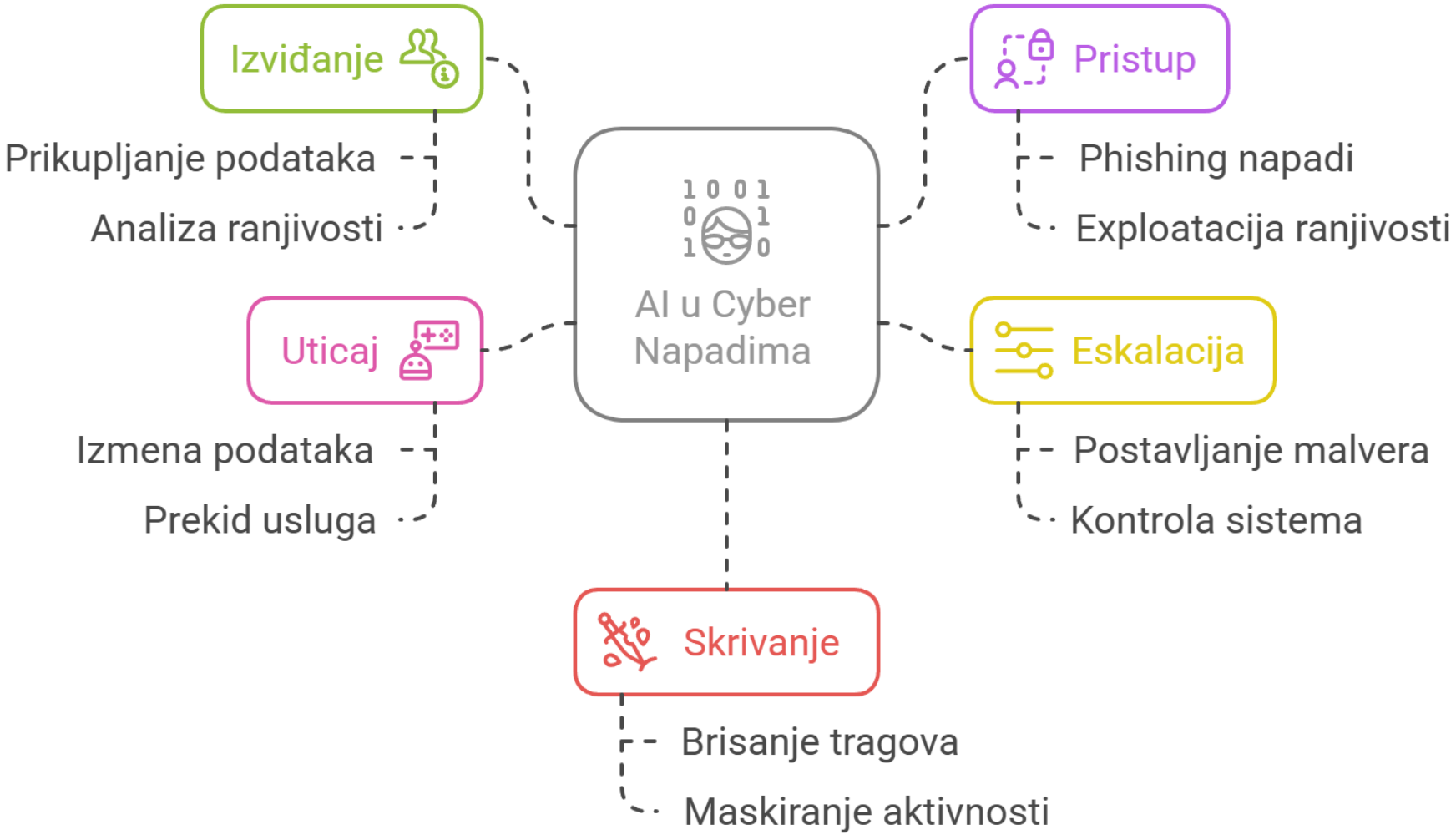
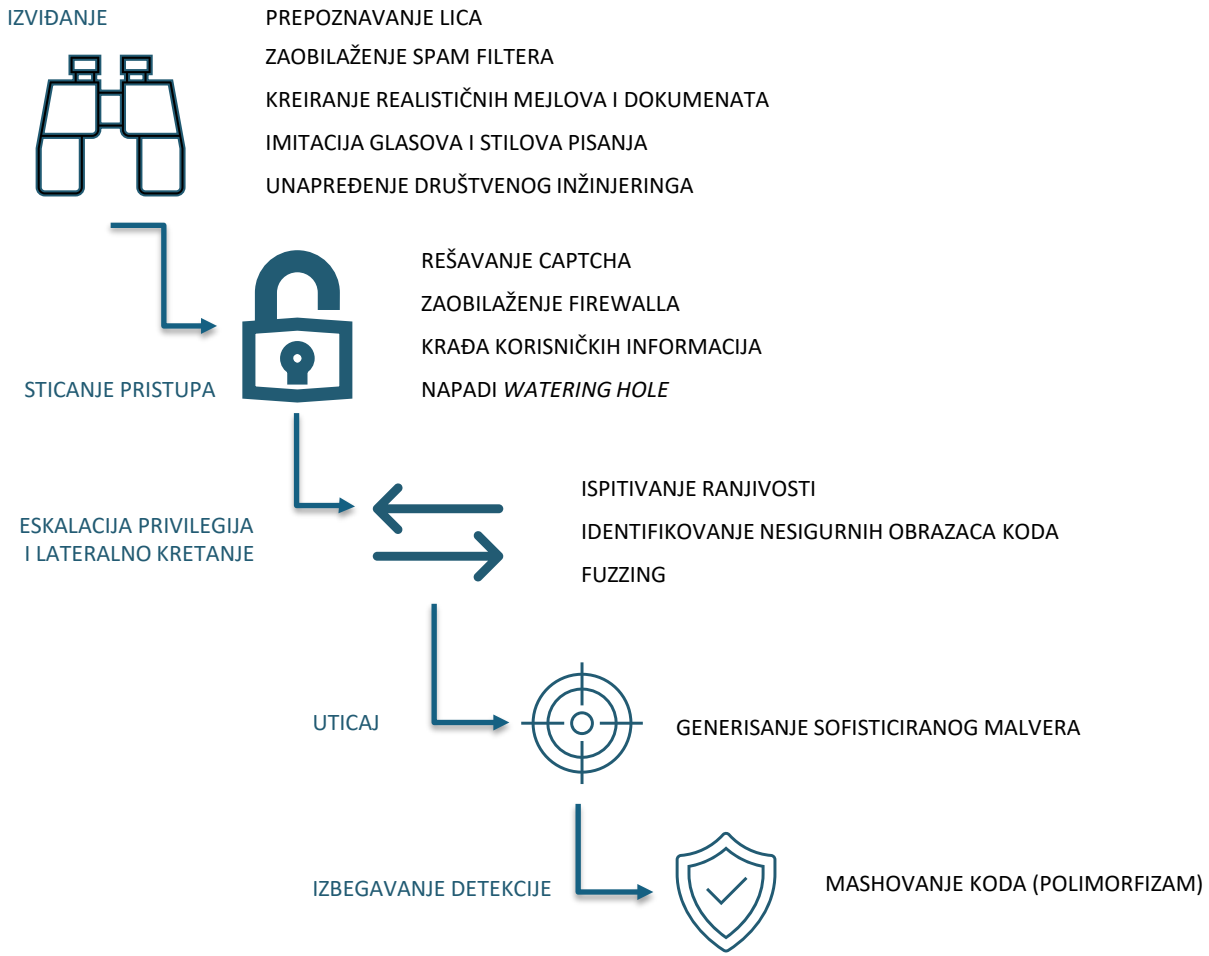
UTICAJ AI NA CYBER BEZBEDNOST

