Dvojaka uloga umjetne inteligencije u informacijskoj sigurnosti kao izvora prijetnje i odgovora na prijetnju

Crnogorac, Snježana

Sveučilište u Zagrebu

Ekonomski fakultet

Specijalistički diplomski stručni studij

Elektroničko poslovanje u privatnom i javnom sektoru

DVOJAKA ULOGA UMJETNE INTELIGENCIJE U INFORMACIJSKOJ SIGURNOSTI KAO IZVORA PRIJETNJE I ODGOVORA NA PRIJETNJU

Diplomski rad

Snježana Crnogorac

Zagreb, srpanj 2023.

Sveučilište u Zagrebu

Ekonomski fakultet

Specijalistički diplomski stručni studij

Elektroničko poslovanje u privatnom i javnom sektoru

DVOJAKA ULOGA UMJETNE INTELIGENCIJE U INFORMACIJSKOJ SIGURNOSTI KAO IZVORA PRIJETNJE I ODGOVORA NA PRIJETNJU

Diplomski rad

Student: Snježana Crnogorac

JMBAG studenta: 0661050194

Mentor: prof. dr. sc. Mario Spremić

Zagreb, srpanj 2023.

IZJAVA O AKADEMSKOJ ČESTITOSTI

Izjavljujem i svojim potpisom potvrđujem da je završni/diplomski/poslijediplomski specijalistički rad, odnosno doktorski rad isključivo rezultat mog vlastitog rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu, a što pokazuju korištene bilješke i bibliografija.

Izjavljujem da nijedan dio rada nije napisan na nedozvoljen način, odnosno da je prepisan iz necitiranog rada, te da nijedan dio rada ne krši bilo čija autorska prava.

Izjavljujem, također, da nijedan dio rada nije iskorišten za bilo koji drugi rad u bilo kojoj drugoj visokoškolskoj, znanstvenoj ili obrazovnoj ustanovi.

Zagreb, 29.08.2023.

(mjesto i datum)

(vlastoručni potpis studenta)

STATEMENT ON THE ACADEMIC INTEGRITY

I hereby declare and confirm by my signature that the final / graduate / postgraduate specialist work or doctoral thesis is the sole result of my own work based on my research and relies on the published literature, as shown in the listed notes and bibliography.

I declare that no part of the work has been written in an unauthorized manner, i.e., it is not transcribed from the non-cited work, and that no part of the work infringes any of the copyrights.

I also declare that no part of the work has been used for any other work in any other higher education, scientific or educational institution.

Zagreb, 29.08.2023.

(Place and date)

CmoS.

(Personal signature of the student)

SAŽETAK i KLJUČNE RIJEČI

Predmet ovoga rada je razmatranje dvojake uloge umjetne inteligencije u informacijskoj sigurnosti. Radom se nastoji dokazati tezu da je upravo umjetna inteligencija tehnologija čije karakteristike podjednako predstavljaju izvor prijetnji, kao i odgovor na prijetnje.

Ciljevi rada:

C1. Identificirati karakteristike umjetne inteligencije koje je čine značajnim izvorom prijetnje i uspješnim odgovorom na prijetnju informacijskoj sigurnosti,

C2. Ispitati mogućnosti umjetne inteligencije da donese unaprjeđenja u točnosti pri detekciji prijetnji, skraćenju vremena istraživanja prijetnji, automatizaciji odgovora i implementaciji proaktivnih mehanizama zaštite,

C3. Ocijeniti mogućnosti umjetne inteligencije da se nosi s izazovima s kojima su suočeni stručnjaci iz područja informacijske sigurnosti (previše zadataka i podataka, ograničeno vrijeme i vještine),

C4. Donijeti zaključke o opravdanosti ulaganja u zaštitne mehanizme temeljene na umjetnoj inteligenciji te rizicima koji mogu proisteći iz neiskorištavanja potencijala umjetne inteligencije u ovome području.

U istraživanju se koriste kvalitativne metode istraživanja te se prikupljaju isključivo sekundarni podaci. Osnovna istraživačka metoda kojom će se ostvariti navedeni ciljevi rada jest analiza recentne znanstvene i stručne literature te sinteza rezultata te analize. Ostale metode dokazivanja su analiza dostupnih studija slučaja i analiza izvješća o učinkovitosti komercijalnih sigurnosnih rješenja.

Analizom dostupne znanstvene i stručne literature zaključeno je da svi kibernetički napadi koji koriste AI, kao i svi obrambeni mehanizmi temeljeni na AI, dijele specifične karakteristike te su uočena preklapanja između karakteristika koje predstavljaju izvor prijetnji i karakteristika koje čine AI učinkovitim odgovorom na prijetnje.

Mogućnosti umjetne inteligencije da donese unaprjeđenja u točnosti pri detekciji prijetnji, skraćenju vremena istraživanja prijetnji, automatizaciji odgovora i implementaciji proaktivnih mehanizama zaštite su značajne. Značajne su i mogućnosti AI da se nosi s izazovima s kojima su suočeni stručnjaci iz područja informacijske sigurnosti, s izuzetkom potpuno autonomnog donošenja sigurnosnih odluka.

Dokazano je da ulaganja u AI u području sigurnosti organizacijama donose različite kvantificirane, ali i nekvantifirane koristi, no AI nije rješenje koje odgovara svim organizacijama. U kojoj mjeri

AI unaprjeđuje informacijsku sigurnost ovisi o tome koliko efikasno organizacije upravljaju svojim AI rješenjima.

Ključne riječi: umjetna inteligencija, AI, strojno učenje, kibernetički napadi, obrambeni mehanizmi

ABSTRACT AND KEYWORDS

The aim of this paper is to consider the dual role of artificial intelligence in information security. The paper tries to prove the thesis that artificial intelligence is a technology whose characteristics are equally a source of threats, as well as a response to threats.

Objectives of the work:

C1. Identify the characteristics of artificial intelligence that make it a significant source of threat and a successful response to threats to information security,

C2. Explore the capabilities of artificial intelligence to deliver improvements in threat detection accuracy, reduce threat investigation time, automate responses, and implement proactive protection mechanisms,

C3. Assess the capabilities of artificial intelligence to deal with the challenges information security professionals are faced with (too many tasks and data, limited time and skills),

C4. Draw conclusions about the justification of investing in protective mechanisms based on artificial intelligence and the risks that may arise from not using its potential in cyber security.

Qualitative research methods are used and only secondary data is collected. The basic research method used to achieve the stated objectives is the analysis of recent scientific and professional literature and the synthesis of the obtained results. Other methods of research are the analysis of case studies and reports on the effectiveness of commercial security solutions.

The analysis of the available scientific and professional literature concluded that all cyber attacks that use AI, as well as all defense mechanisms based on AI, share specific characteristics, and overlaps were observed between the characteristics that represent the source of threats and the characteristics that make AI an effective response to threats.

The potential for artificial intelligence to deliver improvements in threat detection accuracy, reduce threat research time, automate responses, and implement proactive protection mechanisms is significant. The possibilities of AI to cope with the challenges faced by security experts, with the exception of fully autonomous security decision-making, are also significant.

Investments in AI in the field of security have been proven to bring various quantified and unquantified benefits to organizations, but AI is not a solution that fits all organizations. The extent to which AI improves information security depends on how effectively organizations manage their AI solutions.

Keywords: artificial intelligence, AI, machine learning, cyber attacks, defense mechanisms

Sadržaj

1. UVOD

1.1. Predmet i cilj rada

Zadržavanje konkurentnosti i opstanak na tržištu današnjice zahtjevaju digitalnu transformaciju poslovanja. Intenzivna primjena digitalnih tehnologija iziskuje ne samo prilagodbu dinamičnom poslovnom okruženju, već i nužne promjene poslovnih modela. Kompanije s potpuno novim, inovativnim poslovnim modelima, u mogućnosti su vrlo brzo mijenjati svoje poslovne strategije, ali i prilagođavati se konkurenciji i zadržavati čvrstu poziciju na tržištu, stvarajući proizvode ili usluge koje imaju disruptivan utjecaj na način poslovanja i kvalitetu života.

U svom govoru kao predsjednik Društva za povijest tehnologije Melvin Kranzberg je 1985. godine izrekao tvrdnju koju je prozvao prvi zakon tehnologije: "Tehnologija nije ni dobra ni loša; niti je neutralna" (Kranzberg, 1985). Uz dokazan pozitivan utjecaj na inovativnost te stvaranje poslovnih i društvenih vrijednosti, intenzivna primjena informacijskih, a posebno novih digitalnih tehnologija, organizacije i pojedince izlaže i sasvim novim i neočekivanim *cyber* rizicima. Ako znamo da su ti rizici stalno prisutni i neizbježni, a bolje upravljanje njima čuva vrijednost ne samo informatičkih ulaganja, već i cjelokupnog poslovanja (Spremić, 2017), tada možemo zaključiti da u fokusu *cyber* sigurnosti više ne može biti samo identificiranje prijetnji, već njihovo predviđanje i sprječavanje.

U vremenu u kojem je ulaganje u digitalne tehnologije nužnost, razborito je razmotriti mogu li nas iste tehnologije koje nas rizicima izlažu, ujedno i zaštiti. Predmet ovoga rada je razmatranje dvojake uloge umjetne inteligencije u informacijskoj sigurnosti. Radom se nastoji dokazati tezu da je upravo umjetna inteligencija tehnologija čije karakteristike podjednako predstavljaju izvor prijetnji, kao i odgovor na prijetnje u opisanome poslovnom okruženju.

Ciljevi rada:

C1. Identificirati karakteristike umjetne inteligencije koje je čine značajnim izvorom prijetnje i uspješnim odgovorom na prijetnju informacijskoj sigurnosti

C2. Ispitati mogućnosti umjetne inteligencije da donese unaprjeđenja u točnosti pri detekciji prijetnji, skraćenju vremena istraživanja prijetnji, automatizaciji odgovora i implementaciji proaktivnih mehanizama zaštite

C3. Ocijeniti mogućnosti umjetne inteligencije da se nosi s izazovima s kojima su suočeni stručnjaci iz područja informacijske sigurnosti (previše zadataka i podataka, ograničeno vrijeme i vještine)

C4. Donijeti zaključke o opravdanosti ulaganja u zaštitne mehanizme temeljene na umjetnoj inteligenciji te rizicima koji mogu proisteći iz neiskorištavanja potencijala umjetne inteligencije u ovome području.

1.2. Metode istraživanja i izvori podataka

Rad je rezultat samostalnog istraživanja studenta. U istraživanju se koriste kvalitativne metode istraživanja te se prikupljaju isključivo sekundarni podaci.

Osnovna istraživačka metoda kojom će se ostvariti navedeni ciljevi rada, a posebno donijeti ocjene mogućnosti umjetne inteligencije da se nosi s izazovima informacijske sigurnosti te zaključci o isplativosti ulaganja u zaštitne mehanizme temeljene na umjetnoj inteligenciji i rizicima neulaganja, jest analiza recentne znanstvene i stručne literature te sinteza rezultata te analize. Ostale metode dokazivanja su analiza dostupnih studija slučaja i analiza izvješća o učinkovitosti komercijalnih sigurnosnih rješenja.

1.3. Sadržaj i struktura rada

U uvodnome poglavlju iznose se tema rada i osnovna istraživačka pitanja, opisuju se metode prikupljanja i analize podataka te se iznosi struktura i sadržaj rada.

U poglavlju 2 iznose se teorijske spoznaje o umjetnoj inteligenciji, njezinom razvoju i osnovnim granama, sa naglaskom na one relevantne za područje informacijske sigurnosti (*Machine learning, Deep learning, Computer vision, Natural language processing*). U poglavlju 3 iznose se teorijske spoznaje o ulozi umjetne inteligencije u digitalnoj transformaciji poslovanja i stvaranju disruptivnih poslovnih modela, uz razmatranje prednosti i rizika koji iz njezine primjene proizlaze.

U poglavlju 4 prikazuju se rezultati analize onih karakteristika umjetne inteligencije koje je čine izvorom prijetnje te se razmatraju metode kibernetičkih napada temeljenih na umjetnoj inteligenciji (*AI-powered phishing attacks, Data Poisoning, Deepfakes, Automated Exploit Generation, GAN-based attacks, Adversarial attacks* itd.) i njihove implikacije na poslovanje, a temeljem rezultata provedene analize znanstvene literature i dostupnih studija slučaja.

U poglavlju 5 prikazuju se rezultati analize onih karakteristika umjetne inteligencije koje je čine tehnologijom pogodnom za razvoj obrambenih mehanizama te se opisuju načini implementacije tih mehanizama (*Endpoint Protection Platforms, Self-configuring Networks, Intent-based Network Security, Cognitive Security* itd.) i iznosi ocjena njihove djelotvornosti temeljem rezultata provedene analize znanstvene literature i dostupnih studija slučaja. U ovome poglavlju razmatra se i nužnost zaštite samih obrambenih rješenja temeljenih na umjetnoj inteligenciji, kao preduvjet njihove djelotvornosti.

U zaključnom poglavlju iznose se najvažniji rezultati navedenih metoda istraživanja, daju odgovori na istraživačka pitanja te nove spoznaje autora o istraživačkoj temi.

2. UMJETNA INTELIGENCIJA

2.1. Pojam umjetne inteligencije

Ne postoji univerzalno prihvaćena akademska definicija inteligencije. Prema jednoj od mnogih definicija, koju je iznio 1979. Lloyd G. Humphreys, inteligencija je rezultanta procesa stjecanja, pronalaženja i pohranjivanja u pamćenje, a potom kombiniranja, uspoređivanja i korištenja informacija i konceptualnih vještina u novim kontekstima (Humphreys, 1979). Drugu definiciju inteligencije su 1994. godine potpisala pedeset i dva istraživača toga znanstvenog područja, a 3 godine kasnije u svom editorijalu u časopisu Intelligence iznijela ju je Linda S. Gottfredson. Inteligencija je vrlo uopćena, široko definirana mentalna sposobnost koja, između ostaloga, uključuje sljedeće misaone procese: sposobnost planiranja i rješavanja problema, apstraktnog razmišljanja, poimanja složenih ideja i zaključivanja, kao i brzog učenja te učenja iz iskustva. Inteligencijom ne možemo smatrati puko učenje iz knjiga, uske akademske vještine ili vještine polaganja ispita, već širu i dublju sposobnost razumijevanja okoline- to je način na koji pojedinac shvaća stvari, pridaje im smisao ili smišlja što učiniti (Gottfredson, 1997).

Sasvim očekivano, ne postoji ni precizna, univerzalno prihvaćena definicija umjetne inteligencije, a neki teoretičari smatraju da joj je upravo to omogućilo propulzivan razvoj.

Umjetna inteligencija (engl. *Artificial Intelligence- AI*) se obično definira kao grana računalne znanosti koja se bavi razvojem algoritama i tehnika koje mogu simulirati ili čak replicirati sposobnosti ljudskog uma (Vassilopoulos i Georgopoulos, 2010).

Postoji, dakle, tendencija da se umjetna inteligencija opisuje na antropomorfan način, način koji je nužno povezuje s čovjekom, njegovim mozgom, živčanim sustavom, voljom, savješću, a u novije vrijeme čak i emocijama (Afsah, 2022). Međutim, već Nils Nilsson predložio je definiciju koja inteligenciju strojeva ne povezuje nužno sa ljudskom.

Naime, umjetna inteligencija jest posvećena tome da strojeve učini inteligentnima, ali nisu samo ljudi inteligentan entitet. Inteligentan je svaki entitet koji primjereno i s predviđanjem funkcionira u svom okruženju (Nilsson, 2010). To mogu biti ljudi, životinje i neki strojevi. Postoji prošireni kontinuum ili spektar na kojem su poredani inteligentni entiteti, s time da se strojevi i mnoge životinje nalaze na primitivnom kraju tog kontinuuma, dok su ljudi na njegovom drugom kraju. Ljudi su sposobni reagirati na senzorne inpute, opažati, sintetizirati i sažimati informacije, rasuđivati, postizati ciljeve, razumjeti i generirati jezik, dokazivati matematičke teoreme, igrati izazovne igre, stvarati umjetnost i glazbu, pa čak i pisati povijest. Budući da funkcioniranje na odgovarajući način i s predviđanjem zahtijeva različite sposobnosti, ovisno o okruženju, mogli bismo reći da postoji i više različitih kontinuuma inteligencije te da ni na jednome od njih nema oštrih ili jasnih prekida (Nilsson, 2010).

Stanfordova Stogodišnja studija o umjetnoj inteligenciji također naglašava da je ljudska inteligencija bila samo inspiracija za tehnologiju za koju je malo vjerojatno da će kopirati ljudski mozak. Prema studiji, različiti oblici inteligencije (ljudska, životinjska ili strojna) ne razlikuju se po vrsti, već po stupnju skalabilnosti, brzine, autonomije i općenitosti (Stone et al., 2016). Mogli bismo reći da ne postoje različite vrste inteligencije, već širok spektar inteligencije u kojem ljudski mozak nema istaknuto mjesto, osim što, zbog svoje svestranosti, ostaje logičan izbor za vrednovanje napretka umjetne inteligencije. Studija definira umjetnu inteligenciju kao znanost i skup računalnih tehnologija koje su nadahnute ljudskom inteligencijom, ali obično djeluju sasvim drugačije od načina na koji ljudi osjećaju, uče, razmišljaju i poduzimaju akcije (Stone et al., 2016).

U lipnju 2023. godine u svome gostovanju na podcastu Diary of a CEO, Mo Gawdat, bivši Chief Business Officer Google X, otišao je još korak dalje, rekavši da je strojna inteligencija nastala tek kada smo joj prestali nametati ljudsku inteligenciju. Umjetna inteligencija se razvija upravo tako da se računalu ne daju rješenja problema do kojih je došao čovjek, već mu se prepusti da rješenje nađe samo. Po tome se njezin razvoj razlikuje od razvoja dotadašnjeg softvera, gdje je čovjek prvo riješio problem, a potom dao instrukciju računalu da problem rješava na isti način (Gawdat, 2023).

Definiciju umjetne inteligencije moguće je operacionalizirati ako je izvedemo iz posla kojim se bave njezini istraživači. Stanfordova studija definira umjetnu inteligenciju prvenstveno kao granu računalne znanosti koja proučava svojstva inteligencije sintetiziranjem inteligencije (Stone et al., 2016). Iako je pojava umjetne inteligencije ovisila o brzom napretku hardverskih računalnih resursa, u fokusu istraživačke zajednice je donedavno bio softver. Tek u novije vrijeme, napredak u izgradnji hardvera skrojenog za računalstvo temeljeno na neuronskim mrežama odgovarajuću važnost za napredovanje umjetne inteligencije daje vezi između hardvera i softvera (Stone et al., 2016).

Jedna od najvažnijih karakteristika umjetne inteligencije je svakako sposobnost učenja. Ona omogućava strojevima da prepoznaju obrasce, donose zaključke, uče iz iskustva i poboljšavaju se samostalno (Alawadhi et al., 2022).

Učenje strojeva može biti nadzirano, nenadzirano ili pojačano. Nadzirano učenje se koristi za prepoznavanje obrazaca u skupovima podataka koji su već označeni, dok nenadzirano učenje koristi algoritme koji se samostalno prilagođavaju bez izričite upute za učenje. Pojačano učenje je učenje kroz interakciju s okolinom, gdje računalni program dobiva povratne informacije o svojim postupcima i odlukama (Burgard, 2022).

Inteligencija ostaje složen fenomen čiji su različiti aspekti privlačili pažnju mnogih grana znanosti, uključujući psihologiju, ekonomiju, neuroznanost, biologiju, inženjerstvo, statistiku i lingvistiku (Stone et al., 2016). U sljedećem poglavlju bit će prikazano kako se zajednica istraživača umjetne inteligencije u njezinom razvoju koristila napretkom svih ovih znanstvenih grana, a potom je umjetna inteligencija u većini njih našla svoju primjenu.

