# Primjena umjetne inteligencije u kibernetičkoj sigurnosti, statusi, zablude i izazovi

Doc. dr. sc. Stjepan Groš
Voditelj Laboratorija za informacijsku sigurnost i privatnost
Fakultet elektrotehnike i računarstva Sveučilišta u Zagrebu

Inovacijski centar Nikola Tesla

# Napomene (Disclaimer)

**Mišljenja i stavovi iznesena u ovoj prezentaciji su isključivo moja te se baziraju na moji trenutnim znanjima. S njima se moj poslodavac može, ali i ne mora složiti. Nadalje, kako stičem nova znanja mišljenja i stavovi su podložni izmjenama.**

**Kada govorimo o sigurnosti postoje dvije klase problema**

- Problemi koji su postojali i prije Interneta, ali Internet je omogućio nove načine njihova provođenja
- Problemi koji su se pojavili pojavom Interneta (i općenitije računala)
  - Uglavnom se bavimo ovom skupinom problema u ovoj prezentaciji

# AI ili strojno učenje(ML) ili duboko učenje (DL)?

**Umjetna inteligencija predmet je istraživanja već 70+ godina.**

**Zadnji značajna promjena desila se prije ~15 godina s dubokim učenjem**

Stanje tehnologije je omogućilo takav razvoj, ali koncepti su već od prije poznati

**Dvije su posljedice takvog stanja**

Primjena AI u sigurnosti nije nešto novo, već desetljećima se eksperimentira s tom tehnologijom

Kada **danas** govorimo o AI, onda uglavnom mislimo na DL

**Zbog svega toga ja se prvenstveno bavim u svojoj prezentaciji DL-om u sigurnosti**

# Klase stručnjaka – iz perspektive sigurnosti

## Inženjeri/znanstvenici koji se bave sigurnošću (SECURITY)

- Relativno mlada disciplina
- Glavni neprijatelj je inteligentno i kreativno biće (odnosno, češće/bitnije grupa)

## Inženjeri/znanstvenici koji se bave upravljačkim sustavima

- Desetljeća (stoljeće?) iskustva u sigurnosti (SAFETY!!!) i pouzdanosti
- „Neprijatelj" je priroda i slučajnost
  - tek sa terorizmom susret s namjernim prijetnjama
- Od namjernih prijetnji tek fizička sigurnost i destrukcija

## Svi ostali

- Nisu obučeni niti za *safety* niti za *security*

***Od sada kada kažem sigurnost mislim na security, osim ako eksplicitno ne kažem drugačije***

# Zašto uopće sigurnost?

**Sigurnost je znanstvena/inženjerska disciplina čija svrha je osigurati da sustavi rade ono za što su namijenjeni u prisutnosti namjerne i nenamjerne prijetnje pri čemu se za to koristi minimum potrebnih resursa.**

„Sustavi" su bilo što, ali ključno je da sadrže informacijsko-komunikacijske tehnologije – cilj sigurnosti je djelovati na ICT

**Kada govorimo o sigurnosti, sve što imamo u sustavu od informacijsko-komunikacijske tehnologije može biti i cilj djelovanja namjernih i nenamjernih prijetnji**

Odnosno slaba točka sustava

# Brzi pregled nekih osnovnih pojmova iz sigurnosti

**Ranjivost (engl. vulnerability) je pogreška ili slabost u dizajnu sustava, implementaciji, upotrebi ili upravljanju koja se može iskoristiti za narušavanje sigurnosti sustava ili informacije.**

Pogreške u programskoj podršci (engl. bugs), propusti u protokolima, kriva upotreba programske podrške ili nekog sustava

**Prijetnja (engl. threat) je bilo kakav događaj koji može iskoristiti ranjivost te na taj način prouzročiti štetu.**

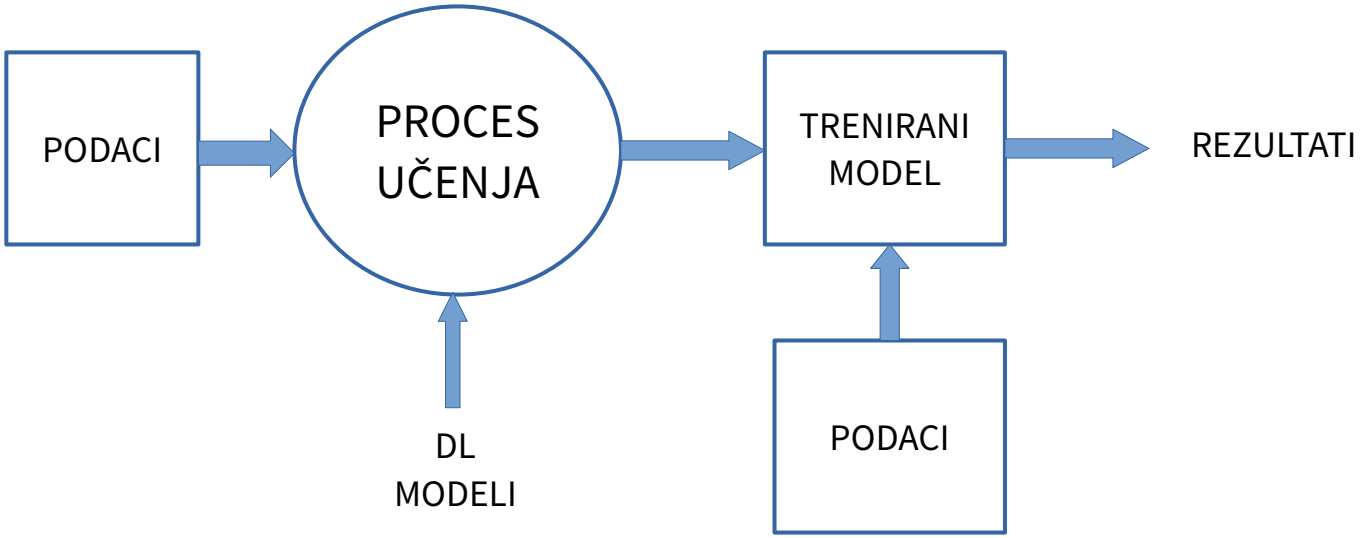
Izvori prijetnji su ljudski (napadači) ili prirodni (potres, nestanak struje);