Predavač: dr Dušan Stefanović

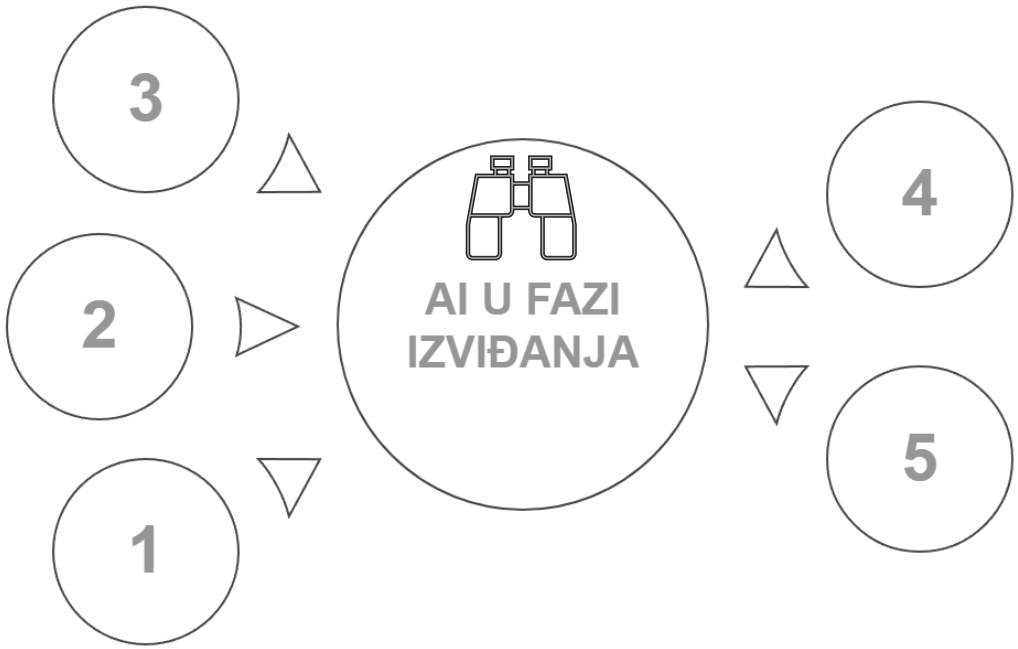
PRIMENA VEŠTAČKE INTELIGENCIJE U NAPADU



PRIMENA VEŠTAČKE INTELIGENCIJE U NAPADU



PRIMENA AI U FAZI IZVIĐANJA



Kreiranje phishinga

AI generiše uverljive lažne komunikacije za prevare.

Zaobilaženje spam filtera

AI koristi prirodni jezik da bi izbegao detekciju spam-a.

Prepoznavanje lica

AI identifikuje pojedince putem slika lica na mreži i putem kamera.

Imitacija glasova

AI replicira ljudske glasove za obmanjujuće audio sadržaje.

Unapređenje društvenog inženjeringa

AI analizira obrasce ponašanja za ciljanje prevara.

RAZUMEVANJE PRETNJI PREPOZNAVANJEM LICA

AI modeli

Tehnologije kao što su OpenFace i DeepFace se koriste za prepoznavanje lica

Javne slike

Slike sa platformi poput LinkedIn-a i Facebook-a se koriste za identifikaciju za lociranje osobe koja ima pristup korporativnim mrežama analizom slika sa konferencija ili javnih mesta.