Naime, sposobnost sustava umjetne inteligencije da donose zaključke, predviđaju trendove i rješavaju probleme na vrlo složen način, daje im potencijal da dramatično promijene način na koji

ljudi obavljaju poslove i komuniciraju sa strojevima, u različitim situacijama i industrijama, kao što su vojna industrija, medicina, financije, oglašavanje, znanost i druge.

2.2. Razvoj umjetne inteligencije

John McCarthy i suautori napisali su "Prijedlog za Ljetni istraživački projekt o umjetnoj inteligenciji u Darthmouthu, 31. kolovoza 1955.", u kojem iznose tezu da se svaki vid učenja ili bilo koja druga odrednica inteligencije može tako precizno opisati da ju je moguće strojno simulirati (McCarthy et al., 2006). Upravo McCarthyju se pripisuje prva uporaba izraza "umjetna inteligencija" u navedenome Prijedlogu te se smatra da je na njegovoj radionici na Ljetnom istraživačkom projektu 1956. godine službeno "rođeno" ovo istraživačko područje.

Mnogi od sudionika njegove radionice (uključujući Arthura Samuela, Olivera Selfridgea, Raya Solomonoffa, Allena Newella i Herberta Simona) pokrenuli su prve značajne projekte "pod zastavom" umjetne inteligencije (u daljnjem tekstu: AI), dajući na taj način novome području identitet i posvećenu istraživačku zajednicu (Stone et al., 2016).

No, mnoge izvorišne ideje ili tehnički preduvjeti razvoja umjetne inteligencije postojali su mnogo ranije.

U osamnaestom stoljeću, Thomas Bayes pružio je okvir za rasuđivanje o vjerojatnosti događaja. U devetnaestom stoljeću, George Boole je pokazao da se logičko rasuđivanje, koje datira još od Aristotela, može izvoditi sustavno, na isti način kao i rješavanje sustava jednadžbi. Do prijelaza u dvadeseto stoljeće, napredak u eksperimentalnim znanostima doveo je do pojave statistike, znanstvene grane koja omogućuje izvođenje zaključaka iz podataka.

Ideja o fizičkom konstruiranju stroja za izvršavanje nizova instrukcija, koja je zaokupila maštu pionira poput Charlesa Babbagea, sazrjela je 1950-ih i rezultirala konstrukcijom prvih elektroničkih računala. Do tog vremena, razvijeni su i prvi primitivni roboti, koji su mogli percipirati stvarnost i autonomno djelovati.

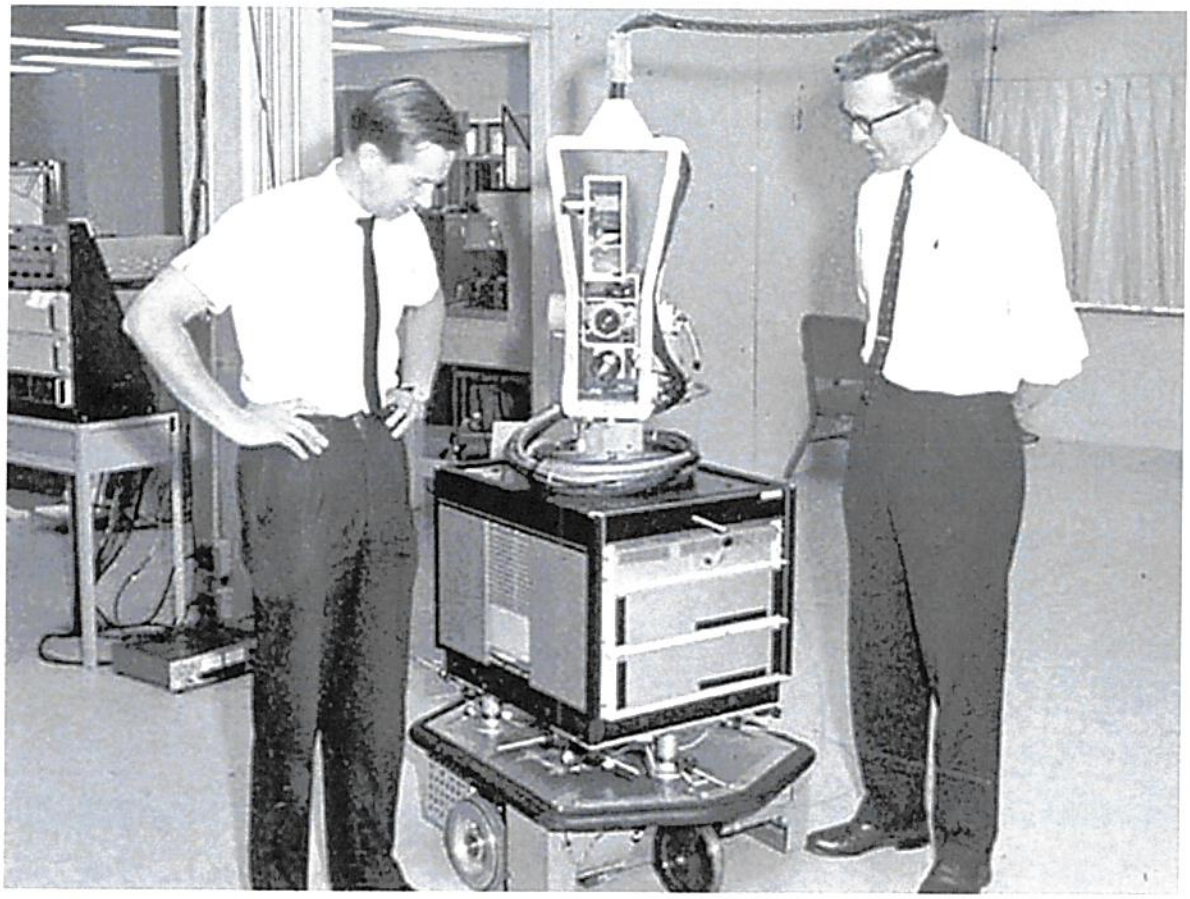
U radu "Računalna tehnologija i inteligencija" Alan Turing još 1950. otvara mogućnost stvaranja računala za simulaciju inteligencije i istražuje mnoge karakteristike danas povezane s umjetnom inteligencijom, uključujući kako se inteligencija može testirati i kako strojevi mogu automatski učiti. On je predložio Turingov test kao način provjere inteligencije računala. Ovaj test i dalje predstavlja važan standard za procjenu sposobnosti AI, a prema njemu, ako osoba ne može razlikovati odgovor računala od odgovora osobe, računalo se smatra inteligentnim.

Između 1950-ih i 1970-ih pojavilo se nekoliko novih područja ključnih za razvoj umjetne inteligencije.

Newell i Simon bili su pioniri heurističkog pretraživanja, učinkovitog postupka za pronalaženje odgovora u velikim, kombinatornim prostorima. Primijenili su ga za konstruiranje dokaza matematičkih teorema kroz svoje programe Logic Theorist i General Problem Solver (Stone et al., 2016).

U području računalnog vida, rani rad Selfridgea na prepoznavanju znakova postavio je osnovu za kasnije složenije primjene kao što je prepoznavanje lica. Do kasnih šezdesetih započeo je rad i na obradi prirodnog jezika.

Shakey, robot na kotačićima izgrađen u SRI Internationalu, pokrenuo je područje mobilne robotike (slika 1.). Samuelov program za igranje dame (engl. *Checkers*), koji se poboljšavao samostalnom igrom, bio je jedan od prvih primjera sustava strojnog učenja.



Slika 1. Nils Nilsson i Sven Wahlstrom s robotom Shakey-em 1960-ih
Izvor: SRI INTERNATIONAL, preuzeto s: https://spectrum.ieee.org/sri-shakey-robot-honored-as-ieee-milestone

Rosenblattov Perceptron, računalni model temeljen na biološkim neuronima, postao je osnova za polje umjetnih neuronskih mreža. Feigenbaum je zagovarao izgradnju ekspertnih sustava, repozitorija znanja skrojenih za specijalizirana područja, kao što su kemija i medicinska dijagnostika (Stone et al., 2016).

Sav ovaj rani konceptualni razvoj umjetne inteligencije pretpostavljao je postojanje simboličkog sustava o kojem se može razmišljati i na kojem se može graditi, dok je nedovoljan naglasak unutar AI zajednice pridavan samoj izgradnji sustava, s izravnim pristupom signalima i podacima iz njegova okruženja. Došlo je i do prenaglašavanja Booleova (točno/netočno) logičkog modela, previđajući potrebu za kvantificiranjem neizvjesnosti (Stone et al., 2016). Iz ovih razloga, do 1980-

ih se polje umjetne inteligencije nije moglo pohvaliti značajnijim praktičnim primjenama, dok se istovremeno financiranje smanjivalo, a interes počeo opadati. Nilsson ovo razdoblje naziva "AI zima" (Nilsson, 2010).

U devedesetima javlja se ideja da je inteligentne sustave vođene podacima iz stvarnog svijeta potrebno graditi iz temelja, a tehnološki napredak, jeftiniji i pouzdaniji hardver učinili su to mogućim. Sposobnost interneta za prikupljanje velikih količina podataka te dostupnost računalnih resursa za pohranu i obradu tih podataka, omogućili su razvoj statističkih tehnika koje izvode rješenja iz podataka (Stone et al. 2016). U ovome desetljeću, razvoj AI usmjeren je na razvoj sustava koji se temelje na neuronskim mrežama.

Nekoliko je čimbenika potaknulo pravu AI revoluciju. Najistaknutiji među njima je sazrijevanje strojnog učenja, dodatno podržano prikupljanjem podataka putem interneta i resursima računalstva u oblaku. Strojno učenje je dramatično unaprijeđeno dubokim učenjem, oblikom adaptivne umjetne neuronske mreže uvježbane korištenjem metode zvane *backpropagation*. Ovaj nagli napredak u izvedbi algoritama je popraćen napretkom u hardverskoj tehnologiji za operacije poput percepcije i prepoznavanja predmeta. Nove platforme i tržišta za proizvode temeljene na podacima te ekonomski poticaji za pronalazak dodatnih proizvoda i tržišta također su doprinijeli razvoju AI tehnologija (Stone et al. 2016).

Možemo zaključiti: potaknuta utjecajem različitih čimbenika, poput povećanja računalnih kapaciteta, dostupnosti podataka i Velikih podataka, revolucije u razvoju algoritama i softvera, napretka u znanju o ljudskom mozgu i postojanja bogate tehnološke industrije sklone riziku (Afsah, 2022), spomenuta teza iz prijedloga za konferenciju u Darthmoutu prestala je biti aspiracija malog kruga znanstvenika i postala stvarnost.

Danas, kada AI tehnologije već prožimaju našu svakodnevicu i društvo, napori istraživača se sa jednostavne izgradnje sustava koji su inteligentni usmjeravaju na izgradnju inteligentnih sustava koji su svjesni ljudi, komuniciraju s ljudima te su za ljude sigurni i pouzdani.

2.3. Podjele umjetne inteligencije

Prema već spomenutoj Stanfordovoj Stogodišnjoj studiji o umjetnoj inteligenciji, AI se tradicionalno može podijeliti na nekoliko grana odnosno područja koja se bave različitim aspektima rješavanja problema. Neka od najvažnijih područja AI navedena su te potkrijepljena primjerima u nastavku.

Pretraživanje i planiranje (engl. *Search and Planning*) bave se razmatranjem ponašanja usmjerenih ka cilju (Stone et al., 2016). Primjer programa u kojima pretraživanje i planiranje igraju ključnu ulogu su programi za igranje šaha, poput Deep Blue, koji razmatraju koji će potezi (ponašanja) dovesti do pobjede (cilja).

Područje predstavljanja znanja i rasuđivanja (engl. *Knowledge Representation and Reasoning*) uključuje transformaciju velikih količina podataka u strukturirani oblik na koji se mogu postavljati pouzdaniji i učinkovitiji upiti. IBM-ov program Watson, koji je pobijedio ljudske natjecatelje u Jeopardy izazovu 2011. godine, uglavnom se temeljio na učinkovitim shemama za organiziranje, indeksiranje i dohvaćanje velikih količina podataka prikupljenih iz različitih izvora (Stone et al., 2016).

Razvojem algoritama, strojno učenje (engl. *Machine Learning*) je AI sustavima omogućilo automatsko poboljšanje njihove izvedbe promatranjem relevantnih podataka. Uz pomoć ovih algoritama, računala mogu prepoznavati obrasce u podacima i donositi odluke temeljene na tim obrascima. Kako je već spomenuto, strojno učenje je pridonijelo razvoju AI u posljednjih nekoliko desetljeća, od jednostavnih tražilica i programa za preporuku proizvoda, do sustava dubokog učenja (engl. *Deep Learning*) poput onih za prepoznavanje govora, otkrivanje prijevara, razumijevanje slika i drugih složenih zadataka koji su se nekoć oslanjali na ljudsku vještinu i prosudbu (Stone et al., 2016). Automatizacija ovih zadataka omogućila je povećanje opsega usluga poput e-trgovine.

Područje sustava s više sudionika (engl. *Multi-Agent Systems*) razmatra pitanje međusobnog djelovanja inteligentnih sustava, koje je posebno važno za online mjesta trgovanja i transportne sustave (Stone et al., 2016).

Područje robotike (engl. *Robotics*) istražuje temeljne aspekte opažanja i djelovanja, a posebno njihove integracije, koji omogućuju robotu da se učinkovito ponaša. Budući da roboti i drugi računalni sustavi dijele živi svijet s ljudima, stručno područje interakcije čovjeka i robota (engl. *Human-Robot Interaction*) je također postalo istaknuto u posljednjim desetljećima (Stone et al., 2016).

Strojna percepcija (engl. *Machine Perception*) uvijek je igrala središnju ulogu u umjetnoj inteligenciji, dijelom u razvoju robotike, ali i kao potpuno samostalno područje istraživanja. Najpoznatiji modaliteti strojne percepcije su računalni vid (engl. *Computer Vision*) i obrada prirodnog jezika (engl. *Natural Language Processing*) (Stone et al., 2016).

Nekoliko novijih područja interesa unutar AI javljaju se kao posljedica rasta interneta. Analiza društvenih mreža (engl. *Social Network Analysis*) istražuje utjecaj socijalnih struktura i odnosa na ponašanje pojedinaca i zajednica (Stone et al., 2016).

Crowdsourcing je još jedna inovativna tehnika rješavanja problema, koja se oslanja na iskorištavanje ljudske inteligencije (obično tisuća ljudi) za rješavanje složenih računalnih problema (Stone et al., 2016).

Navedena područja AI se međusobno preklapaju i često sudjeluju u razvoju složenih sustava koji se koriste u različitim industrijama i segmentima poslovanja, pa tako i u informacijskoj sigurnosti.

Budući da istraživanje AI nastoji razviti sustave i strojeve koji oponašaju funkcioniranje nalik ljudskom, stupanj do kojeg AI sustav uspješno replicira ljudske sposobnosti koristi se kao jedan od kriterija za razlikovanje vrsta umjetne inteligencije. Ovisno o tome kakvo je funkcioniranje stroja u usporedbi s ljudima, u pogledu svestranosti i performansi, pojedini AI sustav se može klasificirati pod jednu od vrsta AI, opisanih u nastavku.

Reaktivni strojevi (engl. *Reactive Machines*) su najstariji oblici AI sustava koji imaju ograničene mogućnosti. Oni oponašaju sposobnost ljudskog uma da reagira na različite vrste podražaja. Nemaju funkciju temeljenu na memoriji, ne mogu formirati sjećanja niti koristiti prethodno stečena iskustva da informiraju svoje sadašnje radnje. Nemaju, dakle, sposobnost "učiti", već se mogu koristiti samo za automatski odgovor na ograničeni skup ili kombinaciju ulaza. Popularan primjer reaktivnog AI stroja je IBM-ov Deep Blue, stroj koji je 1997. godine pobijedio šahovskog velemajstora Garryja Kasparova. Stroj je sposoban prepoznati figure na šahovskoj ploči i znati kakvo će biti njihovo kretanje. Može predvidjeti sljedeći potez svoga protivnika te odabrati svoje najoptimalnije poteze, ali nema koncept prošlosti niti sjećanje na prošle događaje. Mora poštovati pravilo protiv ponavljanja istog poteza tri puta, ali ne može uzeti u obzir odnosno ignorira ostale događaje iz prošlosti (Hintze, 2016).

Googleov proizvod AlphaGo, iako je pobjeđivao vrhunske stručnjake za igru Go, također ne može prognozirati sve moguće buduće poteze. Njegova metoda analize događaja je naprednija od Deep Blue-ove. Za procjenu razvoja igre on koristi tehniku neuronske mreže, no to mu daje prednost isključivo u igranju zadane igre te unutar dodijeljenih mu zadataka. Svoje metode analize on ne može lako promijeniti niti primijeniti na druge situacije. Naprotiv, svaki puta kada se nađu u nekoj situaciji, reaktivni strojevi ponašat će se na potpuno isti način, stoga ih je lako „prevariti" (Hintze, 2016). To je njihova prednost kada je važno osiguravanje pouzdanosti AI sustava, npr. pri razvoju autonomnih automobile, ali je nedostatak kada je potrebno da se strojevi istinski bave svijetom i odgovaraju na njega (Hintze, 2016).

Strojevi ograničene memorije (engl. *Limited Memory*) su strojevi koji, uz sposobnosti reaktivnih strojeva, imaju sposobnost učiti iz povijesnih podataka kako bi donosili odluke. Gotovo sve postojeće primjene AI spadaju u ovu kategoriju. Sustavi dubokog učenja treniraju se velikim količinama podataka za obuku, koje pohranjuju u svoju memoriju kako bi formirali referentni model za rješavanje budućih problema. Umjetna inteligencija za prepoznavanje slika obučava se pomoću tisuća slika za obuku i njihovih oznaka koje koristi kao reference za razumijevanje sadržaja

slike koja joj je predstavljena. Tako na temelju svog "iskustva učenja" označava nove slike s povećanom točnošću (Joshi, 2022).

Samovozeći automobili promatraju brzinu i smjer drugih automobila. Njihova opažanja dodaju se unaprijed programiranim prikazima stvarnoga svijeta, koji uključuju važne cestovne elemente poput oznaka traka, prometnih znakova, semafora, ali i zavoja na cesti. Samovozeći automobil mora raspoznati određene objekte, ne samo u trenutku, već i pratiti ih tijekom vremena. Na taj način zna kada treba promijeniti traku, da ne bi presjekao put drugom vozaču ili udario automobil u blizini. Međutim, ti jednostavni podaci o prošlosti su prolazni i ne spremaju se u memoriju automobila poput iskustva iz kojega on može učiti, kao što ljudski vozači skupljaju iskustvo tijekom godina vožnje (Hintze, 2016).

Dok prethodne dvije vrste umjetne inteligencije postoje i imaju brojne primjene, sljedeće dvije vrste umjetne inteligencije za sada postoje samo kao koncept ili rad u tijeku (Joshi, 2022).

Umjetna inteligencija teorije uma (engl. *Theory of Mind AI*) sljedeća je stepenica u razvoju AI sustava, na čijim su inovacijama istraživači trenutno angažirani. Umjetna inteligencija teorije uma moći će bolje razumjeti entitete s kojima je u interakciji, razlučujući njihove potrebe, emocije, uvjerenja i misaone procese. Iako je umjetna emocionalna inteligencija već industrija u nastajanju i trenutno područje interesa za vodeće istraživače AI, postizanje razine teorije uma zahtijevat će razvoj i u drugim granama AI. To je zato što će, kako bi doista razumjeli ljudske potrebe, strojevi morati percipirati ljude kao entitete čiji umovi mogu biti oblikovani višestrukim čimbenicima (Joshi, 2022). Za potpunu i sigurnu interakciju s ljudima, AI sustavi morat će moći razumjeti da ljudi imaju misli, osjećaje i očekivanja od interakcije te će morati tome prilagoditi svoje ponašanje (Hintze, 2016).

