Umjetna inteligencija u borbi protiv zlostavljanja djece

Zlatan Morić
Sveučilište Algebra
CSC 2024
Zagreb, 10. listopada 2024.

Uvod u problem zlostavljanja djece

- Što je zlostavljanje djece?
  - Fizičko, emocionalno, seksualno zlostavljanje i zanemarivanje
- Globalne statistike
  - Procjena: 3 od 4 djece doživi neki oblik zlostavljanja prije 18. godine (Chaturvedi et al., 2023)
- Utjecaj zlostavljanja
  - Zlostavljanje u djetinjstvu povezano je s razvojem fizičkih i mentalnih bolesti u odrasloj dobi, uključujući migrene, visok krvni tlak, kronični bronhitis, artritis, te smanjenje općeg zdravlja (Afifi et al., 2016).
- Izazovi u prepoznavanju zlostavljanja
  - Ograničeni resursi i nedovoljna obučenost osoblja u mnogim zajednicama.

Uloga umjetne inteligencije u suzbijanju zlostavljanja djece

- Kako AI može pomoći?
  - AI koristi podatke za otkrivanje obrazaca zlostavljanja koje ljudi možda ne prepoznaju.
- Primjene umjetne inteligencije
  - Prepoznavanje zlostavljanja na osnovi medicinskih podataka i ponašajnih obrazaca.
  - Forenzička identifikacija zlostavljanja kroz analizu fizičkih dokaza poput ugriza.
- Konkretni alati
  - Metode prepoznavanja ponašanja zasnovane na neuronskim mrežama, kao što su OpenPose i ST-GCN, pokazale su visoku tačnost u prepoznavanju različitih nivoa fizičkog zlostavljanja djece (Y. Huang et al., 2023)
  - NLP algoritmi za analizu medicinskih zapisa i identifikaciju slučajeva zlostavljanja (Annapragada et al., 2021)
- Prednosti primjene AI
  - Povećana preciznost prepoznavanja zlostavljanja
  - Rano prepoznavanje potencijalnih slučajeva zlostavljanja
  - Ušteda resursa u sustavima zaštite djece

Napredni AI alati za prepoznavanje zlostavljanja djece

- AI u dijagnostičkom slikanju
  - AI tehnologije korištene u pedijatrijskoj radiologiji za otkrivanje ozljeda kod zlostavljanja
  - Primjer: AI pomaže u analizi rendgenskih snimki kako bi se identificirale povrede kod djece koje mogu biti posljedica zlostavljanja (Sorensen et al., 2021)
- AI za forenzičku analizu
  - AI algoritmi koriste se za identifikaciju fizičkih znakova zlostavljanja, poput ugriza ili ozljeda na glavi i vratu
  - Primjer: Forenzički odontolozi koriste AI za prepoznavanje obrazaca ugriza i drugih fizičkih indikacija zlostavljanja (Chaturvedi et al., 2023)

Napredni AI alati za prepoznavanje zlostavljanja djece

- Biometrijska analiza za identifikaciju zlostavljanja
  - Kombinacija prepoznavanja lica i glasa korištena u analiziranju materijala koji sadrži zlostavljanje djece
  - Primjer: Razvijen je softver koji koristi biometrijske podatke iz videa kako bi identificirao žrtve i počinitelje zlostavljanja (Westlake & Brewer, 2022)
- Detekcija zlostavljanja u stvarnom vremenu
  - AI alati za detekciju nasilnih radnji putem video nadzora u stvarnom vremenu
  - Primjer: Primjena algoritama dubokog učenja za automatsko prepoznavanje zlostavljanja u stvarnom vremenu koristeći video nadzor (Chuluunsaikhan et al., 2022)

Primjena AI tehnologija u pedijatrijskoj radiologiji za identifikaciju zlostavljanja

- Ključni problemi u prepoznavanju zlostavljanja djece u radiologiji
  - Fizičke ozljede kod zlostavljanja često su suptilne i teške za prepoznavanje, posebno u ranom stadiju
  - Radiolozi imaju visoku odgovornost u procjeni ozljeda, ali interpretacija slika može biti podložna ljudskim greškama i subjektivnosti
  - Zbog sve veće količine podataka, radiolozi su često preopterećeni, što može rezultirati propuštanjem važnih tragova zlostavljanja

Primjena AI tehnologija u pedijatrijskoj radiologiji za identifikaciju zlostavljanja