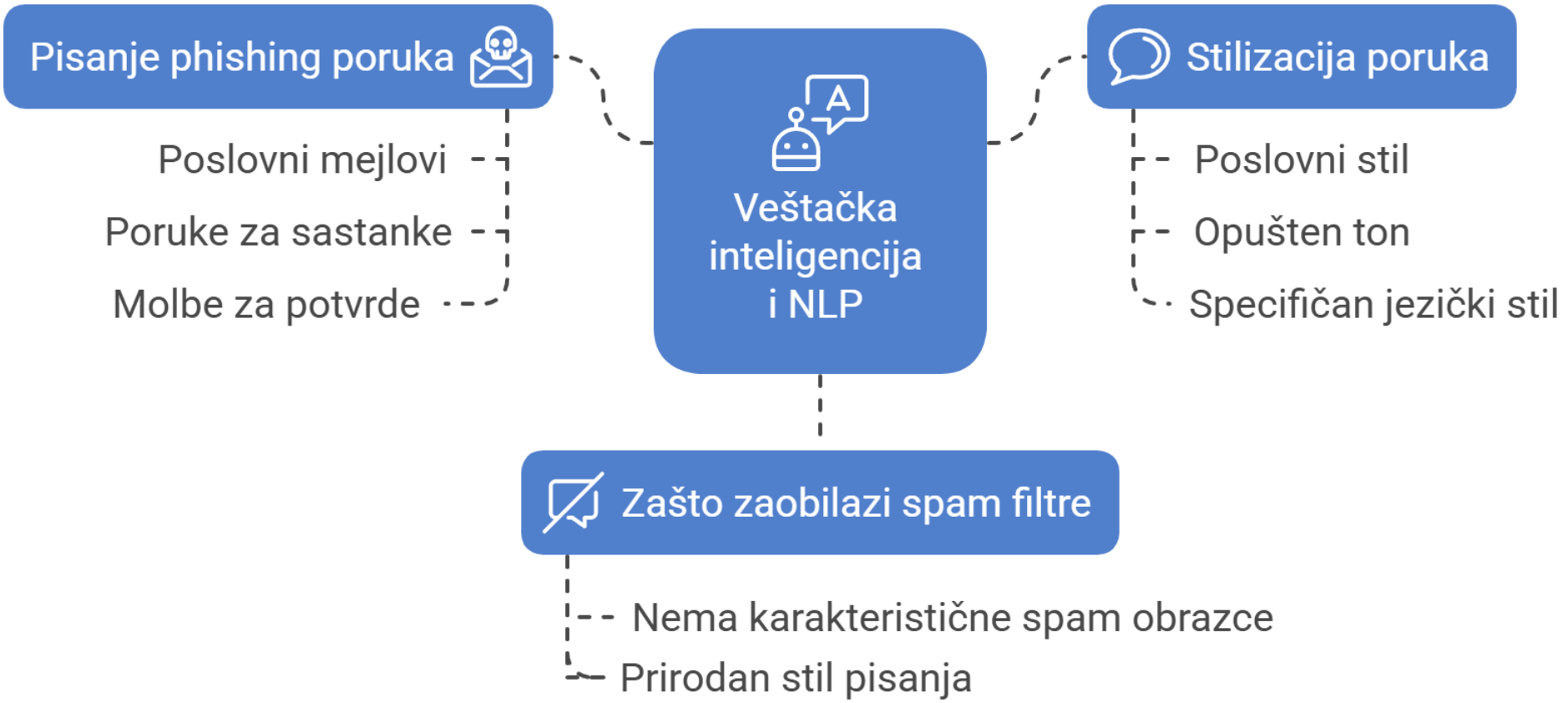
Nadzorne kamere

Video zapisi i slike pomažu u mapiranju kretanja i ponašanja

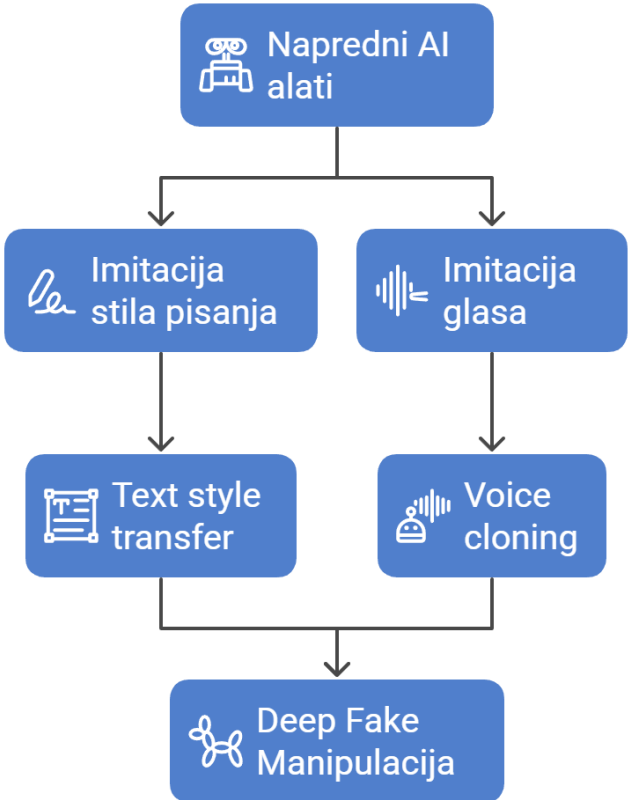
Strategije ciljanje

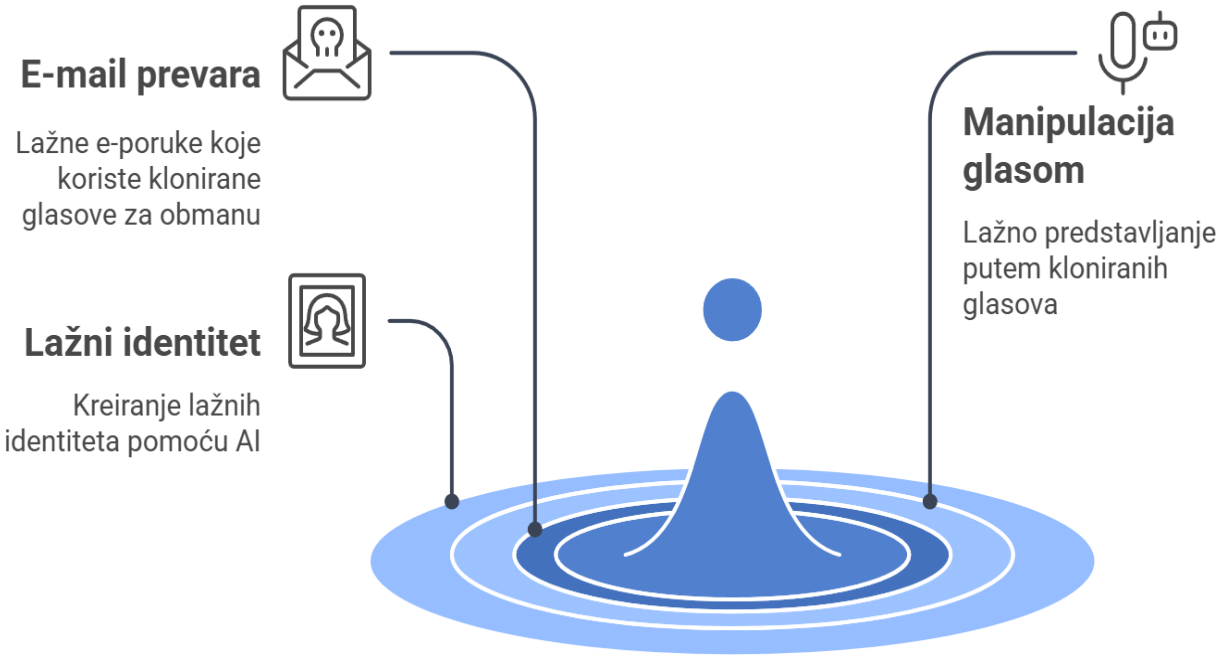
Metode kao što su phishing i socijalni inženjering se koriste za preciznije ciljanje