Samosvjesna umjetna inteligencija (engl. *Self-aware AI*) je posljednja faza razvoja AI, koja trenutno postoji samo hipotetski. To je AI nastala nadogradnjom teorije uma, koja je u tolikoj mjeri evoluirala i postigla toliku sličnost ljudskom mozgu, da je razvila samosvijest. Ova vrsta umjetne inteligencije ne samo da će moći razumjeti i izazvati emocije kod onih s kojima je u interakciji, već će imati i vlastite emocije, potrebe, uvjerenja i potencijalno želje. Iako razvoj samosvijesti AI može potaknuti napredak civilizacije, on također potencijalno može dovesti do katastrofe, jer bi AI svjesna sebe mogla imati ideje, poput samoodržanja, koje bi izravno ili neizravno mogle ugroziti čovječanstvo (Joshi, 2022).

Dodatna podjela umjetne inteligencije, koja se koristi u tehničkom jeziku je podjela AI tehnologija na suženu umjetnu inteligenciju, umjetnu opću inteligenciju i umjetnu superinteligenciju.

Sužena umjetna inteligencija (engl. *Artificial Narrow Intelligence-ANI*) odnosi se na AI sustave koji mogu autonomno izvršiti samo određeni zadatak, koristeći sposobnosti slične ljudskima, ali ne mogu učiniti ništa više od onoga za što su programirani. Imaju vrlo ograničen ili uzak raspon kompetencija. Prema prethodno navedenom sustavu klasifikacije, sužena umjetna inteligencija odgovara svim reaktivnim strojevima i svim sustavima umjetne inteligencije ograničene memorije odnosno uključuje svu postojeću AI, čak i najkompliciraniju koja je stvorena do danas (Joshi, 2022).

Opća umjetna inteligencija (engl. *Artificial General Intelligence-AGI*) sposobnost je AI sustava da uči, percipira, razumije i funkcionira potpuno poput ljudskog bića. Ovi sustavi moći će samostalno razviti složene kompetencije, formirati uzročno-posljedične veze i generalizacije među pojmovima, značajno smanjujući vrijeme potrebno za obuku. Ovo će AI sustave učiniti jednako sposobnima kao što su ljudi, replicirajući složene ljudske kompetencije (Joshi, 2022).

Umjetna superinteligencija (engl. *Artificial Superintelligence-ASI*) vjerojatno će označiti vrhunac istraživanja AI. ASI sustavi će biti iznimno bolji od ljudi u svemu što rade, zahvaljujući većoj memoriji, bržoj obradi i analizi podataka te sposobnosti donošenja odluka. Razvoj AGI i ASI dovest će do scenarija koji se najčešće naziva singularnost (Joshi, 2022). Mogućnost da čovječanstvo ima na raspolaganju tako moćne strojeve može dovesti do utopije, u kojoj je rješena većina današnjih problema, ili pak može ugroziti naš način života, pa i naše postojanje.

Srećom, koliko vremena bude potrebno za dostići ovaj stupanj razvoja AI sustava, toliko vremena ćemo imati za osigurati sigurnost njihova korištenja.

Uz navedene najčešće podjele AI, definirat ćemo i neke novije koncepte, koje im je potrebno pribrojiti. To su umjetna kognicija, kognitivna umjetna inteligencija i umjetna emocionalna inteligencija.

Umjetna kognicija (engl. *Artificial Cognition-AC)* i kognitivna umjetna inteligencija (engl. *Cognitive Artificial Intelligence-CAI*) dvije su specifične pod-domene AI koje su usmjerene na repliciranje ili oponašanje ljudskih kognitivnih sposobnosti. Ne radi se o sinonimima.

Umjetna kognicija se fokusira na razumijevanje i repliciranje ljudskih kognitivnih procesa pomoću AI sustava. Cilj joj je razviti računalne modele i algoritme koji simuliraju ljudsku kogniciju, poput percepcije, zaključivanja, učenja i rješavanja problema (Taylor i Taylor, 2021). S druge strane, kognitivna umjetna inteligencija širi je pojam, koji ne obuhvaća samo replikaciju ljudske kognicije, već i integraciju kognitivnih sposobnosti u AI sustave. Kombinira tehnike AI, kognitivne znanosti, neuroznanosti i srodnih polja, za stvaranje inteligentnih sustava koji mogu percipirati, razumjeti, zaključivati i učiti na način sličan ljudskoj spoznaji. U njenom je fokusu najčešće sposobnost tih sustava da povećavaju svoje znanje (engl. *Knowledge Growing-KG)* (Jagannathan i Parvees, 2022).

Umjetna emocionalna inteligencija (engl. *Artificial Emotional Intelligence*) područje je AI koje se bavi razvojem sustava koji mogu prepoznati i interpretirati ljudske emocije i njihove izraze te na temelju toga prilagoditi svoje ponašanje (Stone et al., 2016).

3. PRIMJENE UMJETNE INTELIGENCIJE U POSLOVANJU

3.1. Uloga umjetne inteligencije u digitalnoj transformaciji poslovanja

Krilatica *„digital or die"* na sažet način izražava u kolikoj mjeri zadržavanje konkurentnosti i opstanak na tržištu današnjice zahtjevaju digitalnu transformaciju poslovanja.

AI igra ključnu ulogu u digitalnoj transformaciji poslovanja jer može doprinijeti automatizaciji procesa, poboljšanju kvalitete proizvoda i usluga, smanjenju troškova, kao i povećanju efikasnosti i produktivnosti. Primjerice, primjena AI u proizvodnji može pomoći u optimizaciji lanca opskrbe, smanjenju troškova proizvodnje i poboljšanju kvalitete proizvoda (Kehayov et al., 2022). Primjena u prodaji može pomoći u personalizaciji marketinških kampanja i boljem razumijevanju potreba kupaca (Haleem et al., 2022), a u zdravstvu može pomoći u prepoznavanju ranih simptoma bolesti, optimizaciji dijagnostičkih postupaka i praćenju zdravlja pacijenata (Lee i Yoon, 2021).

Također, AI može unaprijediti procese analize podataka, učenja iz prethodnih iskustava i donošenja boljih odluka. Na primjer, u financijama će donijeti koristi u predviđanju rizika, poboljšanju procesa kreditiranja i upravljanju financijskim portfeljima (Fares et al., 2022).

Ipak, uz sve prednosti, primjena AI u poslovnom okruženju može predstavljati određene izazove i rizike. Potrebno je uzeti u obzir etičke, pravne i sigurnosne aspekte te osigurati zaštitu podataka i privatnosti korisnika. Potrebno je osigurati adekvatno obrazovanje i edukaciju zaposlenika koji će raditi s AI sustavima te osigurati transparentnost u procesima donošenja odluka koje se temelje na AI algoritmima (Erokhin, 2020).

Usvajanje svih digitalnih tehnologija, a posebno AI, je velik izazov. To usvajanje podrazumijeva brojne, kontinuirane i istodobne prilagodbe organizacijskih resursa, osoblja, kulture i procesa donošenja odluka. Dodatan izazov za organizacije koje usvajaju AI proizlazi iz toga što se AI platforme razlikuju i po opsegu i po složenosti, otežavajući upoznavanje s njima, a time i njihovu implementaciju za postizanje konkurentske prednosti (Holmström, 2022).

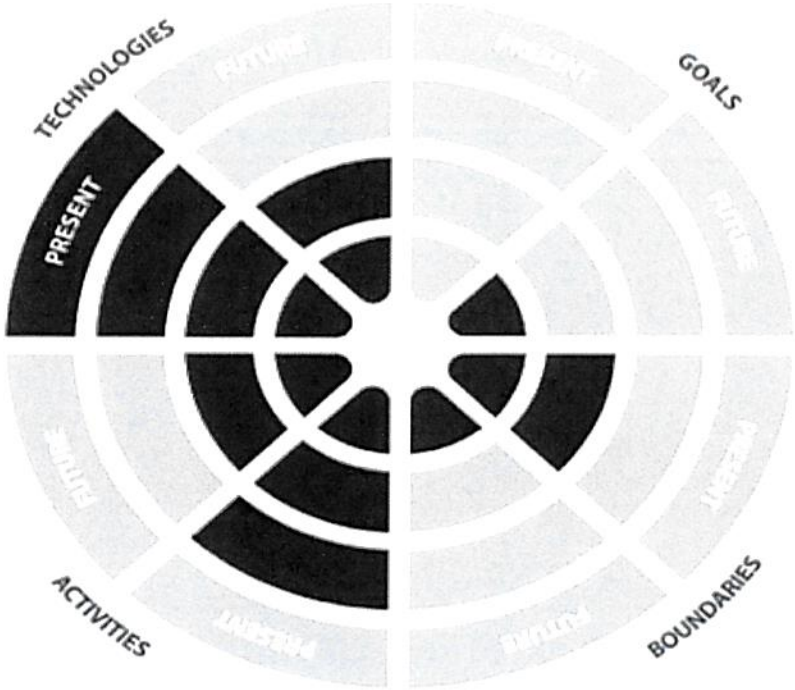
Naime, ako AI sustave definiramo kao racionalne agente koji autonomno odgovaraju na inpute, s malo ili bez intervencije korisnika, izvršavajući zadatke vođeni svojim temeljnim modelima i funkcijama (Bostrom, 2017), tada ti sustavi čine novu vrstu aktera u kontekstu suvremenog organiziranja, pa ih tako trebaju promatrati i menadžeri zaduženi za digitalnu transformaciju. Između ostaloga, menadžeri moraju odlučiti hoće li se usredotočiti na iskorištavanje postojećih tehnologija ili će ulagati u nove AI tehnologije za budućnost. Također se suočavaju s pritiscima brzih promjena kako bi išli ukorak s konkurencijom i zahtjevima ulagača s jedne strane te izbjegavanjem nepromišljenih odluka i uzimanjem u obzir dugoročnih promjena u poslovnom okruženju i društvu s druge strane (Kane et al., 2017).

Vrlo malo poduzeća je prošlo uspješne digitalne transformacije. Kane i suautori su u globalnoj studiji MIT Sloan Management Review 2017. godine otkrili da se samo 25% organizacija transformiralo u digitalna poduzeća, dok ih je 41% na transformativnom putovanju, a 34% je uložilo više vremena pričajući o digitalnoj transformaciji nego radeći na njoj (Kane et al., 2017).

Budući da AI tehnologije imaju kognitivne sposobnosti slične ljudskima, uključujući znanje, učenje, opažanje, osjećanje, djelovanje, komunikaciju i rasuđivanje, njihova primjena može imati dalekosežne posljedice za organizacije i ostale aktere ekosustava, uključujući potrošače, dobavljače, pružatelje usluga na prvoj liniji i druge dionike (Huang et al., 2018). AI platforme će transformirati organizacije na kvalitativno različite načine od drugih tehnologija, stoga je ključno razviti sposobnosti organizacija da se suoče s ovim izazovima odnosno njihovu spremnost za AI (Holmström, 2022).

Jonny Holmström predstavio je okvir koji može pomoći da se odgovori na jedan od prvih izazova: procjenu organizacijske spremnosti za AI odnosno procjenu njezine sposobnosti da implementira AI tehnologije kako bi potpomogla digitalnu transformaciju. Ovaj okvir može olakšati analizu trenutnog sociotehničkog statusa AI u organizaciji i izglede za potpuniju sociotehničku primjenu tehnologije na način da ona stvara dodanu vrijednost (Holmström, 2022).

U fokusu okvira su četiri ključne dimenzije organizacijskog života: tehnologije, aktivnosti, granice i ciljevi (Holmström, 2022). Tehnologije igraju ključnu ulogu u digitalnoj transformaciji, a ostale tri dimenzije odgovaraju trima dimenzijama Aldricheve definicije organizacija kao ciljno orijentiranih, ograničenih entiteta koji sudjeluju u aktivnostima razmjene sa svojim okruženjem preko organizacijskih granica (Aldrich, 2008). Ovaj model analizi spremnosti organizacije za AI pristupa holistički. Nije dovoljno promatrati samo korištenje AI tehnologija u organizaciji, već i koliki broj aktivnosti ključnih za stvaranje vrijednosti one podupiru, u kojoj mjeri utječu na širenje granica organizacije i njezine odnose sa okruženjem te u kojoj mjeri pridonose identitetu organizacije i ostvarenju njezinih strateških ciljeva (Holmström, 2022). Primjer popunjenog *scorecard-a* spremnosti na AI na slici 2. prikazuje sve četiri važne dimenzije predloženog okvira (tehnologije, aktivnosti, granice i ciljevi). Sama primjena tehnologije neće osigurati uspješnu digitalnu transformaciju.



Slika 2.. Primjer popunjenog scorecard-a spremnosti na AI za osiguravajuću kuću
Izvor: Holmström, J. (2022), From AI to digital transformation: The AI readiness framework, Business Horizons, 65(3), 329-339. https://doi.org/10.1016/j.bushor.2021.03.006

3.2. Disruptivni poslovni modeli temeljeni na umjetnoj inteligenciji

U kontekstu digitalne ekonomije često možemo čuti naziv "game-changer". *Game-changers* su u proteklim epohama razvoja društva bili pojedinci progresivnog razmišljanja, koji su svojim izumima ili inovacijama doprinijeli razvoju brojnih područja ljudskog djelovanja (Spremić, 2017).

U današnjem poslovnom okruženju, izrazom *game-changers* nazivamo "cool" kompanije s potpuno novim, dotada nepoznatim, izrazito inovativnim poslovnim modelima. Zahvaljujući svojoj prilagodljivosti, takve kompanije su sposobne vrlo brzo mijenjati ne samo modele poslovanja, već i poslovne strategije, nudeći tržištu originalne proizvode ili usluge koje poboljšavaju kvalitetu života (Spremić, 2017). Kao da je na njih mislio Toffler, kada je još 1970. rekao: "Nepismeni u dvadeset i prvom stoljeću neće biti oni koji ne znaju čitati ili pisati, već oni koji ne mogu učiti, zaboraviti naučeno i ponovno naučiti" (Toffler, 2022).

Ove kompanije su prepoznale da za stvaranje doista disruptivnih inovacija nije dovoljno promijeniti samo način poslovanja i proizvodnje, već predmet promjena mora biti i način razmišljanja (Spremić, 2017). Upravo činjenjem svih tih promjena one mijenjaju pravila igre. Pročišćavanje, prilagodba, revidiranje i redizajniranje poslovnog modela pruža *game-changerima* putokaz za postizanje holističkih ciljeva, iskorištavanjem strateških prednosti AI tehnologija i njihova golemog potencijala za pokretanje novih industrija i disrupciju postojećih (Sewpersadh, 2023).

Ulaganja u pothvate usmjerene na razvoj i korištenje AI porasla su 1800 % u samo šest godina (Sewpersadh, 2023). Poduzeća ulažu očekujući da im AI omogući ulazak u nove poslovne segmente, zadržavanje konkurentske prednosti u trenutnoj industriji ili stvori bolje prilike za monetizaciju. Očekivanja su velika jer je brzina njezina puta od otkrića do komercijalne primjene dosada nezabilježena (Dhanrajani, 2020).

Nedvojbeno je da je uspon AI već doveo do radikalnih promjena u postojećim industrijama i sektorima, pa možemo reći da su algoritmi već sada srž mnogih poslovnih modela i postaju DNK modernih poduzeća (Dhanrajani, 2020). Još radikalnije promjene očekuju se u sljedećih pet do deset godina.

Neki od primjera poduzeća novog doba su: mikrosegmentirane, hiperpersonalizirane platforme za online kupnju, dijeljenje vožnje vođeno GPS-om, kanali za *streaming* vođeni preporukama, *EdTech* tvrtke temeljene na adaptivnom učenju, konverzacijsko planiranje rada vođeno sa AI, itd. (Dhanrajani, 2020).

Novi segmenti i industrije temeljene na AI su u nastajanju. Istovremeno, tradicionalna poduzeća kreću u reinženjering ne samo svojih ulaganja, već i poslovnih modela, procesa i sustava, kako bi novu tehnologiju koristila na načine koji poboljšavaju korisnička iskustva i odnose sa sve svjesnijim i izbirljivijim korisnicima. Mogli bismo reći da AI, budući se usredotočuje na ponašanje kupaca, postojeće organizacije čini humanijima te je više usmjerena na razvoj komercijalnih primjena koje optimiziraju učinkovitost u postojećim industrijama, a manje na razvoj patentiranih algoritama koji bi mogli dovesti do novih industrija (Dhanrajani, 2020).

Stabilna i dugoročna ekonomska dobit od primjene AI mogla bi biti u potrazi za rješenjima složenih i dosada neriješenih problema. Takva rješenja mogu postati temelji novih segmenata postojećih industrija. Prvaci u ekonomiji vođenoj algoritmima bit će poslovni lideri koji usklađuju svoje strategije kako bi povećali AI stručnost i spremnost svojih organizacija, kontinuirano prate najnovije algoritme i redefiniraju poslovne modele da omoguće unovčavanje novih prilika (Dhanrajani, 2020).

Najčešće korišten, ali i dalje najbolji primjer disruptivnog poslovnog modela, jest promjena poslovnog modela taxi industrije.

Tijekom zadnjih desetak godina s korištenja papirnih karti gradova prešli smo na AI, koja nam daje upute za putovanje s jednog mjesta na drugo. Pritom, za razliku od dotadašnjih karti, AI uzima u obzir zatvorene ceste, trenutne uvjete u prometu i vozačeve preferencije, ali se sam čin vožnje ne mijenja suštinski. Vozači taksija koriste digitalne alate i alate temeljene na AI za pronalaženje

učinkovitijih ruta i, zahvaljujući mobilnim tehnologijama, učinkovitiju otpremu. Radimo istu stvar, koristeći isti sustav (postojeću mrežu taksi vozača), ali bolje (Agrawal et al. 2022).

Međutim, Uber i Lyft shvatili su da, uz pouzdane upute i sveprisutne mobilne uređaje, svatko može pružiti uslugu vožnje i stvorili potpuno novi sustav naručivanja vožnji. Prije pet godina u SAD-u je bilo otprilike 200 tisuća profesionalnih vozača taksija i limuzina, a danas je 10 puta veći broj ljudi koji voze samo za Uber (otprilike 3,5 milijuna u SAD-u). Naravno, za integraciju toliko veće radne snage bile su potrebne dodatne inovacije u sigurnosti, praćenju lokacije, cijenama, otpremi i nizu drugih područja (Agrawal et al. 2022).

Slično tome, novi generativni jezični modeli poput ChatGPT-a ili alati koji kreiraju slike iz opisa pisanih prirodnim jezikom poput DELL-E, gotovo svima daju sposobnost da jasno, gramatički ispravno i učinkovito pišu ili kreiraju umjetnost, što će utjecati na to tko je sposoban baviti se kreativnim poslovima i stvoriti nove poslovne modele (Agrawal et al. 2022).

Disruptori tržišta u proizvodnoj industriji i trgovini koncentrirani su na automatizaciju rada korištenjem robotske automatizacije procesa (engl. *Robotic Process Automation-RPA*) i uslužnih robota u globalnim poslovnim uslugama (engl. *Global Business Services-GBS*). Trgovci na malo i proizvođači razvijaju pametne usluge s kojima ulaze u uslužni sektor. Oni transformiraju svoje proizvode ugradnjom softvera za komunikaciju s podatkovnim oblakom, koji se zatim mogu analizirati naprednim analitikama podataka te sukreirati usluge s dodanom vrijednošću. Dva temeljna načela koja vode međunarodne uslužne centre su zadovoljstvo korisnika i kontinuirano poboljšanje kroz inovacije (Sewpersadh, 2023). Primjeri takvih poduzeća su Caterpillar, Michelin, Siemens i Voith Group.

