране и алгоритмична адаптация, но и малигнизира „Синдрома на информационното разстройство“, до степен на автономна инфраструктура от deepfake материали, цели армии от кухи акаунти и ферми за съдържание, засягайки всички сфери на живота и застрашавайки да се превърне в хронично състояние на информационната екосистема с неясен изход. Ето защо е необходим интердисциплинарен метод за разследване на дигиталните измами, обединяващ различни аспекти от разследващата и дейта журналистика, цифровата криминалистика и когнитивната психология. Това е холистичен подход, комбиниращ мониторинг на социалните мрежи за идентифициране на аномалии в комуникационните процеси, задълбочен форензичен анализ за търсене на цифрови артефакти в синтетични медии и прогнозиране на вероятни бъдещи проявления на дезинформация 2.0.

Курсът има за цел да запознае студентите с инструментариума на медийното цифрово разследване чрез употреба на широк набор от детектори за верификация, както и с механизмите за задълбочен анализ на генеративни отпечатъци, спектрограми, мрежов и IP анализ, и извличане на метаданни. Ще бъдат представени също техники като Google Dorking (разширено търсене в Google за извличане на скрити данни) и OSINT (Open Source Intelligence – събиране на информация от публично достъпни източници) работни рамки. Семинарните упражнения ще включват разглеждане на реални казуси, разпознаване на изкуствено генерирани и манипулирани медийни формати и изследване на социалните мрежи за идентифициране на координирани дезинформационни атаки.