RAZUMEVANJE PRETNJI ZAOBILAŽENJEM SPAM FILTERA KORIŠĆENJEM NLP-a

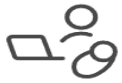


OPONAŠANJE GLASOVA I STILA PISANJA – DEEP FAKE





AI U ANALIZI PONAŠANJA KORISNIKA



Analiza ponašanja zaposlenih

AI analizira online aktivnosti zaposlenih kako bi razumeo obrasce ponašanja

Analiza obrazaca komunikacije

AI proučava kako zaposleni komuniciraju jedni s drugima

Profilisanje zaposlenih

AI stvara profile na osnovu analize ponašanja i komunikacije

Identifikacija ranjivosti

AI identifikuje zaposlene koji su ranjivi na napade socijalnog inženjeringa

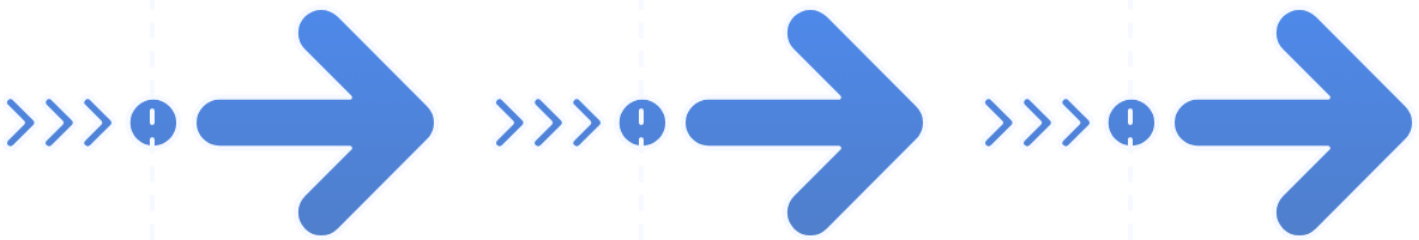
Izvršenje napada

Napadač koristi AI uvide za ciljanje zaposlenih koji su više podložni manipulaciji

STICANJE PRISTUPA

Faza Izviđanja

Napadač prikuplja informacije o meti



Primena AI Tehnika

AI se koristi za automatizaciju i sofisticiranost

Prelazak u Fazu Pristupa

Napadač cilja direktan pristup sistemima ili podacima

TEHNIKE STICANJA PRISTUPA

Automatsko Rešavanje CAPTCHA

Proces automatskog rešavanja CAPTCHA sistema

Prepoznavanje Slabosti Firewall-a

Identifikacija i eksploatacija slabosti firewall-a

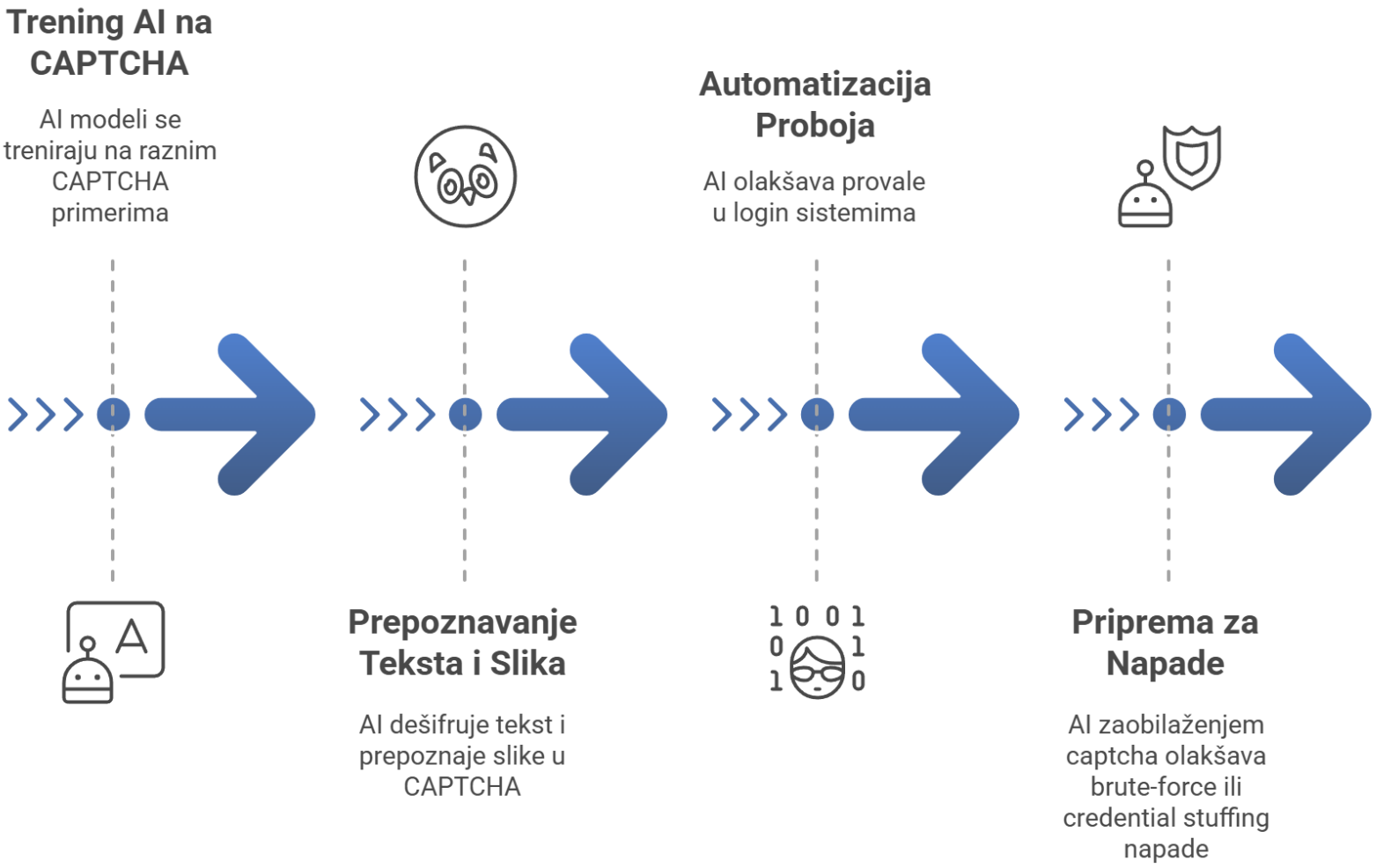
Krađa Korisničkih Podataka

Analiza komunikacije

Watering Hole Napadi

Kompromitovanje često posećenih sajtova

AUTOMATSKO REŠAVANJE CAPTCHA IZAZOVA



PREPOZNAVANJE SLABOSTI FIREWALL-A PRIMENOM AI

Analiza mrežnog saobraćaja
AI modeli analiziraju mrežni saobraćaj i odgove servera da bi identifikovali greške u konfiguraciji slabosti

Istraživanje kombinacija paketa
AI istražuje različite kombinacije paketa i portova korišćenjem reinforcment learning tehnike učenja

Identifikacija anomalija
AI identifikuje anomalije u pravilima zaštite tako što uči koji zahtevi prolaze kroz firewall