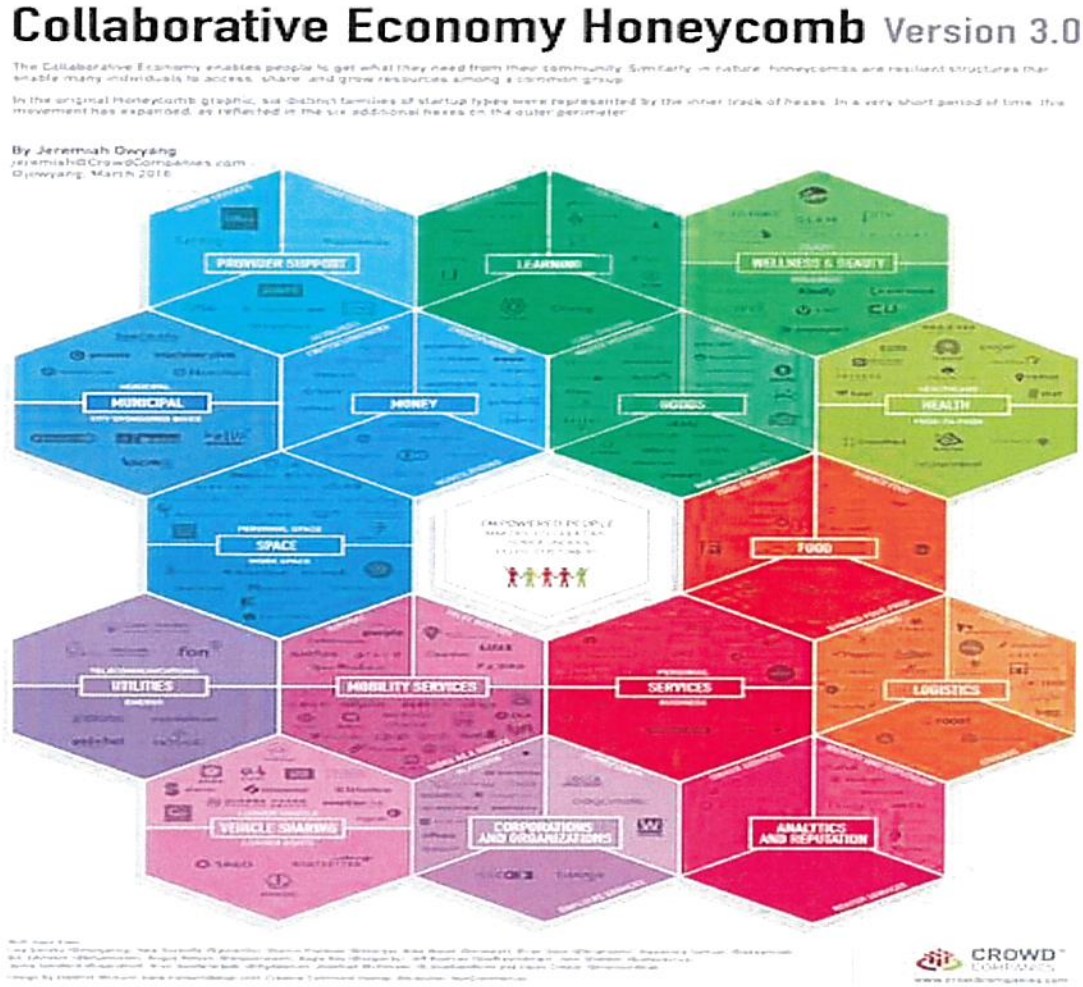
Softbotovi su napredni uslužni roboti s mogućnostima strojnog učenja, koji se koriste za upravljanje i analizu velikih podataka većom brzinom, preciznošću i dosljednošću nego što to ljudi mogu postići. Servisni botovi temeljeni na tehnologijama obrade prirodnog jezika (engl. *Natural Language Processing-NLP*), razumijevanja prirodnog jezika (engl. *Natural Language Understanding-NLU*) i stvaranja prirodnog jezika (engl. *Natural Language Generation-NLG*) razlikuju se od šireg pojma uslužnih robota po svojoj sposobnosti korištenja jezika za razgovor sa klijentima. Poznati su i kao chatbotovi, AI botovi, AI pomoćnici, virtualni pomoćnici/agenti ili digitalni pomoćnici/agenti (Sewpersadh, 2023).

Oni nemaju ljudska ograničenja poput bolesti, zamora, ometanja ili štrajka. Mogu upravljati kvalitetom odnosa s korisnicima rješavanjem svakodnevnih zadataka, kako bi se ljudi mogli posvetiti zadacima kao što su susreti s kupcima i ugovaranje. Ti zadaci donose poboljšane prihode i vrijednost za korisnike (zadovoljstvo korisnika, isporuka usluga i izvedba kontaktnih centara). Integracijom s društvenim medijima, *softbotovi* mogu pristupiti online podacima klijenata i saznati njihove preferencije, osjećaje, stavove i sklonosti te zahvaljujući tome pružiti superiornu uslugu.

Artificial Solutions izvijestio je 2020. godine da je Shell postigao 40-postotno smanjenje broja poziva živim agentima zahvaljujući uslužnim botovima, Emmi i Ethanu. Točno su odgovorili na 97 posto pitanja i riješili 74 posto online upita. U 2019. godini bankarski je sektor ostvario uštedu operativnih troškova od 209 milijuna USD korištenjem servisnih robota, a sektor osiguranja uštede

od 300 milijuna USD u automobilskom, životnom, imovinskom i zdravstvenom osiguranju (Sewpersadh, 2023) .

Dodatno obilježje disruptivnih poslovnih modela jest sukreacija odnosno zajedničko stvaranje vrijednosti. Platforme tvrtkama olakšavaju umrežavanje s unutarnjim i vanjskim okruženjem radi mogućnosti zajedničkog stvaranja i kolaboracije. Putem platformi tvrtke upućuju otvoreni poziv na prijenos znanja, pa korisnici sukreiraju usluge pružanjem uvida tvrtkama. Iskorištavanje društvene povezanosti i odaziva olakšava zajedničko dizajniranje personaliziranih proizvoda, usluga i iskustava. Neki od poslovnih modela koji se temelje na zajedničkom stvaranju vrijednosti su mudrost mnoštva, otvorena inovacija, *crowdsourcing* i *crowdworking* (Sewpersadh, 2023). Amazon Mechanical Turk i Uber primjeri su *crowdworking* filozofije koja koristi digitalne platforme za umrežavanje u sektoru usluga. Tehnologije inovacije usluga koriste pak renomirani brendovi, kao što su Amazon, Netflix, Starbucks i Spotify. Slika 3. prikazuje tzv. "košnicu" kolaborativne ekonomije.



Slika 3. "Košnica" kolaborativne ekonomije
Izvor: Crowd companies, preuzeto s https://www.forbes.com/sites/rawnshah/2015/02/01/how-can-the-collaborative-economy-create-new-local-job-markets/

Navedeni primjeri predstavljaju samo neke od mnogih disruptivnih poslovnih modela temeljenih na umjetnoj inteligenciji, s ključnim elementima koji su transformirali način poslovanja i komunikacije s korisnicima.

Potrebno je također naglasiti da visoko inovativna tvrtka neće ostati disruptor na tržištu ako postane samozadovoljna ili zanemaruje kontinuirano poboljšanje svojih poslovnih procesa. Kako poslovni modeli prelaze iz tradicionalnih u transformativne, naposljetku evoluirajući u one disruptivne, njihovo zastarijevanje je sve brže (Sewpersadh, 2023). Dodatno, nedovoljna znanja, nemogućnost predviđanja posljedica ili pretjerana ulaganja u inovacije mogu dovesti do pretjeranog usvajanja. Iz tog je razloga poželjno uspostaviti strukturu upravljanja strategijom usvajanja tvrtke, koja će procijeniti prikladnost, prihvatljivost, izvedivost i održivost razvoja ili stjecanja inovacija. Upravljanje je ključno za ublažavanje negativnih učinaka pretjeranog usvajanja, složenih ili nekompatibilnih inovacija i digitalnog paradoksa- situacije u kojoj se ne ostvaruje očekivani rast prihoda unatoč dokazanom potencijalu rasta (Sewpersadh, 2023). Jedna od najvažnijih zadaća upravljačkih tijela bit će izraditi smjernice o zaštiti podataka, intelektualnog vlasništva i privatnosti te etičkim pitanjima u upravljanju podacima. Disrupcija ne smije značiti nesigurnost.

3.3. Prednosti i rizici primjene umjetne inteligencije

U ovome potpoglavlju sistematizirat ćemo prednosti primjene AI u poslovanju te im suočiti rizike njene primjene.

Primjena AI dovodi do automatizacije mnogih procesa u poslovanju, smanjenja njihovih troškova te posljedičnog povećanja učinkovitosti. To omogućava tvrtkama da se usredotoče na ključne zadatke i pružanje boljeg korisničkog iskustva. Primjeri uključuju već spomenute chatbotove koji mogu odgovoriti na pitanja korisnika te obraditi njihove zahtjeve, bez ili uz manje ljudskog angažmana te robote u proizvodnji koji mogu ubrzati procese proizvodnje i smanjiti potrebu za radnom snagom.

Zahvaljujući usvajanju AI, moderna poduzeća doživljavaju transformaciju i u pogledu poimanja svrhe i unaprjeđenja prakse mjerenja učinka. Menadžeri se "udružuju" sa strojevima kako bi došli do novih spoznaja o tome što pokreće učinak i kako ga najbolje izmjeriti. Organizacije sve više kombiniraju AI s podacima o izvedbi kako bi generirale i poboljšale ključne pokazatelje učinkovitosti. Čini se da će menadžeri budućnosti koristiti KPI ne samo za praćenje uspjeha poduzeća, već i za redefiniranje i poticanje uspjeha, jer će zahvaljujući AI imati bolje razumijevanje vlastitih kriterija, koje su donedavno rijetko preispitivali (Schrage et al., 2023).

Michael Schrage i koautori na primjeru studije slučaja Google-a, dokazuju da transformacija načina na koji organizacije mjere može fundamentalno promijeniti ono što mjere. 10 puta je vjerojatnije da će tvrtke koje imaju značajne financijske koristi od svojih ulaganja u AI promijeniti način na koji mjere uspjeh u usporedbi s tvrtkama koje ostvaruju manji povrat od svojih ulaganja u AI (Schrage et al., 2023).

Ukratko, organizacije koje koriste AI za generiranje novih metrika ili redefiniranje performansi ostvaruju puno bolje rezultate u odnosu na one koje tu tehnologiju koriste prvenstveno za poboljšanje izvedbe na temelju naslijeđenih metrika ili za puko povećanje performansi. Dok su naslijeđeni KPI-jevi retrospektivni, pametni KPI-jevi gledaju u budućnost; dok su naslijeđeni KPI-jevi usmjereni na fiksne ciljeve, pametni KPI-jevi su prilagodljivi (Schrage et al., 2023). AI algoritmi mogu analizirati odnose između više KPI-jeva te uravnotežiti konkurentske ili komplementarne međuovisnosti.

Dakle, AI ima utjecaj i na to kako organizacije definiraju uspjeh i rast, ne samo na to kako ga mjere.

AI može obraditi i analizirati velike količine podataka brže i preciznije od ljudskih radnika. Zahvaljujući tome može predvidjeti buduće događaje i trendove, dovesti do boljih odluka temeljenih na činjenicama i smanjiti štete koje se događaju zbog ljudske pogreške pri odlučivanju.

Na primjer, naprednim analitikama jednog od najvažnijih KPI-jeva organizacije usmjerenih na korisnika- odljeva korisnika, organizacije mogu identificirati i kontaktirati rizične klijente kako bi ih slanjem automatiziranih standardiziranih ponuda potaknule da ostanu. Danas postoje algoritmi za smanjivanje odljeva kupaca koji precizno određuju koliko truda treba uložiti da se pojedinog kupca zadrži. Na ovaj način kompanije mogu ponude slati najvrjednijim klijentima, a te ponude mogu biti i personalizirane. Zahvaljujući fuziji prediktivne analitike i kreiranja novih strategija za sprječavanje neželjenih poslovnih događaja, donošenje informiranih poslovnih odluka više nije samo posao čovjeka (Schrage et al., 2023).

AI može analizirati velike količine podataka o kupcima i njihovim ponašanjima kako bi stvorio personalizirane ponude proizvoda i usluga, što može poboljšati zadovoljstvo kupaca i povećati vjernost. Primjerice, Netflix korisnicima preporučuje filmove i serije temeljem njihova ponašanja, dok Amazon preporučuje proizvode temeljem njihovih prethodnih kupovina. Nike koristi stečene uvide kako bi stvorio personalizirane tenisice prema zahtjevima kupca, a Tesla za optimizaciju performansi svojih električnih vozila.

CRM temeljen na podacima iskorištava potencijal velikih podataka kako bi se usredotočio ne samo na funkcionalne, već i na dublje psihološke aspekte kupovnog ponašanja. Pametna analitika, kao što je analiza sentimenta, podupire tvrtke u procjeni načina razmišljanja svojih kupaca i učinkovitijoj analizi korisničkog putovanja, sve unutar ograničenja zakona o sigurnosti podataka (Sewpersadh, 2023).

AI se koristi i za stvaranje platformi za dijeljenje odnosno platformi koje povezuju ljude sa sličnim interesima i potrebama. Primjeri uključuju Airbnb koji povezuje goste sa domaćinima koji nude smještaj te Uber koji povezuje vozače i putnike koji trebaju prijevoz (Agrawal et al, 2022).

AI može pomoći u otkrivanju novih prilika za razvoj proizvoda i usluga, poboljšanje postojećih i pružanje novih usluga koje bi mogle ostati nedostupne bez AI.

I zaključno, AI može pomoći u otkrivanju sigurnosnih prijetnji i rizika u poslovanju te pružiti bolju zaštitu od krađe podataka i drugih zlonamjernih aktivnosti. O ovoj primjeni AI bit će više riječi u sljedećem poglavlju.

Uz prednosti, primjena AI u poslovanju nosi i određene rizike i izazove.

AI algoritmi mogu biti vrlo složeni i teški za razumijevanje, što može dovesti do nedostatka transparentnosti i odgovornosti u odlukama koje se temelje na njima (Alawadhi et al, 2022).

AI algoritmi mogu biti diskriminatorni i pristrani prema određenim skupinama ljudi na temelju različitih čimbenika kao što su rasa, spol i etnička pripadnost, posebno ako se algoritmi treniraju na pristranim skupovima podataka. Također, oni mogu biti nepouzdani, ako nisu ispravno trenirani ili ako im nedostaju važni podaci (Wolff, 2020).

Postoje bojazni da će primjena AI dovesti do gubitka radnih mjesta u sektorima u kojima se većina aktivnosti može automatizirati, poput proizvodnje, administracije i financija. Međutim, do tog gubitka radnih mjesta neće doći, bar zasada, zato jer će ljude zamijeniti strojevi, već zato jer će ih zamijeniti drugi ljudi, koji imaju više vještina korištenja AI tehnologijama.

Pri automatizaciji procesa treba učiniti kompromis između mogućih ušteda i utjecaja gubitka radnih mjesta. Općenito, učinci AI tehnologije na model poluge ljudskog kapitala razlikuju se ovisno o skupu vještina ljudi (Sewpersadh, 2023). Primjena AI tehnologija negativno se odražava na niskokvalificirane radnike, ali značajno pozitivno utječe na visokokvalificirane radnike.

Automatizacija procesa pomoću AI dovodi do manje ljudske interakcije, što može smanjiti kvalitetu korisničkog iskustva. Na primjer, ograničenje uslužnih robota jest to što ljudi mogu primijetiti ton glasa, kontekst i podtekst govora na način koji servisni robot ne može savladati. Ovo ograničenje zahtijeva suradnju između uslužnih robota i visokokvalificiranih ljudi, što nas vodi prema mješovitoj radnoj snazi, kao rješenju ovog ograničenja (Sewpersadh, 2023).

Primjeni AI često se pripisuje i ograničenje nedostatka kreativnosti. AI algoritmi se temelje na prethodnim podacima, što može ograničiti kreativnost i inovativnost. Međutim, znamo da inovativnost ne znači nužno sasvim novo rješenje već, pogotovo posljednjih desetljeća, ona znači kombiniranje postojećih rješenja na nov način. U spomenutom gostovanju na podcastu Diary of a CEO, Mo Gawdat ustvrdio je da je i ljudska ingenioznost naravi algoritma- kreativno rješenje je rezultat analize svih poznatih rješenja, eliminiranja onih koja su isprobana te pokušaja rješavanja problema onim rješenjima koja nisu isprobana (Gawdat, 2023). AI već sada to može učiniti.

U nekim sektorima, poput financija i zdravstva, primjena AI može izazvati regulatorne izazove, a tvrtke mogu biti izložene kaznama i drugim sankcijama ako ne poštuju pravila i propise o zaštiti privatnosti i podataka.

I konačno, pored dokazanog pozitivnog utjecaja na inovativnost i stvaranje poslovnih i društvenih vrijednosti, pojačana primjena svih digitalnih tehnologija, pa tako i AI, poslovanje i pojedince izlaže i sasvim novim, neočekivanim, stalno prisutnim i neizbježnim *cyber* rizicima (Spremić, 2017). Upravo intenzivna primjena digitalnih tehnologija, među kojima su i kognitivne tehnologije odnosno AI, jedan je od faktora koji je u posljednjih 15 godina doveo do promjene fokusa sa informacijske na kibernetičku sigurnost (Spremić i Šimunić, 2018).

U tablici 1. sažete su prednosti te rizici i izazovi primjene AI u poslovanju.

Prednosti primjene AI	Rizici i izazovi primjene AI
Smanjenje troškova procesa	Nedostatak transparentnosti
Povećanje učinkovitosti	Gubitak radnih mjesta
Unaprjeđenja prakse mjerenja učinka	Pristranost algoritama
Sposobnost predviđanja trendova	Nepouzdanost algoritama
Unaprjeđenje procesa odlučivanja	Nedostatak odgovornosti u odlukama
Olakšana personalizacija proizvoda i usluga	Nedostatak kreativnosti
Povećano zadovoljstvo i vjernost kupaca	Smanjena kvaliteta korisničkog iskustva
Stvaranje platformi za dijeljenje	Regulatorni izazovi
Stvaranje novih poslovnih prilika	Novi cyber rizici
Unaprjeđenje informacijske sigurnosti	
Smanjenje rizika poslovanja	

Tablica 1. Prednosti i rizici primjene AI

Iz svega navedenoga, razvidno je da se sigurnost podataka korištenjem AI može povećati, kao i ugroziti. U sljedećim poglavljima razmotrit ćemo koje su to karakteristike AI koje je mogu učiniti izvorom prijetnje, ali i podjednako efikasnim i primjerenim obrambenim mehanizmom.

4. UMJETNA INTELIGENCIJA KAO IZVOR PRIJETNJE

4.1. Karakteristike kibernetičkih napada koji koriste umjetnu inteligenciju

Sve kibernetičke napade koji koriste umjetnu inteligenciju možemo obuhvatiti pojmom "ofenzivna umjetna inteligencija" (engl. *Offensive AI*). Radi se o AI primjenama koje omogućuju *cyber* kriminalcima usmjeravanje ciljanih napada neviđene brzine i razmjera koji prolaze nezamijećeni od strane tradicionalnih alata za otkrivanje temeljenih na pravilima (Sparapani i Ruma, 2021).

Analizom znanstvenih članaka zaključeno je da svi ovi napadi dijele određene specifične karakteristike, koje ih čine sofisticiranijima i težima za otkrivanje. Neke od tih karakteristika opisane su u nastavku.