- Kako AI pomaže u prepoznavanju ozljeda
  - Umjetna inteligencija (AI) koristi se za analiziranje radioloških snimki (npr. rendgen, CT, MRI) kako bi automatski detektirala abnormalnosti koje mogu ukazivati na zlostavljanje djece
  - AI alati analiziraju uzorke ozljeda (npr. prijelomi, hematomi) i uspoređuju ih s bazama podataka prethodno identificiranih slučajeva zlostavljanja
  - AI može pomoći u smanjenju ljudske pogreške i omogućuje brže prepoznavanje potencijalno opasnih ozljeda

Primjena AI tehnologija u pedijatrijskoj radiologiji za identifikaciju zlostavljanja

- Specifične primjene AI u pedijatrijskoj radiologiji:
  - AI sustavi su posebno razvijeni za analizu snimki glave, trupa i kostiju kako bi se otkrili prijelomi i druge ozljede koje često nastaju kod fizičkog zlostavljanja djece
  - U istraživanju Sorensen et al. (2021), AI modeli korišteni su za identifikaciju ozljeda koje su karakteristične za zlostavljanje, kao što su specifične vrste prijeloma i hematoma, te su pokazali značajnu točnost u usporedbi s tradicionalnim metodama
  - AI alati su također korišteni za prediktivnu analizu, tj. predviđanje rizika da će dijete ponovno biti zlostavljano na temelju povijesnih radioloških snimki i podataka o pacijentu

Primjena AI tehnologija u pedijatrijskoj radiologiji za identifikaciju zlostavljanja

- Prednosti primjene AI u pedijatrijskoj radiologiji
  - Povećana točnost
    - AI sustavi smanjuju mogućnost ljudske pogreške u analizi snimki
  - Rana identifikacija
    - AI omogućuje prepoznavanje ozljeda u ranoj fazi, što omogućuje bržu intervenciju i sprječavanje daljnjeg zlostavljanja
  - Smanjenje opterećenja za radiologe
    - AI alati omogućuju brzu obradu velikih količina podataka, što smanjuje pritisak na medicinsko osoblje

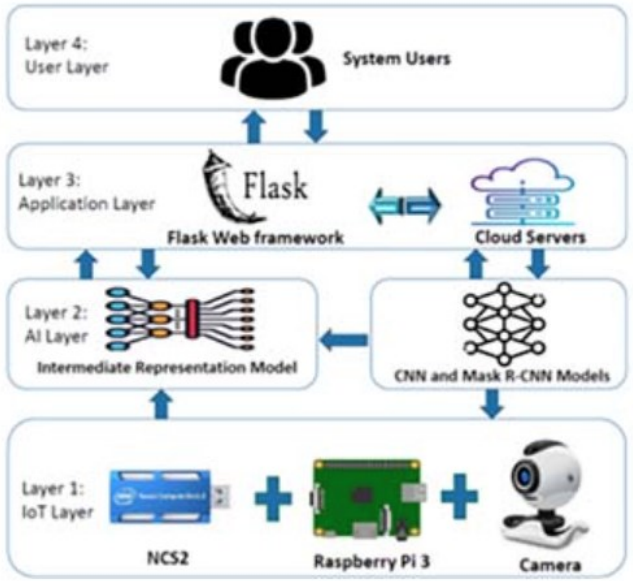
AI alati za detekciju nasilja putem video nadzora

- Sustav za detekciju zlostavljanja u stvarnom vremenu pomoću dubokog učenja, korištenjem video nadzora za automatsko prepoznavanje nasilnih radnji (Chuluunsaikhan et al., 2022)

(a) Normal (b) Abnormal

Detekcija nasilnog ponašanja u javnim prostorima

- „Hawk-Eye” sustav koji koristi duboko učenje za identifikaciju opasnih predmeta i nasilnog ponašanja u realnom vremenu putem kamera u javnim prostorima (Ahmed & Echi, 2021)