Dodatno ljudski izvori mogu biti namjerni (napadači) ili slučajni (nepažnja osobe)

# O sigurnosti dubokog učenja

Prije nego što se pozabavimo primjenom dubokog učenja u sigurnosti, pogledajmo malo sigurnost samog dubokog učenja

Konceptualno, ulazi i način korištenja DL-a izgleda na sljedeći način:




Svaki element može biti mjesto napada!

# Problemi dubokog učenja

## Ključni problemi

- Potrebna velika količina kvalitetnih (označenih) podataka
- Nema teorijskog okvira za dizajn DL sustava
- Za dobijene rezultate ne mogu se dobiti objašnjenja ZAŠTO i KAKO

## Kada se takvi modeli nađu u neprijateljskoj okolini ne postoji „razum” koji bi kontrolirao izlaze

- Primjena takvih modela u *safety-ju* je duboko upitna
- Primjena kod kojih čovjek može na neki način nastradati ili imati problema je također duboko upitna

Nisam se dotaknuo vrlo bitnih problema sa pristranostima, etičnošću, itd.

# Primjena dubokog učenja u poboljšanju sigurnosti

## Vrlo aktivno područje istraživanja

Sigurnosni problemi su prilično teški te još nema zadovoljavajućih rješenja

## Primjenu u sigurnosti možemo sagledati po „vrstama” sigurnosti

Strateška, operativna sigurnost – mizerne primjene umjetne inteligencije

Taktička razina – nešto više primjene

Tehnička razina – najveća primjena

Ofenzivna (napadačka) sigurnost – za sada vrlo malo istraživanja (barem javnog)

Defenzivna (obrambena) sigurnost – puno više istraživanja

Naglasak na prevenciji ranjivosti, sprečavanju napada (anti-spam), otkrivanje incidenata/anomalija, situacijskoj svijesti

# Primjena 1: Detekcija anomalija

## Vrlo primamljiva ideja – želimo da stroj otkrije „neuobičajno” ponašanje

- U osnovi, radi se o vrlo staroj ideji
  - sustavi za detekciju upada (engl. intrusion detection systems)
- Moderna reinkarnacija – User (and Entity) Behavioral Analytics – UBA/UEBA

## U praksi, susreće se s nizom problema (uz opće probleme DL-a)

- Strojno učenje se više koristi za detekciju sličnosti (klasifikacija) nego različitosti
- U sigurnosti pogreške su skupe
- Mrežni promet je vrlo raznolik
- Sustav i napadači se stalno mijenjaju

# Primjena 2: Detekcija zlćudnog koda

**Antivirusna programska podrška tradicionalno koristi potpise za detekciju zlćudnog koda**

Međutim, autori zlćudnog koda vrlo lako mijenja svoje tehničke karakteristike

**Antivirusi gube „bitku” s autorima zlćudnog koda te se traže novi načini detekcije zlćudnog koda – razne tehnike AI-ja**

Statička analiza – pokušaj detekcije zlćudnog koda na temelju binarnog/tekstualnog sadržaja

Dinamička analiza – praćenje poziva funkcija operacijskog sustava (detekcija anomalija!)

**Za sada su te primjene istraživačke i u laboratoriju pokazuju relativno dobre rezultate, ali u praksi ne djeluju**

# Primjena 3: Otkrivanje ranjivosti

## Opća ideja: Ukloniti ranjivosti iz programske podrške i tako onemogućiti napade

Čovjek uvijek ostaje slaba točka, ali to ćemo trenutno zanemariti

## Ova klasa je podskup aktivnosti otkrivanja pogrešaka u kodu

A kao što znamo, radi se o vrlo složenom problemu

# U konačnici, protiv koga se borimo?

**Preskočio sam bitnu informaciju, bitno je protiv koga se borimo!**

## APT

Iza njih stoje nacionalne države – potencijalno neograničeni resursi na raspolaganju

Moje osobno mišljenje – nema tog tehničkog rješenja koje će vas zaštititi od ovog izvora napada

S pozitivne strane, vjerojatno niste cilj tih prijetnji – bar ne direktan

## Kibernetički kriminal

Najbolje krim. grupe još uvijek imaju dosta resursa, ali  manje od APT-a

Agilni i inovativni

## Hakeri

Još manje raspoloživih resursa

## Automatizirane probe

Za ovo ne treba AI/ML/DL, jednostavno se rješava dobrom „sigurnosnom higijenom”

# Umjesto zaključka

**Neosporno je da su AI/ML/DL tehnologije koje duboko utječu na nas i svijet oko nas**

**ALI, AI/ML/DL je „buzzword” zbog kojega se ta tehnologija upotrebljava tamo gdje ne bi trebala, bez promišljanja ili se „klasične” stvari guraju pod AI/ML/DL**

- Startupi generiraju tehnološka rješenja i pokušavaju riješiti (utjecati) na društvene probleme
- Tvrtke na svoje proizvode lijepe AI/ML/DL naljepnice kako bi ih lakše prodali

**U sigurnosti AI/ML/DL još nije pokazala velike napretke te se susreće s vrlo teškim izazovima**

- Jesu li neki od tih izazova fundamentalna ograničenja postojeće tehnologije ostaje za vidjeti