Prva važna karakteristika kibernetičkih napada koji koriste AI jest njihova prilagodljiva i promjenjiva odnosno evoluirajuća priroda. To znači da se napadi pokretani sa AI mogu prilagoditi i samostalno razvijati u stvarnom vremenu, na temelju reakcija ciljanih sustava i promjena njihovih obrambenih mehanizama.

Prilagodljiva i evoluirajuća priroda tih napada proizlazi iz sposobnosti AI algoritama da budu svjesni svog okruženja, iz njega kontinuirano uče te prilagođavaju i poboljšavaju svoje tehnike napada na temelju podataka koje obrađuju ili ishoda koje promatraju (Guembe et al., 2022). Proučimo par primjera prilagodljivosti i evolucije napada.

AI algoritmi mogu učiti iz svojih interakcija s ciljanim sustavima i njihovim obrambenim mjerama. Oni mogu analizirati odgovore i ishode svojih napada te prilagoditi svoje strategije u skladu s time (Sparapani i Ruma, 2021). Na primjer, ako je napad otkriven i blokiran, AI algoritam će analizirati obrambene mehanizme koji su uzrokovali neuspjeh pokušaja napada i razviti nove obrasce napada ili tehnike izbjegavanja, kako bi u budućnosti zaobišao obrambene mehanizme. Proučavanjem obrazaca i ponašanja koji pokreću upozorenja ili obrambene radnje, algoritmi umjetne inteligencije mogu modificirati vektore napada kako bi izbjegli otkrivanje (Sparapani i Ruma, 2021).

AI algoritmi mogu generirati polimorfni zlonamjerni softver koji neprestano mutira i mijenja strukturu koda (Dixon i Eagan, 2019). To omogućuje zlonamjernom softveru da generira nove varijante koje imaju različite potpise i karakteristike, što otežava antivirusnim rješenjima temeljenim na potpisima da ih otkriju i blokiraju (Guembe et al., 2022). AI algoritmi mogu unaprjeđivati zlonamjerni softver tijekom vremena kako bi se suprotstavio najnovijim obrambenim mehanizmima, osiguravajući učinkovitost zlonamjernog softvera u kompromitaciji sustava i njegovu postojanost (Dixon i Eagan, 2019).

AI algoritmi mogu analizirati kontekstualne informacije o ciljanom sustavu i njegovom okruženju i donositi informirane odluke na temelju tih informacija odnosno mogu provoditi kontekstualno odlučivanje (Guembe et al., 2022). Na primjer, napad pokretan sa AI može analizirati ciljanu mrežnu arhitekturu, sigurnosne kontrole, komunikaciju i ponašanja korisnika, kako bi odredio

najučinkovitije vektore i tehnike napada. Ova prilagodljivost omogućuje da se napad prilagodi specifičnoj meti, povećavajući šanse za uspjeh (Dixon i Eagan, 2019).

AI algoritmi korišteni u maliciozne svrhe mogu iskoristiti tehnike kao što je učenje iz uspješnih ishoda (engl. *Reinforcement Learning*) kako bi kontinuirano poboljšavali svoje sposobnosti napada. Mogu se trenirati u simuliranim okruženjima ili kroz napade u stvarnom svijetu, učeći iz uspješnih podviga i prilagođavajući svoje strategije na temelju povratnih informacija. Na primjer, moguće je stvoriti autonomnog agenta koji se ponaša slično kao tester tijekom penetracijskog testiranja-šalje upite ciljanom sustavu, analizira odgovore tog sustava i u njima pronalazi ranjivosti koje je moguće iskoristiti. Na ovaj način napadači mogu kreirati izuzetno opasne *SQL injection* napade. Učenje iz uspješnih ishoda omogućuje napadačima da s vremenom poboljšaju svoje tehnike, čineći napade učinkovitijima i sofisticiranijima (Erdődi et al., 2021).

Prilagodljiva i evoluirajuća priroda napada smanjuje učinkovitost i otežava održavanje tradicionalnih statičkih obrana temeljenih na pravilima te predstavlja značajan izazov za stručnjake kibernetičke sigurnosti (Sparapani i Ruma, 2021). Oni moraju stalno ažurirati svoje obrane i koristiti podjednako napredne tehnike za suzbijanje ovih dinamičnih i inteligentnih napada.

Druga važna karakteristika kibernetičkih napada koji koriste umjetnu inteligenciju jest njihova automatizacija i skalabilnost. AI omogućuje automatizaciju različitih faza napada, uključujući identifikaciju meta (izviđanje), naoružavanje, izvršavanje napada, iskorištavanje ranjivosti, upravljanje i kontrolu itd. (Guembe et al., 2022). To znači da napadi mogu biti brži i učinkovitiji, a napadači mogu ciljati veći broj sustava ili korisnika istovremeno, što može rezultirati kompromitiranjem tih sustava ili krađom osjetljivih informacija.

U nastavku je detaljnije pojašnjenje automatizacije pojedinih faza napada.

AI može automatizirati proces prikupljanja informacija o potencijalnim metama, tzv. izviđanje ili identifikaciju meta. Analizirajući velike količine javno dostupnih podataka, AI algoritmi mogu identificirati ranjivosti, potencijalne ulazne točke i slabosti u infrastrukturi mete (Guembe et al., 2022). Automatizacija ove faze pomaže napadačima da brzo identificiraju vrijedne mete i skrate vrijeme potrebno za izviđanje.

Nakon završetka faze izviđanja, algoritmi umjetne inteligencije mogu se koristiti za automatizirano stvaranje alata za napad i zlonamjernog softvera, tzv. naoružavanje. Kako je prethodno navedeno, AI može generirati sofisticirane i polimorfne varijante zlonamjernog softvera, čineći ih težima za otkrivanje tradicionalnim sigurnosnim sustavima (Erokhin, 2020). Automatizacija ove faze omogućuje napadačima da brzo razviju napade i prilagode ih meti i njezinom okruženju.

AI može automatizirati iskorištavanje ranjivosti u ciljanim sustavima. Iskorištavanjem strojnog učenja, AI algoritmi mogu učiti iz prethodnih napada i samostalno razvijati nove metode za zaobilaženje sigurnosnih kontrola (Guembe et al., 2022). Automatizacija ove faze omogućuje napadačima da identificiraju i iskoriste ranjivosti u velikom broju, značajno povećavajući stopu uspješnosti napada.

AI može automatizirati različite post-eksploatacijske aktivnosti, kao što su lateralno kretanje unutar ciljane mreže, eksfiltracija podataka i ustrajnost (Muppidi et al., 2022). Automatizacija u ovoj fazi

omogućuje napadačima brzo širenje mrežom, izdvajanje osjetljivih informacija i održavanje pristupa tijekom duljeg razdoblja, a da ne budu otkriveni (Guembe et al., 2022).

AI može automatizirati i izbjegavanje sigurnosnih obrana i mehanizama za otkrivanje. Napadači mogu koristiti AI algoritme za razvoj tehnika koje zaobilaze sustave za otkrivanje upada, rješenja za obranu od zloćudnog koda i druge sigurnosne kontrole.

Uz automatizaciju, AI omogućuje napadačima i skaliranje napada na veliki broj sustava istovremeno. I automatizacija i skalabilnost značajno povećavaju brzinu i učinkovitost kibernetičkih napada (Sparapani i Ruma, 2021).

Treća bitna karakteristika kibernetičkih napada koji koriste umjetnu inteligenciju jest poboljšana manipulacija podacima. Pod pojmom manipulacija podacima u ovome kontekstu podrazumijevamo sve radnje upravljanja nad podacima, poput analize, obrade, prepoznavanja uzoraka, generiranja i proširenja podataka.

U nastavku je detaljnije pojašnjenje načina unaprjeđenja pojedinih radnji nad podacima.

AI algoritmi izvrsni su u analizi i razumijevanju velikih količina podataka. Ova mogućnost napadačima omogućuje prikupljanje opsežnih informacija o svojim metama, poput ponašanja korisnika, konfiguracija sustava, obrazaca mrežnog prometa i sl. (Dixon i Eagan, 2019). Sveobuhvatnom analizom ovih podataka, napadači mogu identificirati ranjivosti i slabosti koje se mogu iskoristiti.

AI algoritmi vrlo su vješti u prepoznavanju uzoraka unutar podataka. Iskorištavanjem sposobnosti prepoznavanja, interpretiranja i razumijevanja uzoraka, napadači mogu identificirati specifične podatkovne točke ili nizove kojima se može manipulirati kako bi postigli svoje ciljeve (Guembe et al., 2022). To može uključivati iskorištavanje predvidljivih obrazaca u ponašanju korisnika, mrežnoj komunikaciji ili odgovorima sustava, a moguće je identificirati i korisnike s visokim privilegijama ili pristupom osjetljivim podacima te ih ciljati na sofisticiranije načine, poput ciljanog inženjeringa ili prijevara (Dixon i Eagan, 2019).

AI algoritmi su superiorni u generiranju i proširenju podataka. Oni mogu generirati sintetičke podatke koji su vrlo slični podacima iz stvarnog svijeta (Guembe et al., 2022). Napadači mogu koristiti ovu mogućnost za stvaranje izmijenjenih ili proširenih zlonamjernih podataka ili unosa koji mogu prevariti ili iskoristiti ciljane sustave koji su im meta. Na primjer, AI algoritmi mogu generirati phishing e-poruke realističnog izgleda, sintetičke URL-ove, krivotvorene dokumente, lažne recenzije ili objave na društvenim mrežama te sintetičke *captchas*, kako bi prevarili korisnike ili zaobišli sigurnosne mjere sustava (Guembe et al., 2022).

Napadi vođeni AI mogu manipulirati i modelima i algoritmima strojnog učenja. Iskorištavanjem ranjivosti u procesu obuke ili zaključivanja modela, napadači mogu manipulirati ulazima ili izlazima modela ili ubacivati tzv. suparničke instance, što dovodi do netočnih predviđanja ili odluka (Guembe et al., 2022). To može imati ozbiljne posljedice, posebno u kritičnim područjima poput financija, zdravstva ili autonomnih sustava.

Dakle, napredne mogućnosti AI algoritama u analizi podataka, prepoznavanju uzoraka i generiranju podataka omogućuju napadačima da manipuliraju podacima na sofisticirane i ciljane načine. To povećava njihovu sposobnost da prevare, iskoriste ili izbjegnu otkrivanje unutar ciljanih sustava. Branitelji tih sustava moraju primijeniti robusnu provjeru valjanosti podataka, otkrivanje anomalija i sigurnosne mjere za zaštitu modela kako bi ublažili rizike koje predstavlja pomoću AI omogućena manipulacija podacima u kibernetičkim napadima (Rios, 2021).

Neke metode napada manipulacijom podacima bit će podrobnije objašnjene u sljedećem poglavlju.

Iduća bitna karakteristika kibernetičkih napada koji koriste AI jest nedostatak objašnjivosti i interpretabilnosti odnosno otežano razumijevanje. Neke AI tehnike, kao što su neuronske mreže dubokog učenja, mogu biti vrlo složene, što predstavlja izazov za razumijevanje i tumačenje njihovih procesa donošenja odluka (Erokhin, 2020). Ovaj nedostatak objašnjivosti može otežati prepoznavanje malicioznog koda i rješavanje ranjivosti ili pristranosti u AI sustavima, potencijalno omogućujući napadačima da iskoriste to nerazumijevanje (Guembe et al., 2022). Iako je učinjen značajan napredak na tome da AI postane objašnjivija i da se nerazumljivi modeli ne koriste u područjima visokih uloga, sustavi odgovornosti zahtijevaju više od objašnjenja o tome kako je donesena odluka; oni zahtijevaju normativne prikaze kako i zašto je odluka u skladu s ljudskim vrijednostima (Littman et al., 2021).

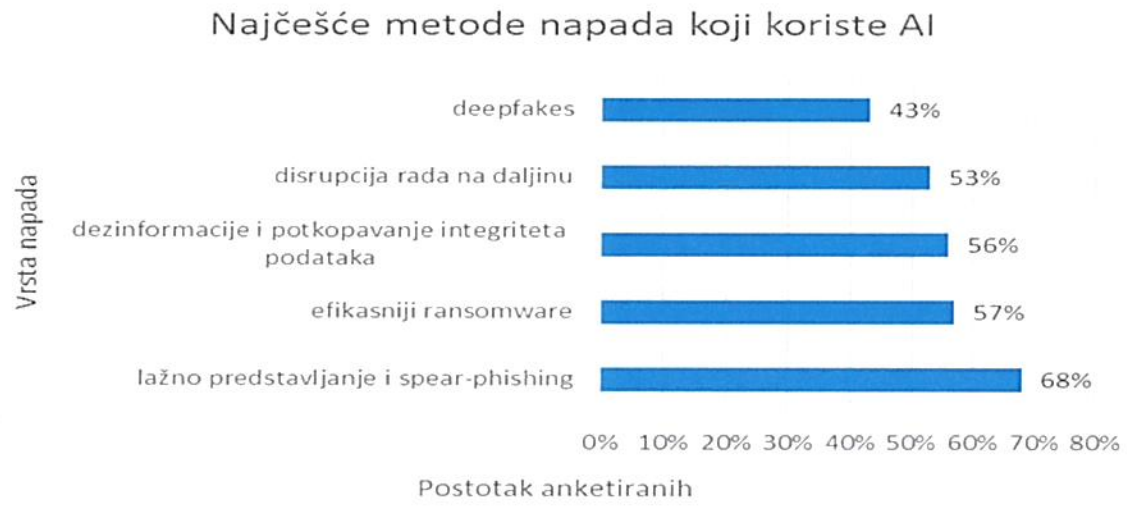
Zlonamjerni akteri mogu zloupotrijebiti i pristupačnost i dostupnost alata i tehnika AI. Oni su široko dostupni za razvoj i implementaciju naprednog zlonamjernog softvera, stvaranje kampanja društvenog inženjeringa ili automatiziranje napada velikih razmjera (Sparapani i Ruma, 2021). Pristupačnost i demokratizacija AI tehnologija mogu osnažiti zlonamjerne aktere znatno sofisticiranijim mogućnostima nego što bi ih imali bez primjene tih tehnologija.

I konačno, među karakteristikama kibernetičkih napada koji koriste AI često se spominje nedostatak etičkog kodeksa. AI nije intrinzično moralno osviještena niti posjeduje etičke principe (Visvizi i Bodziany, 2022). Ako se koristi neetički, može se upotrijebiti za zlonamjerne svrhe, kao što su nadzor, manipulacija ili potkopavanje temeljnih vrijednosti ne samo informacijske sigurnosti, već i temeljnih društvenih vrijednosti (Bird et al., 2020). Ovdje je važno napomenuti da sama po sebi AI nije prijetnja tim temeljnim vrijednostima, već je to potencijalno način na koji se ona koristi.

Analizirane studije dokazuju da je nemogućnost tradicionalnih tehnika kibernetičke sigurnosti da otkriju i ublaže napade potpomognute sa AI izravno povezana s njihovom nesposobnošću da se nose s brzinom, složenošću logike odlučivanja i promjenjivom prirodom (višestrukim varijantama) napada (Guembe et al., 2022). Obrana od kibernetičkih napada omogućenih sa AI zahtjeva višestruki pristup koji kombinira napredne obrambene sustave temeljene na AI sa ljudskom stručnošću i etičnošću. Naime, iste karakteristike koje AI čine potencijalnom prijetnjom mogu se iskoristiti i u obrambene svrhe. AI ima potencijal biti od koristi za kibernetičku sigurnost, kao i potencijal da joj naškodi. AI se koristi i kao mač, za podršku zlonamjernim napadima, i kao štit, za suzbijanje sigurnosnih rizika (Alawadhi et al, 2022). Upravo stoga je postizanje ravnoteže između iskorištavanja sposobnosti AI za obranu i upravljanja rizicima povezanim sa zlonamjernom upotrebom AI stalni izazov za organizacije, sigurnosne timove i istraživače AI (Guembe et al., 2022).

4.2. Metode napada i njihove implikacije

Postoje različite metode napada na informacijsku sigurnost koje koriste AI. U izvješću MIT Technology Review Insights, 309 poslovnih lidera bilo je upitano da navedu na koji način se AI najčešće koristi u maliciozne svrhe usmjerene protiv poslovnih organizacija. Anketirani lideri mogli su odabrati više od jedne vrste/metode napada. Na slici 4. prikazani su rezultati ankete. Najveći broj anketiranih, njih 68%, naveo je AI potpomognuto lažno predstavljanje i *spear-phishing* napade. 57% anketiranih navelo je kao najčešću primjenu AI efikasniji *ransomware*, 56% anketiranih kreiranje dezinformacija i potkopavanje integriteta podataka, 53% anketiranih disrupciju rada na daljinu, a 43% anketiranih kreiranje *deepfakeova* (Sparapani i Ruma, 2021).



Slika 4. Najčešće metode napada koji koriste AI
Izvor: Izrada autora temeljem podataka preuzetih iz: Sparapani, J. i Ruma, L. (2021), Preparing for AI-enabled cyberattacks

U nastavku su navedene i objašnjene neke od najčešćih metoda napada potpomognutih sa AI.

AI poboljšava učinkovitost i sofisticiranost phishing kampanja i kampanja društvenog inženjeringa (Sparapani i Ruma, 2021). To postiže stvaranjem vrlo uvjerljivih i personaliziranih poruka. AI algoritmi mogu analizirati velike količine podataka, posebno o korisnicima s visokim ovlastima, kako bi izradili ciljane phishing e-poruke koje blisko oponašaju komunikacijske obrasce i stilove pisanja sudionika komunikacije. AI se može upotrijebiti i za poboljšanje spear-phishing napada automatiziranjem procesa izrade i distribucije ciljanih phishing e-poruka (Sparapani i Ruma, 2021).

Također, AI se može koristiti za prikupljanje i analizu golemih količina osobnih podataka iz različitih izvora, kao što su platforme društvenih medija, za izradu uvjerljivih napada društvenim inženjeringom. To može uključivati lažno predstavljanje pojedinaca ili korištenje psihološkog profiliranja za manipuliranje žrtvama (Dixon i Eagan, 2019). Na primjer, cyberkriminalci mogu kombinirati društveni inženjering i glasovni phishing pomoću AI softvera koji uči oponašati i

govoriti koristeći snimku ljudskog glasa. Nakon 20 minuta slušanja glasa mete napada ovaj softver je u stanju izgovoriti bilo koju pisanu poruku uvježbanim glasom (Erokhin, 2020). To povećava vjerojatnost uspješnih napada društvenim inženjeringom odnosno vjerojatnost da korisnik nasjedne na lažnu komunikaciju ili otkrije povjerljive podatke.

AI može generirati lažne informacije, slike ili videozapise koji izgledaju autentično. To se može koristiti za manipulaciju korisnicima, širenje dezinformacija ili stvaranje *deepfakeova,* kako bi se manipuliralo javnim mnijenjem i diskreditirale osobe ili organizacije. Ovako stvorene raširene kampanje dezinformiranja mogu imati značajne društvene i političke implikacije (Agarwal i Farid, 2019).

AI može generirati sofisticirani zlonamjerni kod koji je složeniji i zahtjevniji za otkrivanje u odnosu na tradicionalni zlonamjerni kod. Koristi algoritme strojnog učenja za stvaranje zlonamjernog softvera koji se može prilagoditi i učiti iz svog okruženja te uspješno izbjegavati detekciju sigurnosnih mehanizama (Guembe et al., 2022). Ovakav zlonamjerni softver tradicionalnim obranama kibernetičke sigurnosti predstavlja veći izazov za otkrivanje i ublažavanje napada (Dixon i Eagan, 2019).