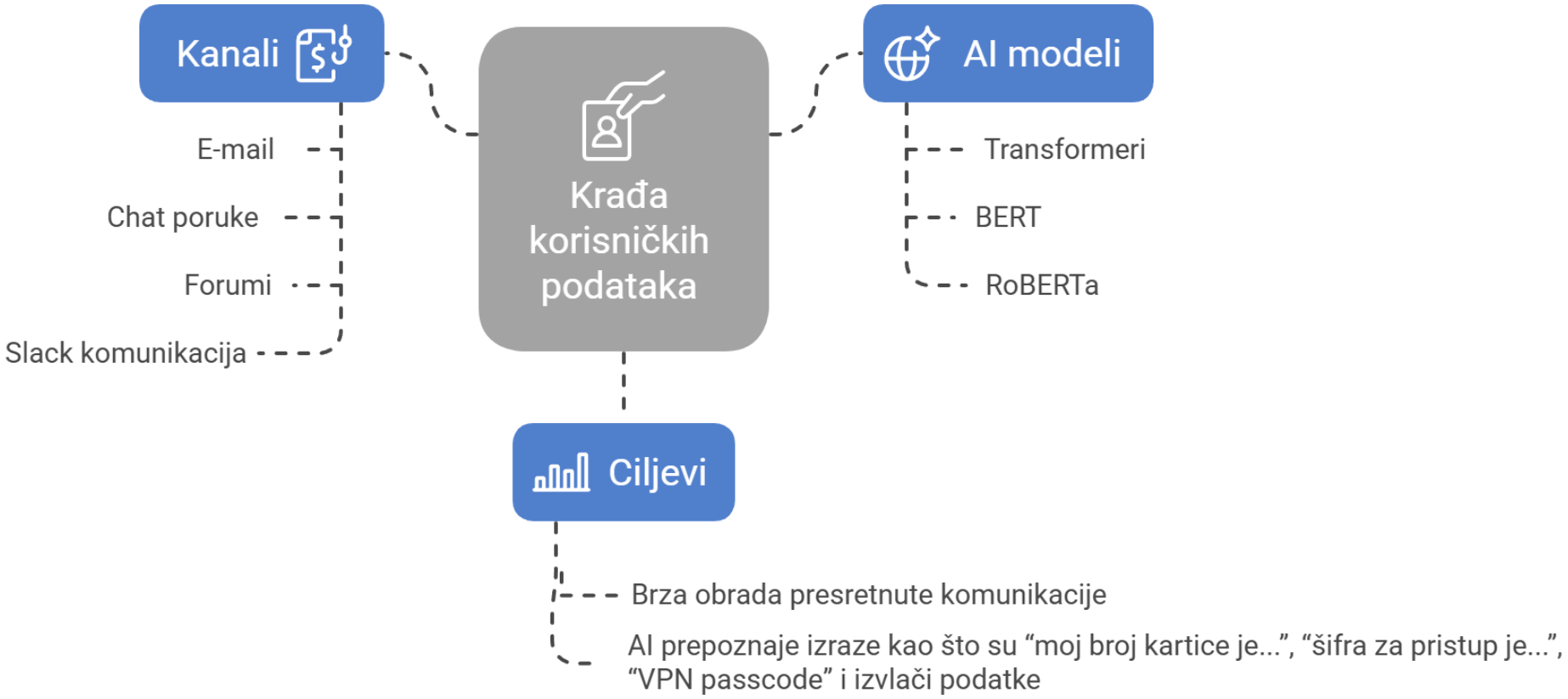
Učenje efikasnih zahteva
AI uči koji zahtevi prolaze kroz firewall

Oblikovanje mrežnog saobraćaja
AI oblikuje svoj mrežni saobraćaj za prolazak kroz firewall

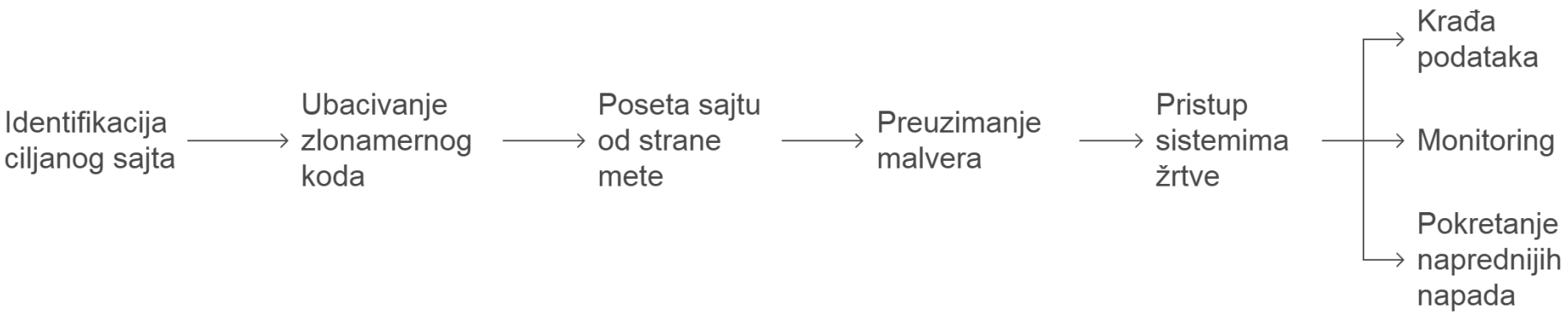
Otkrivanje ranjivih servisa
AI otkriva ranjive servise iza firewall-a