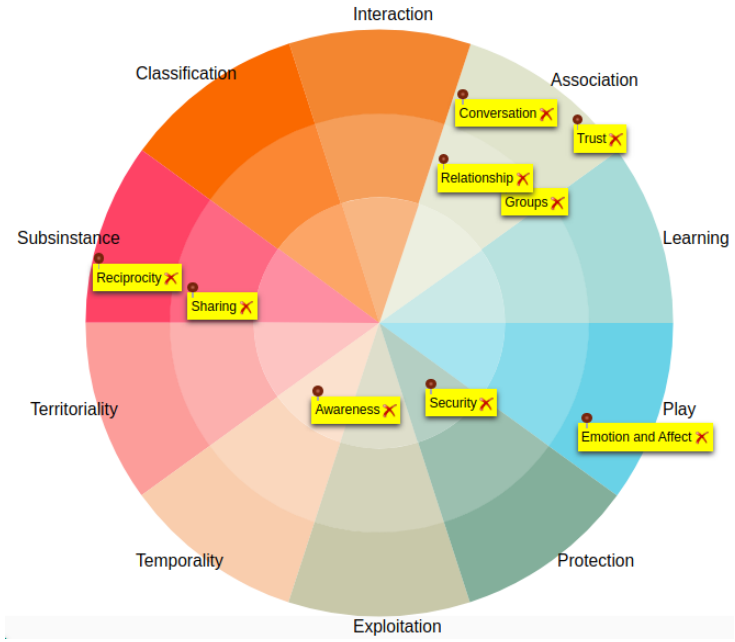


Prepoznavanje nasilja kroz zvučni i video kanal

- Kombinacija prepoznavanja nasilja korištenjem video nadzora i analize zvučnog kanala za identifikaciju zlostavljanja i verbalnog nasilja (Kiani & Kayani, 2022)

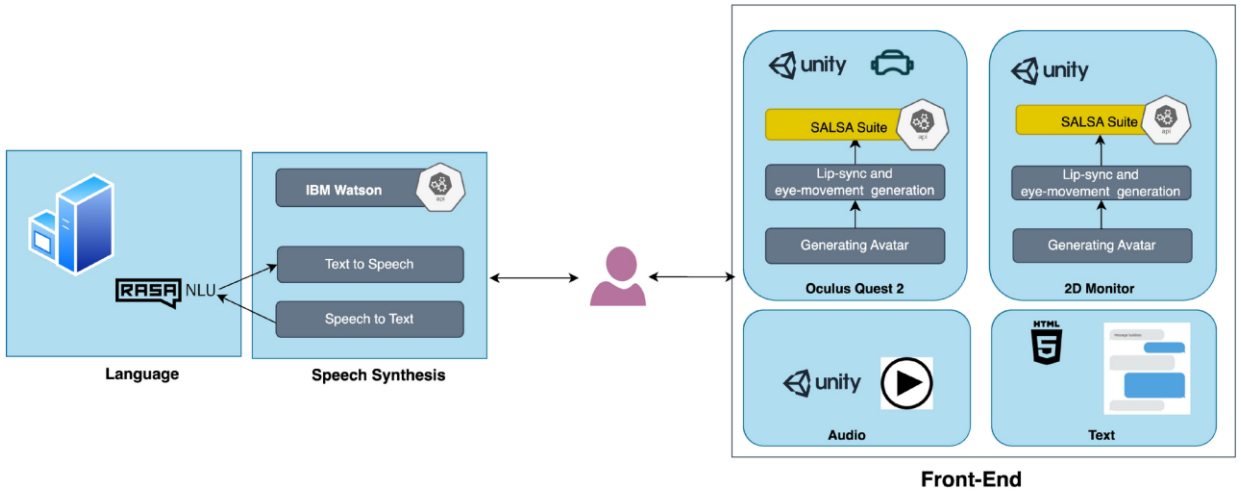
Biometrijska analiza za prepoznavanje predatorskog ponašanja

- AI sustavi koriste biometrijsku analizu za identifikaciju predatorskog ponašanja u video materijalima, prepoznajući lica i glasove koji mogu ukazivati na zlostavljanje (Himi et al., 2020)



Virtualna stvarnost i AI za obuku istražitelja i stručnjaka

- Obuka stručnjaka za vođenje intervjua sa zlostavljanom djecom kroz interaktivne simulacije
- AI-driven avatar u virtualnom okruženju za obuku istražitelja i stručnjaka kroz simulacije intervjua sa zlostavljanom djecom (Hassan et al., 2022)



AI u forenzičkoj analizi za otkrivanje zlostavljanja djece

- Forenzička odontologija i analiza ugriza korištenjem AI tehnologija.
- AI alati pomažu forenzičarima prepoznati obrasce ugriza i analizu mandibularne morfologije za otkrivanje slučajeva zlostavljanja djece (Chaturvedi et al., 2023)