AI koristi strojno učenje za identifikaciju slabosti i ranjivosti u sigurnosnim sustavima i prilagođavanje napada na temelju tih informacija. To može uključivati otkrivanje ranjivosti u mrežnoj infrastrukturi, probijanje lozinki, zaobilaženje višefaktorskog autentifikacijskog sustava ili izvođenje naprednih tehnika za zaobilaženje drugih sigurnosnih provjera, pa i fizičkih kontrola pristupa (Guembe et al., 2022). Dodatno, AI može optimizirati vektore napada određivanjem najučinkovitijih metoda za iskorištavanje određene ranjivosti, kao i automatizirati samo iskorištavanje ranjivosti (engl. *Automated Exploit Generation*) (Guembe et al., 2022).

Primjer ovakvih napada su AI potpomognuti *brute-force* napadi, napadi kojima je cilj pogađanje i probijanje lozinki ili enkripcijskih ključeva. Napadači mogu koristiti AI za optimizaciju i ubrzanje inače dogotrajnih i iscrpljujućih *brute-force* napada (Sparapani i Ruma, 2021), posebno protiv slabih ili ponovno korištenih lozinki. AI algoritmi mogu učiti iz prethodnih obrazaca napada, identificirati potencijalne ranjivosti i generirati učinkovitije kombinacije lozinki.

AI može ubrzati otkrivanje i iskorištavanje prethodno nepoznatih ranjivosti, također poznatih kao ranjivosti nultog dana (engl. *Zero-day Vulnerability*). Algoritmi strojnog učenja analizirat će softver i ponašanje sustava kako bi identificirali potencijalne ranjivosti koje su ljudski analitičari možda propustili (Abeshu i Chilamkurti, 2018). To omogućuje napadačima da iskoriste te ranjivosti prije nego što budu zakrpane.

AI modeli su ranjivi na suparničke napade, kojima se rade suptilne izmjene ulaznih podataka (perturbacije) kako bi se prevario model strojnog učenja. Suparnička AI (engl. *Adversarial AI)* uzrokuje da model strojnog učenja pogrešno tumači ulazne podatke i ponaša se na način koji pogoduje napadaču- daje netočna predviđanja ili poduzme neželjene radnje (Erokhin, 2020). Na primjer, to se događa kada su neuronske mreže dubokog učenja prevarene netočnom identifikacijom ili netočnom klasifikacijom objekata (Erokhin, 2020) ili kada napadač upotrijebi suparničke primjere koje je generirala AI kako bi zaobišao sustave za prepoznavanje slike ili manipulirao algoritmima donošenja odluka temeljenima na AI (Eykholt et al., 2018). Vrlo često

ova vrsta napada kreira se upotrebom generativnih suparničkih mreža (engl. *Generative Adversarial Networks-GAN*). GAN-ovi se sastoje od dvije neuronske mreže, generatora i diskriminatora, koji se treniraju na način da se "natječu"- generator kako bi proizveo realistične sintetičke podatke koji mogu prevariti diskriminatora da ih klasificira kao stvarne, a diskriminator kako bi ispravno klasificirao stvarne i sintetičke uzorke odnosno primjere. Kada napadači žele zaobići sustave za detekciju zlođudnog koda temeljene na strojnom učenju, u GAN arhitekturu dodaju komponentu zamjenskog detektora. Podaci o obuci zamjenskog detektora sastoje se od primjera zlonamjernog softvera iz generatora i benignih programa iz dodatnog benignog skupa podataka koje su napadači prikupili. Cilj zamjenskog detektora je odgovarati detektoru crne kutije. Detektor crne kutije će prvi otkriti podatke o obuci i kao output označiti program kao bezopasan ili zlonamjeran. Te oznake detektora crne kutije koristi zamjenski detektor kako bi generirao informacije za treniranje još efikasnijeg generatora (Hu i Tan, 2022). Ovakvu primjenu GAN-ova autori su nazvali MalGAN, što je kratica za engl. *Malware Generative Adversarial Network*. Perturbacije je moguće učiniti i na fizičkim objektima (Eykholt et al., 2018).

Napadi potpomognuti sa AI mogu manipulirati i podacima o obuci kako bi ugrozili integritet i učinkovitost modela strojnog učenja. Ubacivanjem zlonamjernih ili pogrešnih podataka u proces obuke, napadači mogu manipulirati ponašanjem modela, što dovodi do netočnih rezultata ili pristranih odluka. Ova metoda napada naziva se trovanje podataka (engl. *Data Poisoning*), a na nju su posebno osjetljivi modeli koji uče nenadzirano. Za trovanje podataka potrebno je znatno manje zlonamjernih uzoraka nego za klasičan suparnički napad, a njegovo otkrivanje je gotovo nemoguće bez naprednih ekspertiza iz područja AI tehnologija (Raza, 2021).

Naoružana AI (engl. *Weaponized AI*) može se široko definirati kao: "zlonamjerni algoritmi umjetne inteligencije koji mogu pogoršati performanse i poremetiti normalne funkcije benignih algoritama umjetne inteligencije, istovremeno pružajući tehnološki napredne scenarije napada u kiberprostoru i fizičkim prostorima" (Yamin et al., 2021). Ovaj pojam se odnosi na upotrebu AI tehnologija u razvoju, uvođenju i izvršavanju ofenzivnih radnji. Uključuje integraciju AI algoritama i sustava u oružje, vojne platforme i alate za kibernetičko ratovanje, kako bi se povećala njihova učinkovitost. Zlonamjerni akteri mogu trenirati AI modele za provođenje ofenzivnih radnji, kao što su automatizirano izviđanje i prikupljanje informacija, generiranje zavaravajućih sintetičkih podataka, sabotaža važnih istraživanja, napadi na kritičnu infrastrukturu ili čak autonomno donošenje odluka tijekom procesa napada (Yamin et al., 2021). Ovo dodaje novu dimenziju složenosti cyberprostora, jer, kako je više puta spomenuto, AI sustavi mogu naučiti i optimizirati svoje strategije napada tijekom vremena.

Opisane metode napada sistematizirane su u tablici 2.

Metode napada koje koriste AI
Phishing kampanje
Kampanje društvenog inženjeringa
Kampanje dezinformiranja i *deepfakes*
Sofisticirani zlonamjerni softver
Automatizirano iskorištavanje ranjivosti
Iskorištavanje ranjivosti nultog dana

Suparnički napadi
Trovanje podataka
Naoružana umjetna inteligencija

Tablica 2. Metode napada koje koriste AI

One mogu imati različite implikacije i posljedice na informacijsku sigurnost. Nekoliko najčešćih opisano je u nastavku.

Napadi poput phishinga, generiranja zlonamjernog koda i brute-force napada mogu rezultirati krađom povjerljivih podataka. To može uključivati financijske podatke, korisničke podatke, poslovne tajne ili druge osjetljive informacije. Krađa takvih podataka može dovesti do financijskih gubitaka, ali i reputacijske štete (gubitka povjerenja korisnika) i mogućih pravnih posljedica (Sparapani i Ruma, 2021).

Kada korisnici postanu žrtve napada i krađe podataka, to rezultira gubitkom povjerenja u organizacije ili platforme koje koriste (Dixon i Eagan, 2019), što može imati dugoročne posljedice na njihovu privrženost i lojalnost. Kada organizacije ne uspiju zaštititi povjerljive podatke ili postanu izvor lažnih informacija, njihova reputacija može biti ozbiljno narušena, što će utjecati na njihovu konkurentnost i dugoročni uspjeh.

Uspiju li se napadima poput phishinga napadači domoći financijskih podataka (podataka o kreditnim karticama ili bankovnim računima), to može rezultirati financijskim prijevarama korisnika i značajnim financijskim gubicima za pojedince i organizacije.

Generiranje lažnih informacija, *deepfakeova* i ciljanih napada na korisnike mogu se koristiti za manipulaciju korisnicima i javnim mnijenjem. To može dovesti do širenja dezinformacija, diskreditiranja pojedinaca ili organizacija, političkog utjecanja, pa čak i društvenih nemira (Agarwal i Farid, 2019).

Napadi koji koriste AI mogu za cilj imati i ugrozu nacionalne sigurnosti. To može uključivati krađu državnih tajni, napade na kritičnu infrastrukturu, širenje dezinformacija s ciljem destabilizacije država ili izazivanje društvenih nemira (Dixon i Eagan, 2019).

4.3. Studije slučaja i primjeri

U ovome potpoglavlju bit će izneseni primjeri metoda napada koje koriste umjetnu inteligenciju, prikupljeni analizom dostupnih studija slučaja.

Za razumijevanje prva dva primjera potrebno je definirati pojam *stegomalware*. *Stegomalware* je napredni zloćudni kod koji koristi steganografiju kako bi izbjegao otkrivanje. Zloćudni kod je skriven u benignim nosačima kao što su slike, dokumenti i video zapisi. U novije vrijeme, modeli neuronskih mreža koriste se kao prijenosnici zlonamjernog softvera za napade (Wang et al., 2022).

IBM-ov istraživački tim predstavio je koncept prikrivenog ciljanog napada nazvan DeepLocker. DeepLocker koristi osobne i okolišne atribute svoje mete (npr. lice, glas, geolokaciju i kretanje) za

treniranje modela neuronskih mreža. Na primjer, kada je lice mete napada ulaz, model generira stabilan kriptografski ključ kao izlaz. Ključ šifrira zlonamjerne pakete (engl. *Payloads*). DeepLocker distribuira video povezan sa zlonamjernim softverom u žrtvino okruženje kroz opskrbni lanac. Kada je video pokrenut, zlonamjerni softver snima slike lica mete i unosi ih u model neuronske mreže. Kada meta koristi softver, model ju prepoznaje i ispisuje kripto ključ; u suprotnom ispisuje besmisleni niz brojeva. Nakon toga se zlonamjerni sadržaj može dešifrirati i izvršiti (Wang et al., 2022).

Sigurnosni analitičari ne mogu dobiti ispravan ključ iz modela kada je meta koja dekriptira sadržaj nepoznata, pa nisu svjesni namjere zlonamjernog sadržaja niti mogu identificirati metu napada (Wang et al., 2022).

U implementaciji, napadači prvo moraju uvježbati model za prepoznavanje lica ili nabaviti unaprijed obučeni model i koristiti ga kao prvi model (M1). Kada je model obučen, on će temeljem inputa lica iste osobe generirati slične, ali ne identične vektore značajki. Vektori se koriste za izgradnju drugog modela- M2. Za njegovu obuku, ulaz su vektori značajki mete iz M1, a izlaz je unaprijed definirani kriptografski ključ. Obuka M2 zahtijeva otprilike deset slika lica mete. Kada je i drugi model obučen, napadači spajaju M1 i M2 u konačni model M. Kada je ulaz lice mete, konačni model šalje kripto ključ. Na taj način, uvjet za okidač i informacije o meti skrivene su u modelu neuronske mreže, a namjera zlonamjernog softvera je skrivena u šifriranom *payloadu*. Skrivanje napada kod DeepLockera je ograničeno činjenicom da je šifrirani *payload* moguće prepoznati kao anomaliju u okruženju mete (Wang et al., 2022).

Stegomalware pod nazivom StegoNet predložen je za ugradnju zlonamjernog softvera u modele neuronskih mreža, na način da su parametri modela zamijenjeni ili mapirani bajtovima zlonamjernog softvera. Istovremeno, performanse modela su održane zahvaljujući složenosti modela i njegovoj toleranciji na pogreške. Predloženi scenarij napada StegoNet-om je pokretanje onečišćenja opskrbnog lanca kroz tržišta strojnog učenja. S razvojem strojnog učenje kao usluge (engl. *Machine Learning As A Service-MLaaS*), stvoreno je tržište na kojem napadači mogu distribuirati zagađene modele putem MLaaS pružatelja usluga (Liu et al., 2020).

Općenito, prednosti skrivanja zlonamjernog softvera u modelima neuronskih mreža su: izbjegavanje otkrivanja (nakon skrivanja, zlonamjerni softver se ne može rastaviti i njegove karakteristike se ne mogu izdvojiti), prikrivanje (zbog redundantnih neurona i izvrsne generalizacije, modificirani modeli mogu zadržati svoje izvorne performanse), visok kapacitet (moguće je ugraditi veliki zlonamjerni softver), neovisnost o ranjivostima sustava (isporuka putem kanala za ažuriranje modela iz opskrbnog lanca ili alternativnim načinima koji ne privlače pažnju korisnika) te univerzalnost (s porastom popularnosti neuronskih mreža) (Wang et al., 2022).

StegoNet nije uspio ponuditi sve ove prednosti, pa nije u većoj mjeri privukao pažnju zajednice IT sigurnosti. StegoNet ima nisku vrijednost stope ugrađivanja-bez degradacije točnosti modela ona je tek oko 15%, što znači da se u male ili srednje velike modele ne može ugraditi veliki zlćudni kod. Također, značajno smanjuje performanse manjih modela, a zahtjeva i značajan napor (obuku ili permutaciju indeksa) da se ugradi ili ekstrahira zlonamjerni softver (Wang et al., 2022).

Da je performanse *stegomalwarea* koji se distribuira kroz modele neuronskih mreža moguće unaprijediti, dokazali su Wang i koautori u svojoj studiji EvilModel 2.0 (Wang et al., 2022). Uspjeli su ugraditi zlonamjerni softver u modele neuronskih mreža s visokom vrijednošću stope ugrađivanja i niskim utjecajem na performanse. Autori predlažu tri metode ugrađivanja: MSB (engl. *Most Significant Byte*) rezervaciju, brzu zamjenu i polovičnu zamjenu. Njihovim korištenjem ugradili su 19 uzoraka zlonamjernog softvera u 10 modela neuronskih mreža i analizirali njihovu izvedbu kako bi ocijenili predložene metode (promatrajući stopu ugrađivanja, učinak na izvedbu i potreban napor). Apstrahiranjem svakog cilja napada u zasebni vektor obilježja, unaprijedili su mogućnosti skrivanja zloćudnog koda u odnosu na DeepLocker. Ponudili su i mogući scenarij napada u 5 koraka koje provodi napadač. Prvi korak je priprema dobro uvježbanog modela neuronske mreže i zlonamjernog softvera za specifične zadatke. Napadači mogu dizajnirati i trenirati vlastite modele ili preuzeti dobro obučene modele iz javnih repozitorija. Jednako tako, mogu ili razviti vlastiti zlonamjerni softver ili ga kupiti, no u svakom slučaju moraju dobro procijeniti strukturu i veličinu modela kako bi donijeli odluku koliko zlonamjernog softvera može biti ugrađeno. Nakon toga ugrađuju softver u model, koristeći spomenute metode. Kada završe s ugradnjom, napadači bi trebali procijeniti izvedbu modela, kako bi se osiguralo da nema velikog pada performansi. Ako on postoji, treba ponoviti ugradnju ili promijeniti model ili zloćudni kod. Ako je izvedba zadovoljavajuća, dizajniraju okidač u skladu sa specifičnim zadacima. Pretvaraju izlaz srednjeg sloja modela u vektore značajki, kako bi pronašli mete i aktivirali ciljani napad. Propagacija EvilModela može se vršiti učitavanjem u javna spremišta, poslužitelje za pohranu podataka u oblaku, tržišta neuronskih mreža itd., kroz onečišćenje opskrbnog lanca ili slične pristupe koji će osigurati da modeli budu isporučeni zajedno s benignim aplikacijama. Zadnji korak je aktivacija zloćudnog koda. On se izdvaja iz modela neuronske mreže i izvršava kada ispuni unaprijed definirani uvjet na krajnjim uređajima (Wang et al., 2022).

Moguće protumjere za ublažavanje prijetnji od strane *stegomalwarea* su podešavanje veličine parametara (smanjenje preciznosti brojeva); modificiranje modela neuronske mreže finim podešavanjem, skraćivanjem ili kompresijom modela (čime se prekida struktura zlonamjernog softvera i sprječavanje njegovog oporavka); zaštita opskrbnog lanca poboljšanom provjerom identiteta korisnika (predstavljanjem provjerenih dobavljača certifikatima koji se izdaju uz modele) i omogućavanjem učitavanja modela samo provjerenim korisnicima te detekcija zlonamjernog softvera u modelu neuronske mreže (kroz promatranje stope preklapanja parametara izvučenih iz različitih slojeva) (Wang et al., 2022).

Druga analizirana studija slučaja odnosi se na *deepfake* tehnologiju. Dobro poznati primjer *deepfakea* jest viralni video s bivšim predsjednikom SAD-a Barackom Obamom. Video, koji su kreirali BuzzFeed i komičar Jordan Peele, demonstrira potencijal tehnologije *deepfake* za generiranje videozapisa realističnog izgleda, koji mogu oponašati obrasce govora, izraze lica, kretnje i manire pojedinaca, u svrhu manipuliranja javnim mnijenjem. U videu je imitacija Baracka Obame od strane Jordana Peelea kombinirana s *deepfake* tehnologijom kako bi izgledalo kao da Obama osobno izgovara riječi koje je izgovorio komičar, uključujući psovke. Ovaj video je primjer lakoće s kojom se mogu stvoriti uvjerljivi, ali izmišljeni videozapisi te ističe važnost svijesti o mogućnostima zlouporabe *deepfake* tehnologije, kao i važnost odgovornosti medija i kritičkog razmišljanja građana u vremenu u kojem se manipulirani sadržaj može širiti neslućenom brzinom.

Za stvaranje *deepfakeova,* koji su kovanica engleskih riječi "deep learning" i "fake" (umjetan, izmišljen), koristi se već spomenuta AI tehnika GAN ili generativne suparničke mreže. GAN mogu koristiti podatke (kao što su slike ljudskih lica) za proizvodnju novih stvari (kao što su impresivno realistične slike ljudskih lica ili videozapisi) (Agarwal i Farid, 2019). Postoji zabrinutost zbog mogućnosti zlouporabe ove tehnologije za širenje lažnih medijskih sadržaja u svrhu dezinformiranja, klevetanja pojedinaca, manipuliranja javnim mišljenjem i političkog utjecanja.

Zbog toga je, pred američke izbore 2020. godine, Hany Farid, profesor i stručnjak za forenziku slika na koledžu Dartmouth, počeo izrađivati softver koji može uočiti političke krivotvorine ili autentificirati autentične videozapise koje se proglašava lažnima. Farid i njegov student, Shruti Agarwal, razvijaju način razlikovanja stvarne osobe od njezine lažne verzije koji nazivaju "mekom biometrijom" (Agarwal i Farid, 2019). Korištenjem automatiziranih alata za proučavanje sati i sati autentičnih videozapisa ljudi poput bivših i sadašnjih američkih predsjednika, otkrivaju veze između pokreta glave, govornih obrazaca i izraza lica. Te se korelacije koriste za izradu modela pojedinca, tako da se, kada izađe novi video, model može upotrijebiti za određivanje ima li pojedinac prikazan u videu govorne obrasce, pokrete glave i izraze lica koji odgovaraju stvarnoj osobi pojedinca (Agarwal i Farid, 2019).

Slika 5. Usporedba originalnog videa sa 3 deepfake videa mekom biometrijom
Izvor: Agarwal, S. i Farid, H. (2019), Protecting World Leaders Against Deep Fakes

U svojim intervjuima sam Farid je iznio da se ovaj i slični alati teško prilagođavaju brzini kojom *deepfakeovi* postaju sve sofisticiraniji- npr. ako istraživači pronađu način kako detektirati nerealističan obrazac treptanja, već za mjesec dana napadači će stvoriti videozapise koji imaju sasvim realističan obrazac treptanja.

Posebno zabrinjava činjenica da je na ovakav način moguće zaštiti demokracije i njihove najistaknutije političare ili pak druge poznate javne osobe, čije su stotine govora dostupne za analizu i treniranje modela, ali nije moguće zaštititi građane. I građani mogu postati žrtve zlonamjernih kampanja diskreditacije i klevetanja, pa je moguće generirati lažne pornografske sadržaje, manipulirati snimkama nadzornih kamera i sl., sa ciljem da se pojedinca lažno optuži ili mu se učini reputacijska šteta. U siječnju 2020. FBI je upozorio da je *deepfake* tehnologija dosegla točku u kojoj je moguće stvoriti umjetne osobe koje bi mogle proći biometrijske testove (Sparapani i Ruma, 2021).