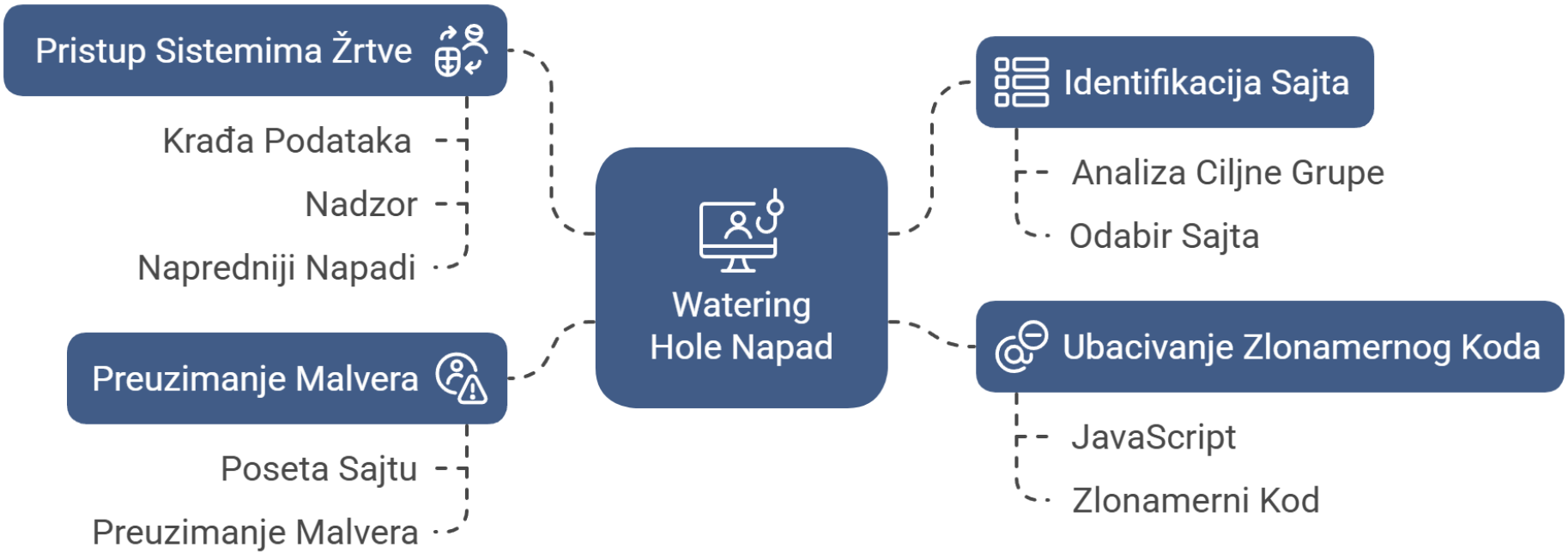
Priprema za eksploataciju
AI priprema strategije za eksploataciju slabosti

KRAĐA KORISNIČKIH PODATAKA ANALIZOM KOMUNIKACIJE



Watering Hole napad





Upotreba veštačke inteligencije u Watering Hole napadima

Upotreba mašinskog učenja od strane napadača

- **Analiza saobraćaja** — Analiziranje obrazaca internet saobraćaja za zlonamerne aktivnosti.
- **Identifikacija ciljeva** — Identifikovanje sajtova koje često posećuje ciljana grupa.
- **Profilisanje korisnika** — Praćenje interesovanja i ponašanja korisnika radi kreiranja profila.

AI-omogućena personalizacija zlonamernog sadržaja

- **AI generiše lažne forme** — AI stvara uverljive lažne forme, upozorenja ili zahteve.
- **AI generiše adaptivne exploit skripte** — AI razvija skripte koje se prilagođavaju korisnikovom sistemu.
- **Skripte se prilagođavaju operativnom sistemu** — Skripte se menjaju kako bi odgovarale korisnikovom operativnom sistemu.
- **Skripte se prilagođavaju browseru** — Skripte se menjaju kako bi odgovarale korisnikovom browseru.
- **Skripte se prilagođavaju IP lokaciji** — Skripte se menjaju kako bi odgovarale korisnikovoj IP lokaciji.

Upotreba veštačke inteligencije u Watering Hole napadima

Izbegavanje otkrivanja u sandboxu

Napadači treniraju modele

Napadači razvijaju modele za prepoznavanje sandboxa.

Model detektuje sandbox

Model uspešno identifikuje sandbox okruženje.

Kod se ne aktivira

Kod se ne izvršava kako bi se izbeglo otkrivanje.

Vrste deepfake sadržaja

Deepfake video zapisi

Realistične video poruke sa deepfake licima.

AI-generisane izjave

Lažne AI izjave ljudi od autoriteta.

Primer veštačke inteligencije u Watering Hole napadima

1 **Analiza profila**

Napadači koriste AI za analizu LinkedIn profila zaposlenih.

2 **Identifikacija foruma**

Napadači otkrivaju forum koji zaposleni često koriste.

3 **Kompromitovanje foruma**

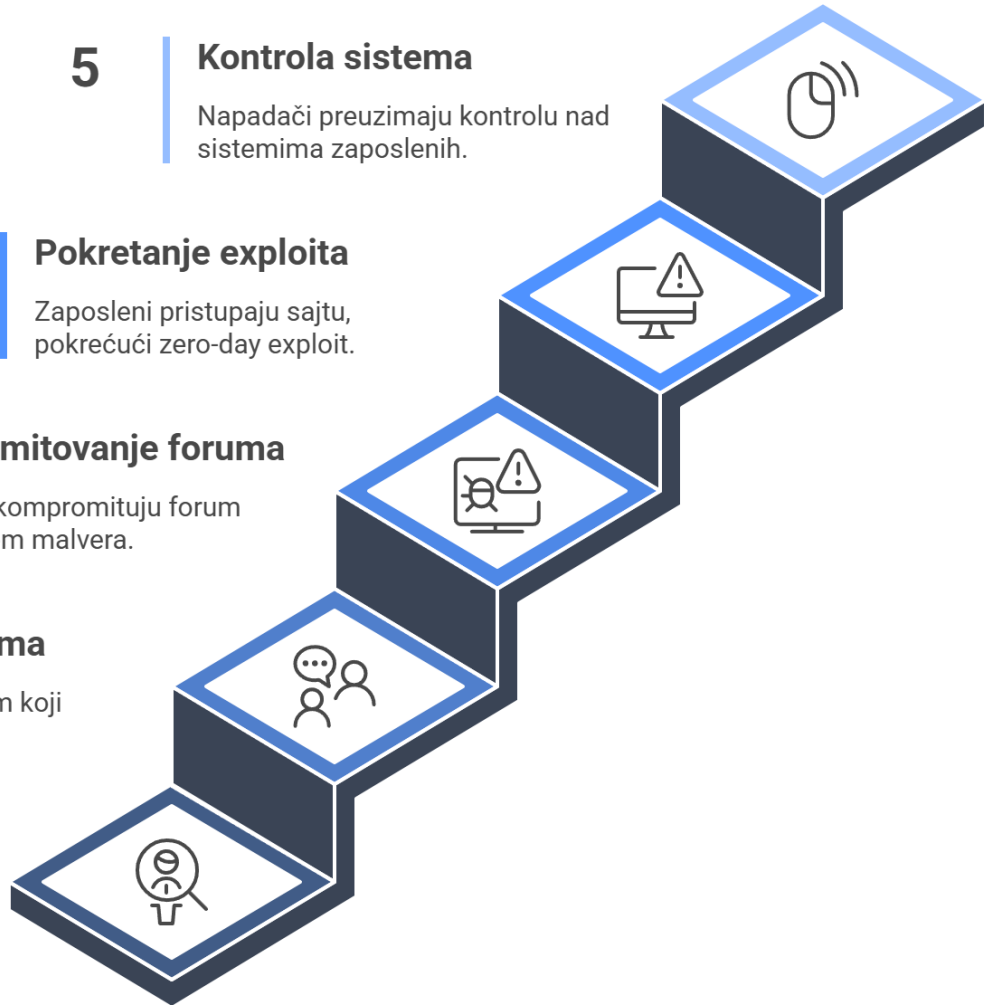
Napadači kompromituju forum ubacivanjem malvera.