Dijagnostički alati za otkrivanje fizičkog zlostavljanja djece

- SPUTOVAMO-R2 (Screening Tool)
  - Upitnik razvijen za prepoznavanje specifičnih ozljeda koje upućuju na fizičko zlostavljanje. Korišten za identifikaciju ozljeda kod male djece
- TEN-4 (Bruise Screening Tool)
  - Alat za prepoznavanje modrica u djece ispod 4 godine. Specifično dizajniran da identificira modrice na područjima koja nisu tipična za ozljede kod normalnih padova (Torzo, Ears, Neck)
- Child Abuse Risk Evaluation (CARE)
  - Alat koji pomaže u procjeni rizika od zlostavljanja na temelju medicinskih i socijalnih faktora, usmjeren na procjenu ponašanja roditelja i djece
- ESCAPE (Early Screening for Child Abuse Predictive Evidence)
  - Ovaj alat koristi kliničke indikatore za ranu detekciju zlostavljanja i uključuje analizu anamneze, ozljeda i ponašanja roditelja

Dijagnostički alati za otkrivanje fizičkog zlostavljanja djece

- PediBIRN (Pediatric Brain Injury Research Network Algorithm)
  - Usmjeren na prepoznavanje traumatskih ozljeda glave koje su rezultat zlostavljanja. Ovaj alat analizira ozbiljnost ozljeda i povezanost sa zlostavljanjem
- CAFES (Child Abuse Fatality Evaluation Study)
  - Ovaj alat koristi se za retrospektivnu analizu slučajeva smrti djece kako bi se utvrdilo je li zlostavljanje bilo prisutno
- LUCAS (Likelihood of Unexplained Child Abuse Scoring)
  - Skor koji omogućuje procjenu vjerovatnosti zlostavljanja na temelju više faktora kao što su neobjašnjene ozljede, povijest zlostavljanja i reakcija roditelja.

Pregled primjene AI u detekciji i sprječavanju zlostavljanja djece

- Od 400 pregledanih radova, sedam je ispunilo kriterije, ali svi su imali visok rizik od pristranosti zbog malih uzoraka i nedostatka valjanih modela.
- Korišteni su umjetne neuronske mreže, konvolucijske neuronske mreže, te tehnike obrade prirodnog jezika za analizu podataka o zlostavljanju
- AI algoritmi mogu identificirati zlostavljanje, ali je potrebna daljnja validacija i veći uzorci (Lupariello et al., 2023)

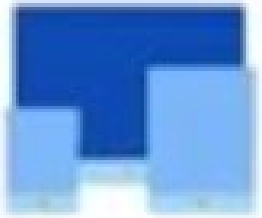
Initial Investigation

Search / Seizure
Warrant Execution

Know Your Tools

Preparation for Digital
Forensic Examination

Victim & Suspect
Identification



Device Attribution &
Corroboration Evidence

Standard
Documentation

Export / Report /
Update

Rastuća prijetnja AI-generiranog materijala seksualnog zlostavljanja djece

- Glavni problem
  - Generativna umjetna inteligencija omogućuje stvaranje novih oblika materijala seksualnog zlostavljanja djece (CSAM), koji je teško razlikovati od stvarnog sadržaja
- Vrste AI-generiranog CSAM-a
  - Tekst u sliku
    - Stvaranje fotorealističnih slika djece
  - Tekst u razgovor
    - AI chatbotovi simuliraju eksplicitne razgovore s djecom
  - Manipulacija slika
    - Neškodljive slike pretvaraju se u eksplicitne sadržaje
- Posljedice
  - AI-generirani CSAM povećava internetsku seksualnu eksploataciju djece, pri čemu počinitelji koriste AI za stvaranje i širenje štetnih sadržaja putem popularnih platformi poput Instagrama i TikToka

Izazovi u otkrivanju AI-generiranog CSAM-a

- Trenutne metode otkrivanja
  - Provođenje zakona
    - Policijske snage teško prate napredak AI tehnologija, a gotovo 50% zaplijenjenog materijala sada uključuje AI-generirane sadržaje
  - Privatni sektor
    - Tehnološke tvrtke, uključujući platforme društvenih medija, ključni su u širenju AI-generiranog CSAM-a, no suočavaju se s poteškoćama u provođenju učinkovitih sigurnosnih mjera
- Pravni izazovi
  - Spora prilagodba zakonodavstva
    - Zakonodavci i tijela za provođenje zakona zaostaju u donošenju zakona koji bi kaznili počinitelje koji stvaraju ili distribuiraju AI-generirani CSAM
  - Manipulacija dokazima
    - Počinitelji koriste AI kako bi modificirali stvarni CSAM i učinili ga da izgleda kao da je generiran umjetnom inteligencijom, otežavajući istrage