Primjer značajne studije slučaja koja se odnosi na korištenje suparničkih napada za manipuliranje AI sustavima jest istraživanje koje su proveli istraživači sa Sveučilišta Washington i Sveučilišta Michigan (Eykholt et al., 2018). Pokazali su kako se suparnički napadi mogu koristiti za manipuliranje sustavima detekcije znakova zaustavljanja koji se koriste u autonomnim vozilima. U svojoj studiji, istraživači su stvorili fizičke zavaravajuće znakove zaustavljanja, apliciranjem pažljivo dizajniranih naljepnica ili modifikacija na znakove zaustavljanja u stvarnom svijetu. Ove izmjene su imale za cilj zbuniti algoritme za otkrivanje objekata koji se koriste u sustavima autonomnih vozila. Unatoč tome što ljudsko oko percipira znakove za zaustavljanje kao nepromijenjene, algoritmi umjetne inteligencije pogrešno su ih klasificirali, uzrokujući da autonomna vozila pogrešno protumače znakove ili ih uopće ne prepoznaju. Ovo istraživanje istaknulo je ranjivost sustava za otkrivanje objekata temeljenih na AI na suparničke napade u scenarijima stvarnog života. Implikacije mogu biti značajne jer bi neovlašteni znakovi za zaustavljanje mogli dovesti do opasnih situacija te ugroziti sigurnost u prometu i brojne živote (Eykholt et al., 2018).

Ova studija slučaja pokazala je važnost robusnosti i sigurnosti AI sustava, posebno u domenama kritičnim za sigurnost, kao što su autonomna vozila ili kritična infrastruktura. Naglašena je potreba osiguravanja da su AI modeli obučeni i testirani protiv širokog spektra potencijalnih suparničkih napada, kako bi se poboljšala njihova otpornost i pouzdanost u stvarnim uvjetima (Eykholt et al., 2018).

5. UMJETNA INTELIGENCIJA KAO ODGOVOR NA PRIJETNJU

5.1. Karakteristike obrambenih mehanizama temeljenih na umjetnoj inteligenciji

Sve obrambene mehanizme koji koriste umjetnu inteligenciju možemo obuhvatiti pojmom "defenzivna umjetna inteligencija" (engl. *Defensive AI*). Za razliku od tradicionalnih statičkih obrambenih mehanizama temeljenih na pravilima, koji se oslanjaju na povijesne podatke o napadima, obrambena AI uči što je normalno za organizaciju i može otkriti nenormalnu, potencijalno zlonamjernu aktivnost čim se pojavi, čak i ako nikad prije nije viđena (Sparapani i Ruma, 2021).

Analizom znanstvenih članaka zaključeno je da svi obrambeni mehanizmi koji su temeljeni na umjetnoj inteligenciji dijele određene specifične karakteristike, koje ih čine sofisticiranijima i učinkovitijima od tradicionalnih obrambenih mehanizama. Neke od tih karakteristika opisane su u nastavku.

Prva važna katakteristika obrambenih mehanizama temeljenih na AI jest poboljšana manipulacija podacima (Muppidi et al., 2022). Pod pojmom manipulacija podacima u ovome kontekstu podrazumijevamo sve radnje upravljanja nad podacima, poput analize, obrade, prepoznavanja uzoraka, kontekstualizacije podataka, generiranja izvješća i slično.

U nastavku je detaljnije pojašnjenje načina unaprjeđenja pojedinih radnji nad podacima.

AI je sposobna brzo obraditi velike količine strukturiranih i nestrukturiranih podataka kako bi prepoznala i odgovorila na prijetnje u stvarnom vremenu (Muppidi et al., 2022). To uključuje analizu mrežnog prometa, logova sustava, ponašanja korisnika, sigurnosnih događaja i feedova, istraživačkih radova i izvješća o incidentima te drugih relevantnih izvora podataka.

AI može unaprijediti sigurnost redovitim izvođenjem otkrivanja i klasifikacije osjetljivih podataka (*on premise*, na krajnjim točkama, u tranzitu i u oblaku). AI tehnologije omogućuju organizacijama korištenje izvornih podataka i metapodataka za stvaranje sigurnosnog konteksta za svaku njihovu interakciju te za razumijevanje gdje su pohranjeni najosjetljiviji podaci, tko ima pristup (i kako), tko im pristupa (i kada) i koje radnje obavlja s njima. To može pomoći kao podrška nadzoru i kontroli pristupa vrlo osjetljivim podacima, ali i u ispunjavanju standarda privatnosti podataka i usklađenosti s propisima (Muppidi et al., 2022).

AI je izvrsna u prepoznavanju obrazaca i otkrivanju odstupanja od normalnog ponašanja. Analitika ponašanja korisnika i entiteta (engl. *User and Entity Behaviour Analysis-UEBA*) može naučiti tipične obrasce korisničkih aktivnosti, mrežnog prometa ili ponašanja sustava, što joj omogućuje prepoznavanje anomalija ili indikatora ugroženosti (RSA, 2015). Izgradnjom profila ponašanja i primjenom algoritama strojnog učenja, ovi sustavi mogu otkriti prijetnje iznutra i identificirati ugrožene račune. AI algoritmi tako mogu znatno ranije otkriti sumnjive aktivnosti i anomalije koje mogu ukazivati na kibernetički napad, poput pokušaja neovlaštenog pristupa, neuobičajenog korisničkog ponašanja, eskalacije privilegija (Abeshu i Chilamkurti, 2018), eksfiltracije podataka (Dixon i Eagan, 2019) ili nenormalnog prometa mreže.

Poboljšane analitičke sposobnosti i napredna inteligencija prijetnji mogu unaprijediti upravljanje informacijskom sigurnošću i na druge načine. To uključuje generiranje izvješća o sigurnosnim događajima, statističku analizu i vizualizaciju podataka te pružanje preporuka za jačanje sigurnosti. Platforme za obavještavanje o prijetnjama koje pokreće AI mogu pomoći sigurnosnim timovima da, zahvaljujući kontekstualiziranim i obogaćenim podacima, skrate vrijeme potrebno za istrage, ublaže rizike, budu bolje informirani o najnovijim prijetnjama te proaktivno poboljšaju svoju obranu, uz smanjenje operativnih troškova i izbjegavanje reputacijske štete (Muppidi et al., 2022).

S poboljšanom manipulacijom podacima su povezane poboljšana detekcija i klasifikacija anomalija, koje će biti detaljnije opisane u sljedećem potpoglavlju.

Sljedeća bitna karakteristika jest prediktivnost obrambenih mehanizama temeljenih na AI. Ona je povezana s prethodnim karakteristikama. U prošlosti je kibernetička sigurnost bila fokusirana na zaštitu infrastrukture i reagiranja na prijetnje. Korištenjem AI ona prelazi s proaktivnog na prediktivni pristup (Gregory, 2021). AI algoritmi ne samo da mogu proaktivno tražiti potencijalne prijetnje i ranjivosti unutar mreže organizacije već, analizirajući povijesne podatke, AI može identificirati obrasce koji prethode napadima i predvidjeti potencijalne buduće prijetnje. To, uz proaktivne obrambene mjere, omogućuje organizacijama da budu korak ispred napadača. Tvrtke mogu koristiti AI za prediktivnu inteligenciju rizika na četiri načina: donošenje odluka povezanih s rizikom, prepoznavanje rizika, praćenje prijetnji te otkrivanje i automatizacija procesa upravljanja rizicima (Gregory, 2021).

Upravo činjenica da korištenjem strojnog učenja u informacijskoj sigurnosti, njegovim uvođenjem i treniranjem, organizacija zauzima prediktivni pristup sigurnosti smatra se jednom od najvećih prednost njegove primjene (Erokhin, 2020).

Obrambene mehanizme temeljene na AI također odlikuje automatizacija i autonomnost (Sparapani i Ruma, 2021). Kao što je moguće automatizirati sve faze napada, tako je moguće automatizirati i sve obrambene procese- procese zaštite, prevencije, otkrivanja i odgovora (Parham, 2022). AI potpomognuti obrambeni mehanizmi mogu automatski reagirati na prijetnje i izvršavati potrebne akcije bez potrebe za ljudskim intervencijama. Na primjer, mogu blokirati ili izolirati zaražene uređaje ili mrežne segmente, pokretati zakrpe i ažuriranja, preusmjeravati promet kroz sigurnosne sustave radi dodatne provjere itd. (Muppidi et al., 2022). O automatizaciji sigurnosnih zadataka bit će više riječi u sljedećem potpoglavlju.

Kao i napadi potpomognuti sa AI, i obrambeni mehanizmi su prilagodljive i evoluirajuće prirode. Ova karakteristika proizlazi iz načina na koji AI uči. AI sustavi sposobni su generalizirati između zadataka—to jest, prilagoditi znanja i vještine koje je sustav stekao za jedan zadatak novim situacijama, s malo ili nimalo dodatne obuke. U novije vrijeme u fokusu istraživača su poboljšane tehnike dubokog učenja poput neprestanog učenja, čiji je cilj sve što je naučeno u prošlosti stalno primjenjivati na nove zadatke, a u učenju novih zadataka, izbjegavati uništavanje onoga što je već naučeno (Littman et al., 2021). U sigurnosti to znači da AI uči i iz iskustva i iz novih podataka o prijetnjama i prilagođava se promjenama u tehnologiji i taktikama napadača. To uključuje prepoznavanje obrazaca i trendova u podacima vezanim uz sigurnost, kontinuirano ažuriranje baza podataka o prijetnjama, usvajanje novih sigurnosnih tehnologija i postupaka te prilagodbu na nove

vrste napada (Muppidi et al., 2022). To sigurnosnim stručnjacima i njihovim mehanizmima omogućuje da ostanu učinkoviti čak i u okruženju nepredvidivih i dinamičkih prijetnji.

Daljnja karakteristika AI jest fleksibilnost i sposobnost integracija. Dobro implementirana AI je fleksibilna i omogućuje integraciju s različitim sigurnosnim alatima i sustavima kako bi se ostvarila cjelovita zaštita informacijske infrastrukture. To uključuje integraciju s vatrozidima, sustavima za detekciju intruzije, antivirusnim programima, sigurnosnim upravljačkim sustavima itd. (Muppidi et al., 2022). Isto je važno jer bi dobra strategija informacijske sigurnosti trebala osigurati da je usvajanje AI usklađeno sa strateškim sigurnosnim ciljevima organizacije te da je AI implementirana tamo gdje može donijeti najveće unaprjeđenje performansi. Primjenu AI potrebno je kombinirati s ostalim sigurnosnim mjerama, kao što su obuka osoblja, ažuriranje sigurnosnih politika, primjena najboljih praksi i redovito ažuriranje tehničkih komponenti sustava (Muppidi et al., 2022).

Nakon iznošenja karakteristika kibernetičkih napada koji koriste AI i karakteristika obrambenih mehanizama temeljenih na AI, sažeto ćemo ih prikazati u tablici 3., zbog jednostavnije usporedbe i uočavanja preklapanja.

Karakteristike AI kibernetičkih napada	Karakteristike AI obrambenih mehanizama
Prilagodljiva i evoluirajuća priroda	Prilagodljiva i evoluirajuća priroda
Automatizacija i skalabilnost	Automatizacija i autonomnost
Nedostatak objašnjivosti i interpretabilnosti	Prediktivnost
Poboljšana manipulacija podacima	Poboljšana manipulacija podacima
Pristupačnost i demokratizacija	Fleksibilnost i sposobnost integracija
Nedostatak etičkog kodeksa	

Tablica 3. Karakteristike AI napada i AI obrambenih mehanizama

Odgovor na AI napade je upravo korištenje AI. Slično kao što akteri prijetnji svakim napadom postaju pametniji, korištenjem AI za zaštitu i obranu i naši sustavi svakim napadom postaju učinkovitiji- kao što je Svjetski ekonomski forum jezgrovito rekao, samo AI može pobijediti AI u vlastitoj igri (Gregory, 2021). Iako AI ima snažne obrambene sposobnosti, važno je osigurati njezinu pravilnu implementaciju i kontinuirani nadzor. Redovita ažuriranja i fina podešavanja su neophodni za rješavanje prijetnji u nastajanju, izbjegavanje lažnih pozitivnih rezultata i ublažavanje pristranosti koje mogu utjecati na učinkovitost AI u informacijskoj sigurnosti. Osim toga, ljudska stručnost i dalje je ključna za tumačenje odgovora i preporuka koje daje AI te donošenje informiranih odluka (Dixon i Eagan, 2019).

5.2. Implementacija i djelotvornost obrambenih mehanizama temeljenih na umjetnoj inteligenciji

Posljednjih godina bilježi se značajan porast broja poslovnih transakcija, ulaganja rizičnog kapitala i korporativnih ulaganja u umjetnu inteligenciju i kibernetičku sigurnost. Preko 400 poslovnih transakcija primilo je ukupno 10 milijardi dolara ulaganja (Alawadhi et al., 2022). Od 2017. do 2019. broj investicija se više nego utrostručio, odražavajući kontinuirani rast tržišta AI u

kibernetičkoj sigurnosti. Očito je da organizacije prepoznaju prednosti primjene AI i smatraju ulaganja opravdanima.

Slika 6. prikazuje osnovne razloge za usvajanje AI u informacijskoj sigurnosti, prema istraživanju IBM Institute for Business Value (IBV), a temeljem anketiranja 1000 menadžera odgovornih za sigurnost informacijskih i operativnih tehnologija iz 16 industrija i 5 globalnih regija.

Slika 6. Razlozi usvajanja AI u sigurnosti
Izvor: izrada autora temeljem podataka IBM Institute for Business Value (IBV) preuzetih iz Muppidi, S., Fisher, L., Parham, G. (2022), AI and automation for cybersecurity: How leaders succeed by uniting technology and talent

Razlozi zašto neke organizacije oklijevaju s usvajanjem AI mogu biti organizacijski, poput misije ili vizije organizacije, organizacijske kulture, svijesti menadžmenta o informacijskoj sigurnosti (Kane et al., 2017); ekonomski, poput nedostatka financijskih sredstava za ulaganje u AI (Kant i Johannsen, 2022) ili tehnički, poput kompleksne postojeće infrastrukture, puno *legacy* sustava, nedovoljnog broja stručnjaka (Kant i Johannsen, 2022). Dodatan razlog je nemogućnost menadžmenta da procijeni jesu li ulaganja u moderne tehnologije opravdana i u kolikoj mjeri. Najčešći razlog za to jest pomanjkanje znanja o tim tehnologijama, zbog čega menadžeri ne mogu procijeniti opravdanost troškova, a još manje razumjeti benefite koje nije moguće kvantificirati, bar ne u kratkome roku (Forrester Consulting, 2019). Kako je ranije spomenuto, nedostatak znanja znači nerazumijevanje načina na koji AI donosi odluke i postiže rezultate i posljedično nedostatak povjerenja u te rezultate.

Implementacija obrambenih mehanizama temeljenih na AI ima određene preduvjete, kako bi bila uspješna. U nastavku je navedeno nekoliko ključnih preduvjeta implementacije.

Prvi preduvjet svakako su kvalitetni podaci. Implementacija AI mehanizama zahtijeva pristup kvalitetnim podacima o sigurnosnim prijetnjama, povijesnim incidentima, logovima sustava, korisničkim aktivnostima i drugim relevantnim izvorima podataka. Ti podaci trebaju biti dobro strukturirani, ažurni, pouzdani i relevantni za ciljeve obrane te je važno osigurati njihovu

dostupnost, autentičnost i integritet, kako bi AI imala pouzdane informacije za analizu i učenje (Krishnappa, 2015).

Naime, jedan od važnijih ciljeva primjene AI u informacijskoj sigurnosti jest riješiti ono na što stručnjak potroši određeno vrijeme, u vrlo kratkom vremenu. Dobiveni model mora moći izvršiti potrebne radnje na isti način na koji to radi savjetnik za informacijsku sigurnost, uključujući i u odsutnosti takvih stručnjaka. Svaki zadatak koji zahtijeva stručnu intervenciju može se modelirati korištenjem AI tehnika, ako su dostupni odgovarajući podaci. Ako se značajke povezane sa zadatkom mogu identificirati i mogu se prikupiti podaci koji predstavljaju te značajke, tada se zadatak može riješiti pomoću metoda strojnog učenja. U području informacijske sigurnosti važno je postići pravilno strukturiranje i obradu velikog broja ulaznih podataka za učinkovit rad automatiziranih sigurnosnih alata, jer razina kvalitete sigurnosnih alata temeljenih na AI ovisi o cjelovitosti podataka za obuku. I ovaj proces može biti optimiziran korištenjem mogućnosti umjetne inteligencije (Erokhin, 2020).

U kolikoj mjeri su dostupnost, kvaliteta i integritet podataka važni, govori činjenica da, ako je abnormalna aktivnost prisutna u podacima za treniranje, postoji mogućnost da će je AI kvalificirati kao normalnu i stoga je neće otkriti kada se ponovo pojavi (Erokhin, 2020).

Važno je imati pristup relevantnim algoritmima i modelima strojnog, posebno dubokog, učenja koji su prilagođeni za primjenu u sigurnosnim scenarijima. To može uključivati algoritme za klasifikaciju, detekciju anomalija i upada, predikciju prijetnji itd. Rekurentna i konvolucijska neuronska mreža (engl. *Recurrent Neural Network-RNN, Convolutional Neural Network- CNN*), višeslojni perceptron (engl. *Multi-layer Perceptron-MLP*) i dugotrajno kratkoročno pamćenje (engl. *Long–short-term Memory-LSTM*) popularni su pristupi koji se koriste u modeliranju dubokog učenja u području sigurnosti. Uz to se koriste različite metode nenadziranog učenja, kao što su autoenkoder (engl. *Autoencoder-AE*), duboka mreža uvjerenja (engl. *Deep Belief Network - DBN*), ograničeni Boltzmannovi strojevi (engl. *Restricted Boltzmann Machines-RBM*), generativne suparničke mreže (engl. *Generative Adversarial Network-GAN*) itd., kao i hibridni pristupi odnosno kombinacije ovih pristupa (Sarker et al., 2021).

Daljnji preduvjet je adekvatna infrastruktura koja omogućuje prikupljanje, obradu i pohranu podataka potrebnih za implementaciju AI mehanizama. AI mehanizmi zahtijevaju dovoljno snage obrade i resursa kako bi se mogli učinkovito nositi s velikim količinama podataka i složenim analitičkim postupcima (Krishnappa, 2015). Adekvatna infrastruktura uključuje brze i efikasne mehanizme za odabir, prikupljanje, agregaciju i predobradu podataka, robusne baze podataka za njihovu pohranu, visokokapacitetne procesore i poslužitelje, infrastrukturu oblaka, GPU-akceleraciju ili druge oblike ubrzanja obrade, veliku propusnost mreže i druge potrebne alate i tehnologije (Quach, 2021).

Implementacija AI mehanizama trebala bi biti integrirana s postojećom sigurnosnom infrastrukturom organizacije (Kant i Johannsen, 2022). To uključuje integraciju s vatrozidima, postojećim sustavima za detekciju intruzije, antivirusnim programima, sustavima upravljanja incidentima i drugim relevantnim sigurnosnim alatima, kako bi se postigla cjelovita i koherentna sigurnosna strategija (Muppidi et al., 2022). Postojeća sigurnosna infrastruktura kompatibilna s AI tehnologijama je zato također preduvjet (Kant i Johannsen, 2022). Osim toga, važno je osigurati

dobru koordinaciju između različitih sigurnosnih sustava kako bi se izbjegla konfuzija i poboljšala ukupna sigurnost.