4 **Pokretanje exploita**

Zaposleni pristupaju sajtu, pokrećući zero-day exploit.

5 **Kontrola sistema**

Napadači preuzimaju kontrolu nad sistemima zaposlenih.

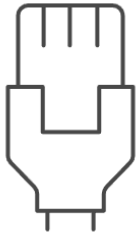


PRIMENA VEŠTAČKE INTELIGENCIJE U ODBRANI

VRSTE MREŽNIH BARIJERA (FIREWALL)

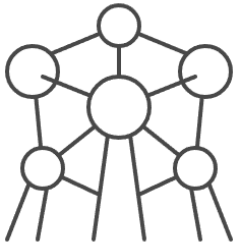
Firewall-ovi na nivou hosta

Nalaze se na krajnjim uređajima, integrisani u operativne sisteme.



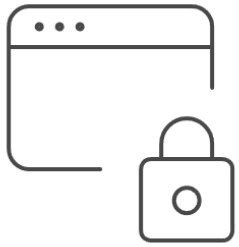
Mrežni firewall-ovi

Hardverski uređaji koji filtriraju saobraćaj na nivou mreže.



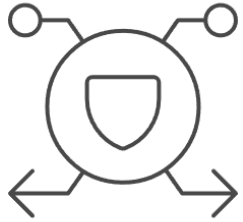
Next Generation Firewall

Napredni firewall sa mogućnostima dubinske inspekcije paketa.



Firewall-ovi za zaštitu veb aplikacija

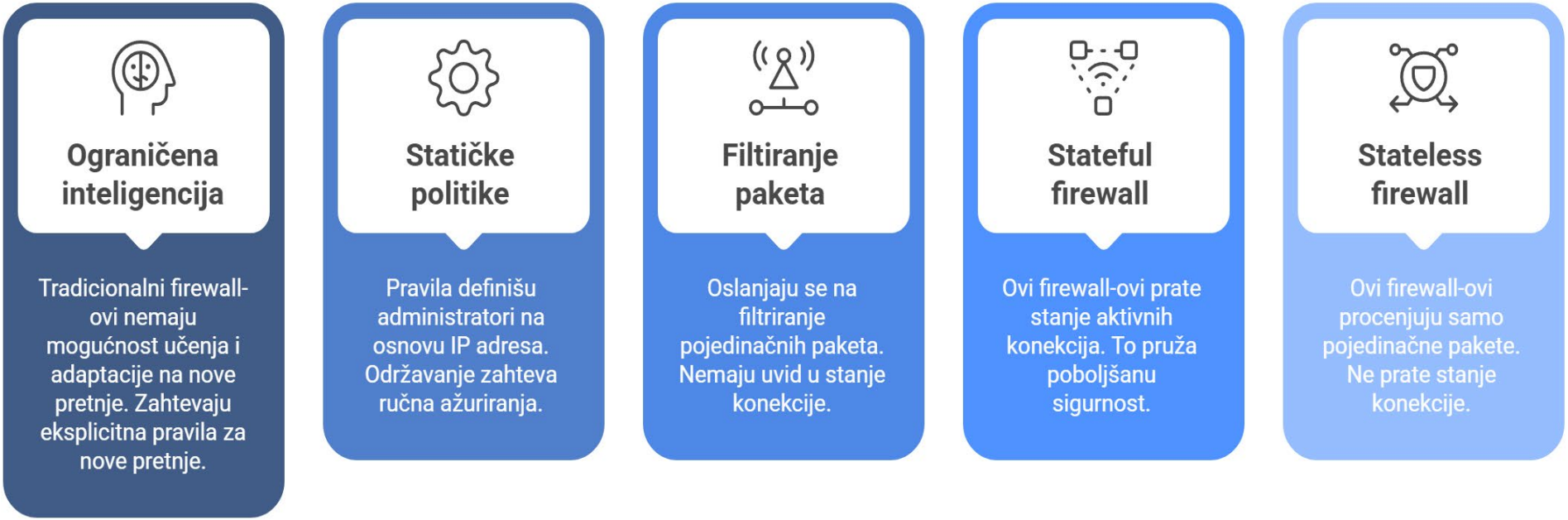
Štite veb aplikacije od zlonamernih napada.



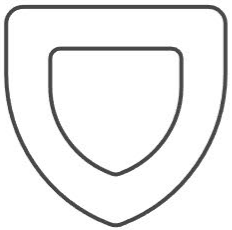
Sistem objedinjene pretnje

Kombinuje više bezbednosnih funkcija u jedan sistem.

TRADICIONALNA FIREWALL ZAŠTITA

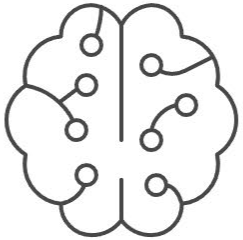


FIREWALL ZAŠTITA SA UKLJUČENOM VEŠTAČKOM INTELIGENCIJOM



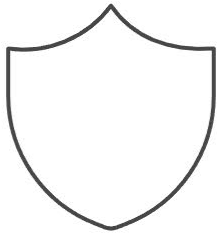
Dinamični i adaptivni

Koriste AI za prilagođavanje pretnjama.



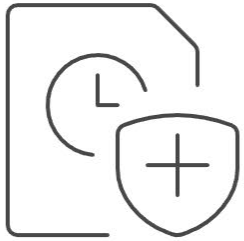
Automatsko učenje

Uče iz istorijskih podataka, ažuriraju automatski pravila



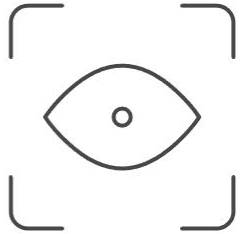
Zaštita od nepoznatih pretnji

Bolje su u zaštiti od nepoznatih napada. jer ne zavise isključivo od poznatih potpisa