Preporuke za suzbijanje AI-generiranog CSAM-a

- Privatni sektor
  - AI programeri moraju ugraditi zaštitne mjere koje sprječavaju generiranje eksplicitnog sadržaja
  - Društvene platforme trebaju provoditi strože moderiranje i metode detekcije
- Države
  - Zakonodavstvo mora biti ažurirano kako bi obuhvatilo AI-generirani CSAM i omogućilo suradnju između AI programera i tijela za provođenje zakona
- Provođenje zakona
  - Policijske snage moraju usvojiti AI alate za prepoznavanje i detekciju AI-generiranog CSAM-a u stvarnom vremenu te surađivati na međunarodnoj razini
- Skrbnici
  - Potrebno je educirati roditelje i učitelje o rizicima AI-generiranog CSAM-a te poticati otvorenu komunikaciju s djecom o sigurnosti na internetu

Je li generiranje AI CSAM-a manje štetno od stvarnog CSAM-a?

- Argumenti za AI-generiran sadržaj
  - Nema stvarne žrtve: AI-generirane slike i videozapisi ne uključuju stvarnu djecu, pa ne dolazi do izravnog zlostavljanja
  - Mogućnost preusmjeravanja kriminalnih radnji: Počinitelji bi mogli koristiti AI-generirane materijale umjesto stvarnog zlostavljanja djece, što potencijalno smanjuje stvarne zločine
- Argumenti protiv AI-generiranog sadržaja
  - Normalizacija zlostavljanja: AI-generirani sadržaj može pridonijeti normalizaciji pedofilskih sklonosti i potaknuti stvarne zločine
  - Zamagljivanje granice: AI-generirani CSAM može biti teško razlikovati od stvarnog materijala, što otežava provođenje zakona
  - Kreiranje novih tržišta: AI-generirani sadržaj može stvoriti nove "potražnje" za materijalima, povećavajući interes za CSAM umjesto smanjenja

Metode detekcije CSAM-a

- Baze podataka sa slikovnim hashovima
  - Koriste se za identifikaciju poznatih slika zlostavljanja kroz jedinstvene hash kodove. Primjeri: PhotoDNA od Microsofta
- Web-crawleri
  - Automatizirani alati koji pretražuju web stranice kako bi pronašli i indeksirali CSAM sadržaj. Primjeri: Project Arachnid (Kanada) i IWF crawler (UK)
- Algoritmi vizualne detekcije
  - Koriste se za prepoznavanje novog i nepoznatog CSAM-a putem vizualnih značajki (boja, tekstura) i dubokog učenja
- Duboko učenje
  - CNN (konvolucijske neuronske mreže) koriste se za prepoznavanje slika i videa s CSAM sadržajem. Ovi algoritmi su najučinkovitiji u detekciji novih i nepoznatih slučajeva.

Zaključci

- AI tehnologija je ključna za prepoznavanje zlostavljanja djece
  - AI sustavi pružaju alate za ranu detekciju zlostavljanja kroz analizu medicinskih podataka, radioloških snimki i ponašajnih obrazaca
- Prednosti primjene AI u zaštiti djece
  - AI omogućuje veću preciznost, ranu identifikaciju potencijalnih slučajeva zlostavljanja i smanjuje pritisak na stručnjake, omogućujući bržu intervenciju
- Izazovi i budućnost
  - Potrebno je razviti bolje modele za prepoznavanje zlostavljanja, s naglaskom na smanjenje pristranosti, povećanje točnosti te poboljšanje pravnih okvira kako bi AI tehnologija bila učinkovita u globalnom kontekstu.
- Dilema ili ne oko AI-generiranog CSAM-a
  - AI-generirani sadržaji ne uključuju stvarnu djecu, no mogu normalizirati pedofiliju i potaknuti stvarno zlostavljanje, te otežati provođenje zakona jer se stvarni sadržaji mogu prikriti kao AI-generirani
  - Zakonodavstvo kasni s prilagodbom, što omogućuje počiniteljima da manipuliraju stvarnim materijalima koristeći AI i izbjegnu kaznene posljedice


Thank you for
your attention!
Q&A