Zaštita u realnom vremenu

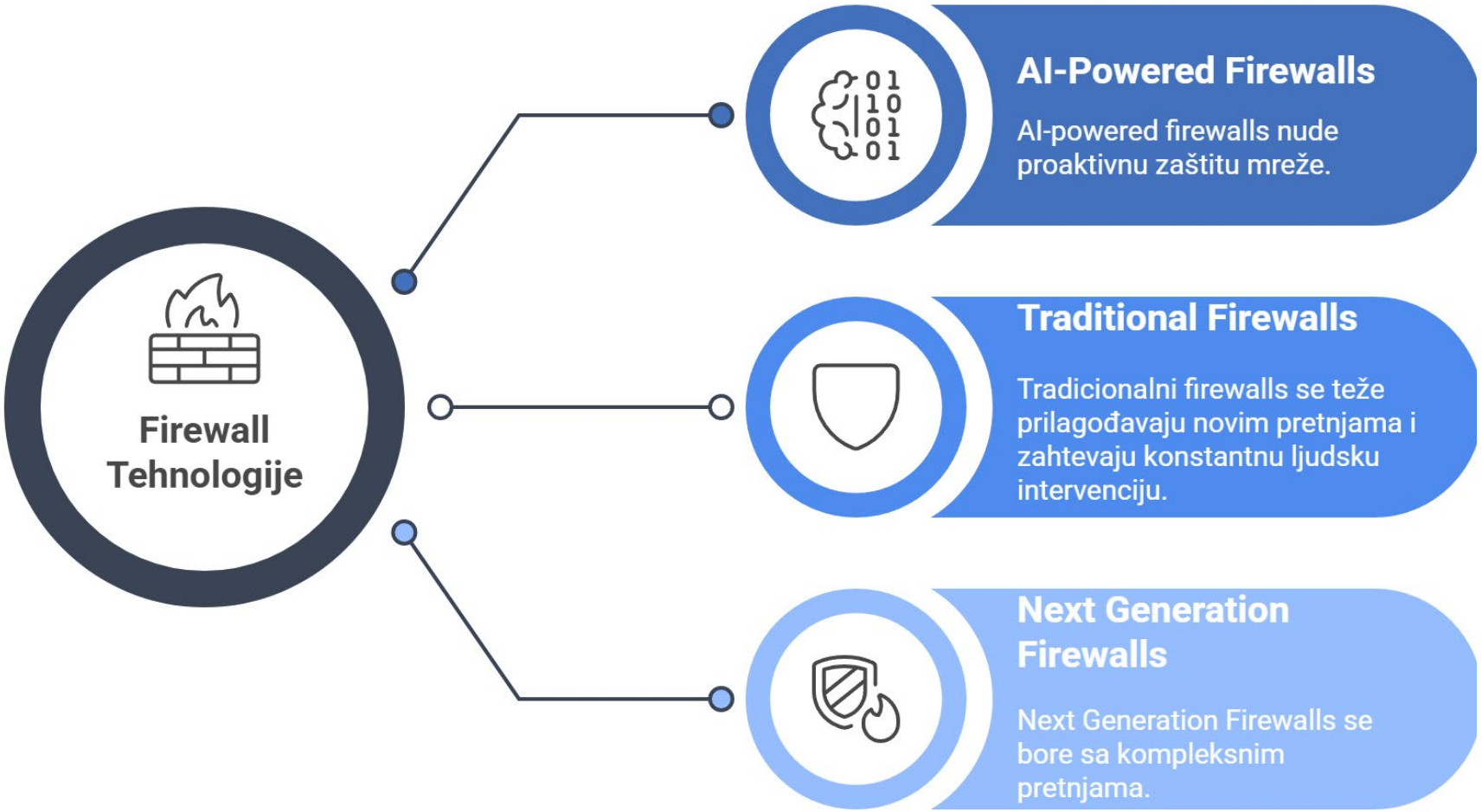
Automatski otkrivaju i reaguju na pretnje.



Svesni konteksta

Razmatraju komunikacione sesije i sadržaj podataka.

FIREWALL + AI


Firewall Tehnologije
AI-Powered Firewalls
AI-powered firewalls nude proaktivnu zaštitu mreže.
Traditional Firewalls
Tradicionalni firewalls se teže prilagođavaju novim pretnjama i zahtevaju konstantnu ljudsku intervenciju.
Next Generation Firewalls
Next Generation Firewalls se bore sa kompleksnim pretnjama.

VAŽNOST FILTRIRANJA MEJLOVA

Fišing napad

Najčešći vid sajber pretnje

Vektor napada

Uglavnom se obavlja putem elektronske pošte

Prevencija

Filtriranje omogućava zaštitu korisnika

INDIKATORI ZA EMAIL PRETNJE

Zlonamerne IP adrese i URL-ovi

Identifikuje i blokira opasne internet adrese

Nedosledna gramatika i pravopis

Prepoznaje e-poštu sa greškama koje ukazuju na spam

Sumnjive ključne reči

Otkriva i označava potencijalno štetne reči

Prekomerna upotreba specijalnih karaktera

Označava e-poštu sa previše neobičnih simbola i emotikona

Nepouzdani prilozi

Sprečava otvaranje potencijalno štetnih datoteka

ALGORITMI ZA IDENTIFIKACIJU PRETNJE

Identifikacija crvenih zastavica

Prepoznavanje potencijalnih indikatora spama

Algoritamska procena

Algoritmi analiziraju e-poštu za crvene zastavice

Različite interpretacije

Različiti algoritmi mogu različito tumačiti zastavice

ULOGA AI U FILTRIRANJU E-POŠTE

Detekcija spama

Identifikuje i filtrira neželjene mejlove

Detekcija fišinga

Otkriva i blokira prevare putem mejlova

Detekcija malvera

Otkriva i neutrališe zlonamerne fajlove u mejlovima

Kategorizacija mejlova

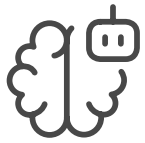
Organizuje mejlove u relevantne kategorije

Prioritizacija mejlova

Određuje važnost mejlova za pažnju

Analiza sadržaja

Analizira sadržaj mejlova za relevantne informacije



Prilagodljivo učenje

Poboljšava tačnost filtriranja tokom vremena

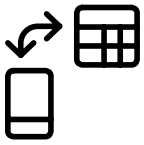
AI ALGORITMI ZA FILTRIRANJE E-POŠTE

Algoritam zasnovan na Sličnosti

Upoređuje dolazne mejlove sa prethodno pohranjenim mejlovima

Algoritam zasnovan na Uzorcima

Koristi šablone spam mejlova za procenu novih poruka



Adaptivni Algoritam

Prilagođava se i menja kategorije podataka vremenom

RIZICI AI FILTRIRANJA E-POŠTE

AI filtriranje nosi određene rizike – može doći do pogrešnog označavanja sigurnih mejlova kao opasnih i obrnuto.

Fišing mejlovi često pokušavaju da izgledaju kao da dolaze iz pouzdanih izvora, pa dobro osmišljene poruke mogu proći kroz filter.

Lažno pozitivni rezultati (false positive)

Označavanje legitimnih mejlova kao spam

Lažno negativni rezultati (false negative)

Dozvoljavanje fišing mejlovima da prođu