Sljedeći preduvjet je stručnost kadrova u području sigurnosti i AI tehnologija (Sparapani i Ruma, 2021). Potrebno je imati tim stručnjaka koji razumiju nove sigurnosne prijetnje, nove tehnike obrane, algoritme strojnog učenja i dubokog učenja te razvoj i primjenu AI rješenja. Kada dio sigurnosnih zadataka preuzme AI, najtraženiji će biti stručnjaci sposobni razvijati sigurnosne AI mehanizme, kontrolirati kvalitetu njihove izvedbe i utemeljenost odluka donesenih korištenjem algoritama (Dixon i Eagan, 2019). Nedavno istraživanje IBM-a pokazalo je da je 34% rola u IT sigurnosti doživjelo promjene zahtjeva u pogledu potrebnih vještina, od kojih je 35% bilo izravno ili neizravno potaknuto usvajanjem AI (Muppidi et al., 2022). Većina stručnjaka osjeća da joj manjka znanja potrebnog za upravljanje AI rješenjima i njihov daljni razvoj. Također smatraju da je za informacijsku sigurnost i za razvoj zaposlenika bolje nabavljati, i za njih educirati, najnovija AI rješenja, a ne, kao što je ponekad slučaj, starija i jeftinija rješenja (Alawadhi et al., 2022).

Organizacije mogu doskočiti problemu pomanjkanja broja talenata na način da automatizaciju ne promatraju samo kao način optimizacije troškova, već kao način stvaranja specijalizacija i boljeg radnog iskustva, pomažući zaposlenicima da prošire svoje vještine (Muppidi et al., 2022).

Ovi preduvjeti su ključni za uspješnu implementaciju obrambenih mehanizama temeljenih na AI i osiguravaju visoku razinu sigurnosti informacijske infrastrukture.

Uz preduvjete postoje i neki zahtjevi dobrih praksi za uspješnu primjenu AI u informacijskoj sigurnosti. U nastavku je navedeno nekoliko ključnih zahtjeva.

Implementacija obrambenih AI mehanizama zahtijeva visok stupanj pouzdanosti (Krishnappa, 2015). Obrambeni sustavi moraju biti zaštićeni od napada, manipulacija ili iskorištavanja od strane napadača. Također je važno osigurati da AI mehanizmi ne stvaraju lažne pozitivne ili lažne negativne rezultate, jer to može dovesti do gubitka povjerenja u sustav (Krishnappa, 2015). Stoga je potrebno sustavno praćenje i evaluacija performansi AI mehanizama kako bi se utvrdila njihova učinkovitost i uspješnost u otkrivanju, sprječavanju i neutraliziranju sigurnosnih prijetnji. To uključuje praćenje lažno pozitivnih i lažno negativnih rezultata, ažuriranje algoritama i modela temeljenih na povratnim informacijama te kontinuirano poboljšavanje sustava (Muppidi et al., 2022).

Implementacija AI mehanizama zahtijeva i pažnju prema etičkim pitanjima te društvenom, ekonomskom i političkom kontekstu. Mogli bismo reći da je implementacija ne samo tehnički, već sociotehnički izazov. Potrebno je osigurati da se obrambeni mehanizmi ne koriste na način koji krši privatnost korisnika ili stvara diskriminaciju. Također je važno osigurati odgovornost i transparentnost u funkcioniranju AI sustava kada oni donose važne odluke, kao i razumijevanje tih odluka od strane svih dionika. Uspješno implementirani AI sustavi su promišljeno integrirani u postojeće društveno i organizacijsko okruženje i prakse (Littman et al., 2021).

I konačno, obrambeni AI mehanizmi moraju biti sigurni i zaštićeni. Ako se oslanjamo na algoritme strojnog učenja za otkrivanje kibernetičkih napada i odgovor na njih, ključno je da ti algoritmi budu zaštićeni od smetnji, ugrožavanja ili zloporabe. Što više ovisimo o AI sustavima za kritične funkcije

i usluge (autonomna vozila, zdravstvo, financijske usluge, energetika), to će napadači imati veći poticaj da ciljaju te sustave, a svaki uspješan napad na njih imat će tim teže posljedice (Wolff, 2020).

Vlade poduzimaju korake za reguliranje visokorizičnih AI aplikacija i promicanje odgovorne upotrebe AI, ali na strani napada, najpogubnije upotrebe se množe, a troškovi razvoja padaju, čineći bilo koji oblik zaštite ogromnim izazovom (Abeshu i Chilamkurti, 2018). Zbog utjecaja AI, posebno dubokog učenja, na kibernetičku sigurnost, krajolik prijetnji će nedvojbeno postati raznovrsniji, s pojavom novih prijetnji i prilagodbom postojećih. Ne samo da će AI sustavi postati ranjiviji na manipulaciju, već će napadači također imati pristup novim i učinkovitijim rutama napada. Dodatno, AI modeli otvorenog koda mogu i sami biti podložni hakiranju zbog nedovoljnih mjera kibernetičke sigurnosti za sprječavanje krađe osjetljivih podataka (Alawadhi et al., 2022).

Kako bi se smanjio rizik suparničkih napada pojavljuju se novi pristupi obrani. Prvi je suparničko strojno učenje. Jednako kao što ga koriste napadači, mogu ga koristiti i sigurnosni stručnjaci. Njegov je cilj učiniti sustave strojnog učenja otpornima na zlonamjerne napade proučavanjem napada koji se stalno razvijaju, ali i obrana od tih napada. Suparničko strojno učenje uglavnom koristi teoriju igara za modeliranje sukoba između sustava temeljenih na učenju i njihovih protivnika. Međutim, obzirom suparničko strojno učenje pretpostavlja da branitelji i napadači dijele neke informacije i znanje, što nije slučaj u informacijskoj sigurnosti, razvija se noviji pristup pod nazivom suparnička analiza rizika. Ovaj pristup naglasak daje predviđanju. Razvija modele o tome kako napadači napadaju i reagiraju te koristi to znanje da predvidi kako bi mogli napasti u budućnosti, bez pretpostavki o zajedničkom, općem znanju (Rios Insua et al., 2020).

Kako se organizacije u svojim sigurnosnim strategijama više oslanjaju na AI rješenja, tako će svoje sigurnosne stručnjake morati više educirati za obranu upravo tih rješenja. Obzirom automatizacija sigurnosnih zadataka oslobađa vrijeme stručnjaka, a dio zadataka zahvaljujući njoj može preuzeti i drugo IT osoblje, ključni sigurnosni talenti u organizaciji moraju se posvetiti proučavanju suparničkog strojnog učenja, suparničke analize rizika, zaštite AI sustava od trovanja podataka, sprječavanju curenja podataka ML modela itd. Istraživanje suparničkog strojnog učenja pokazalo je da pokušaj da se AI modele učini otpornijima na trovanje podataka i suparničke unose često uključuje izgradnju modela koji otkrivaju više informacija o pojedinačnim podatkovnim točkama koje se koriste za treniranje tih modela. Kada se osjetljivi podaci koriste za treniranje modela, ovo stvara novi skup sigurnosnih rizika, rizik da će protivnici moći pristupiti podacima o obuci ili zaključiti podatke o obuci iz samog modela. Pokušaj zaštite modela od ove vrste napada zaključivanjem može ih učiniti osjetljivijima na suparničke napade i obrnuto. To znači da je održavanje sigurnosti AI sustava osjetljiva ravnoteža između ova dva različita, ali povezana skupa rizika (Wolff, 2020).

Očekuje se da bi veći stupanj pouzdanosti i sigurnosti mogla donijeti i bolja regulacija te standardizacija razvoja AI rješenja. Ako vlade propišu potrebu da o procesu razvoja AI sustava mora postojati detaljna dokumentacija, time će potaknuti razvoj učinkovitih tehnika testiranja i revizije, kao i programa certifikacije koji pružaju jasne smjernice razvojnim programerima i korisnicima umjetne inteligencije. Ove bi revizije iskoristile znanja stečena o suparničkom strojnom učenju i curenju podataka ML modela, kako bi se modeli obavezno testirali na ranjivosti

i procijenila njihova ukupna robusnost i otpornost na različite oblike napada. Regulacija bi također razjasnila tko je odgovoran kada AI sustavi uzrokuju štetu zbog nedostatka odgovarajućih sigurnosnih mjera i koje su odgovarajuće kazne za tu štetu (Wolff, 2020).

Kada govorimo o standardizaciji, nužno je spomenuti rad Međunarodne organizacije za standardizaciju (engl. *International Organization for Standardization-ISO*). ISO i Međunarodna elektrotehnička komisija (engl. *International Electrotechnical Commission-IEC*) u standardu ISO/IEC TR 24028:2020 Informacijske tehnologije- Umjetna inteligencija- Pregled pouzdanosti umjetne inteligencije adresirali su u ovome radu analizirane karakteristike AI, načine njihova ublažavanja na strani prijetnje i načine njihova maksimalnog iskorištavanja na strani obrane. Ovaj dokument razmatra teme povezane s pouzdanošću u AI sustavima, uključujući pristupe za uspostavljanje povjerenja u AI sustave kroz transparentnost, objašnjivost, mogućnost kontrole itd., inženjerske zamke i tipične povezane prijetnje i rizike za AI sustave, s mogućim tehnikama i metodama ublažavanja te pristupe za procjenu i postizanje dostupnosti, otpornosti, pouzdanosti, točnosti, sigurnosti, zaštite i privatnosti AI sustava (ISO, 2020).

Nakon ovoga dokumenta, nastao je dokument ISO/IEC TR 24030:2021 Informacijske tehnologije- Umjetna inteligencija- Slučajevi korištenja, koji je zbirka ukupno 132 slučaja korištenja AI u različitim domenama. Cilj dokumenta je prikaz primjenjivosti rada na standardizaciji AI u različitim industrijama i područjima djelovanja, stvaranje temelja i referenci za rad na standardizaciji AI, dijeljenje prikupljenih slučajeva korištenja kao potpore radu na standardizaciji te uspostavljanje suradnje dionika na prikupljanju novih tehničkih zahtjeva za AI, s ciljem ubrzanja znanstvenih i tehnoloških dostignuća (ISO, 2021). Trenutno se radi na drugoj verziji ovoga dokumenta, a u svibnju 2023. objavljeno je i prvo izdanje povezanog tehničkog izvješća ISO/IEC TR 27563 Sigurnost i privatnost u slučajevima korištenja umjetne inteligencije- Dobre prakse. Ovaj dokument opisuje najbolje prakse za procjenu sigurnosti i privatnosti u slučajevima korištenja AI objavljenima u ISO/IEC TR 24030. Dokument daje ukupnu procjenu sigurnosti i privatnosti u AI sustavima od interesa, opisuje rizike sigurnosti i privatnosti, kontrole sigurnosti i privatnosti te iznosi načine osiguranja sigurnosti i privatnosti i pripadajuće planove njihova osiguranja (ISO, 2023). Iznesenim smjernicama potrebno se voditi pri implementaciji AI sustava u opisanim i svim drugim industrijama te u svim oblicima poslovnih, vladinih i drugih organizacija.

Opisani preduvjeti uspješne implementacije i zahtjevi dobrih praksi sistematizirani su u tablici 4.

Preduvjeti uspješne implementacije	Zahtjevi dobrih praksi
Kvalitetni podaci	Visok stupanj pouzdanosti algoritama
Pristup relevantnim algoritmima i modelima	Pažnja prema etičkim pitanjima
Adekvatna infrastruktura	Pažnja prema društvenom, ekonomskom i političkom kontekstu
Postojeća sigurnosna infrastruktura kompatibilna s AI tehnologijama	Visok stupanj sigurnosti i zaštićenosti algoritama
Stručnost kadrova	Bolja regulacija i standardizacija razvoja AI rješenja

Tablica 4. Preduvjeti uspješne implementacije i zahtjevi dobrih praksi

U nastavku će biti opisane konkretne primjene obrambenih mehanizama temeljenih na umjetnoj inteligenciji u informacijskoj sigurnosti.

AI mehanizmi mogu se koristiti za nadzor mreže i krajnjih točaka. Donedavno je jedina opcija nadzora bila prikupljanje i analiza zapisa iz kritičnih sustava. Pristup otkrivanju prijetnji orijentiran na zapise ostavlja mnoge "slijepe točke" koje sofisticirani protivnici mogu iskoristiti. Nasuprot njemu, sigurnost vođena inteligencijom ima za cilj eliminirati slijepe točke pružanjem sveobuhvatne vidljivosti, kako na mreži tako i na krajnjim točkama, kao što su poslužitelji i računala zaposlenika. Vidljivost se postiže uključivanjem mogućnosti mrežnog hvatanja paketa (engl. *Packet-capture*), koje se odnosi na snimanje, parsiranje, normaliziranje, analiziranje i ponovno sastavljanje cjelokupnog podatkovnog prometa na svakom sloju mrežnog stoga (RSA, 2015).

Sigurnosna rješenja vođena inteligencijom također pružaju duboku vidljivost aktivnosti na krajnjim točkama, uključujući poslužitelje i prijenosna računala. Potrebno je analizirati i ono što se događa u memoriji računala i ono što je pohranjeno na fizičkom disku te usporediti pokrenute procese s datotekama na disku, kako bi se dobio uvid u to jesu li izvršne datoteke i procesi krajnje točke legitimni ili su zlonamjerno ubačeni. Duboki rendgenski pregled krajnjih točaka i automatizirano otkrivanje sumnjivih aktivnosti ključni su za brže prepoznavanje i istraživanje prijetnji. Pritom nema oslanjanja na operativne sustave i hipervizore, koji i sami mogu biti kompromitirani (RSA, 2015).

AI mehanizmi mogu se koristiti za automatizirano otkrivanje iskorištavanja ranjivosti (engl. *Automated Exploit Detection*). Različite vrste napada nastoje iskoristiti različite vrste ranjivosti. Npr. neki od R2L (engl. *Remote to User Attacks*) napada kao što su *imap*, *named* i *sendmail* iskorištavaju prekoračenje međuspremnika (engl. *Buffer Overflow*) u mrežnim programima, dok drugi napadi kao što su *dictionary*, *ftp-write* i *guest* iskorištavaju slaba ili pogrešno konfigurirana sigurnosna pravila. R2L napadi dovode do napada eskalacijom privilegija zvanih U2R (ngl. *User to Root*) napadi, koji su uzrokovani loše kodiranim sistemskim programima koji rade kao *root*. Neki napadi ove kategorije, kao što su *loadmodule* i *Perl*, mogu iskoristiti slabosti u provjeri naziva staze. Određene kategorije DoS (engl. *Denial of Service*) napada, kao što su *apache2*, *back* i *syslogd*, iskorištavaju softverske greške u mrežnim *daemonima* (Abeshu i Chilamkurti, 2018).

Dakle, automatizirano otkrivanje iskorištavanja odnosi se na proces identificiranja i otkrivanja specifičnih softverskih ranjivosti, ranjivosti u konfiguraciji sustava ili konfiguraciji sigurnosnih pravila, koje se mogu iskoristiti za kompromitiranje sustava. Uključuje skeniranje softverskog koda i konfiguracije sustava te analizu tih i drugih relevantnih podataka za prepoznavanje poznatih ranjivosti koje bi napadači mogli iskoristiti ili obrazaca zlonamjernih aktivnosti i ponašanja. Cilj automatskog otkrivanja iskorištavanja je proaktivno identificirati i zakrpati ranjivosti prije nego što se mogu iskoristiti odnosno u stvarnom vremenu. U tu svrhu, algoritmi strojnog, a posebno dubokog učenja, mogu se uvježbati na velikim skupovima podataka poznatih ranjivosti i pripadajućih eksploatacija, za razvoj modela koji ih mogu automatski otkriti, klasificirati i zakrpati. Ovakvi sustavi, koji kombiniraju razne alate, tehnike i stručno znanje za stvaranje potpuno autonomnih sustava za otkrivanje ranjivosti, generiranje mogućih iskorištavanja i softversko

krpanje, bez ljudske intervencije, često se nazivaju i *cyber* sustavi rasuđivanja (engl. *Cyber Reasoning Systems*) (Brooks, 2019).

Korištenjem strojnog učenja moguće je postići znatno bolje rezultate u detekciji i klasifikaciji softverskih ranjivosti te je prema Gartneru detekcija ranjivosti primjena AI koja trenutno ima najveću tržišnu zrelost (Kant i Johannsen, 2022).

Kombinacija algoritama strojnog i dubokog učenja i analize anomalija omogućuje i automatiziranu detekciju upada (engl. *Automated Intrusion Detection*) koji se ne bi mogli lako prepoznati konvencionalnim sigurnosnim sustavima. Primarni cilj automatizirane detekcije upada je otkriti i odgovoriti na pokušaje neovlaštenog pristupa, infekcije zlonamjernim softverom ili druge sumnjive aktivnosti u stvarnom vremenu. Uključuje korištenje različitih tehnika, kao što je statističko otkrivanje anomalija, otkrivanje na temelju potpisa i analiza ponašanja, kako bi se otkrili pokazatelji potencijalnog upada ili ugrožavanja (Krishnappa, 2015). Na primjer, automatizirani AI alati mogu analizirati mrežni promet i logove, otkriti kada netko skenira portove od glavnog računala do glavnog računala ili šalje velike količine podataka koje nisu planirane (Erokhin, 2020). Na ovaj način moguće je otkrivati i blokirati zlonamjerne aktivnosti, napade usmjerene na aplikacije, DDoS napade i druge sigurnosne prijetnje. Sve relevantne informacije se agregiraju, analiziraju i unose u model dubokog učenja. To omogućuje da se pouzdano predvidi vjerojatnost da će određena vrsta prometa biti zlonamjerna. Ovu analizu treba izvesti brzo, svodeći vrijeme između prepoznavanja i reakcije na minimum (Erokhin, 2020). Strojno učenje može generirati i statistički profil normalnog korisnika, aktivnosti uređaja ili web mjesta. Sofisticirani napadači mogu zaobići takve statične pristupe nadzoru modificiranjem redaka koda, postavljanjem novog virtualnog stroja u javnom oblaku ili registriranjem nove internetske domene kao naredbeno-kontrolne ili "ispuštajuće" stranice (RSA, 2015). No teško će zaobići ili prevariti bihevioralnu analitiku, koja omogućuje prevenciju štete od napada koji nisu primijećeni standardnim alatima za upravljanje prijetnjama, uključujući one temeljene na zlouporabi legitimnih validacijskih podataka (Erokhin, 2020). Sigurnosni sustavi vođeni inteligencijom utvrđuju kako "dobro" ponašanje unutar određenog IT okruženja izgleda praćenjem i učenjem raznih aktivnosti strojeva i ljudi. Na primjer, koji se portovi na poslužiteljima obično koriste za vanjsku komunikaciju, koje su uobičajene lokacije s kojih se korisnici prijavljuju i uobičajene navike zaposlenika. Sigurnosni analitičari označavaju aktivnosti za koje se primijeti da su izvan uobičajenog ponašanja. Ako analitičari odbace događaj kao lažno pozitivan, sigurnosni alati mogu učiti iz tog iskustva i manje je vjerojatno da će označiti buduća ponavljanja kao prijetnju (RSA, 2015).

AI mehanizmi mogu identificirati čak i napredne prijetnje i zlonamjerni softver koji koristi sofisticirane tehnike skrivanja ili promjene svojih oblika, poput naprednih ustrajnih prijetnji (engl. *Advanced Persistent Threats*) i steganografije, a kombinirajući tehnike strojnog učenja, analize anomalija i heurističke analize mogu otkriti i blokirati čak i nove i nepoznate prijetnje (RSA, 2015).

AI mehanizmi mogu se koristiti za automatizirano prepoznavanje i odgovaranje na sigurnosne incidente (engl. *Automated Incident Response*). Ovisno o vrsti incidenta, moguće je primijeniti odgovarajuće algoritme strojnog učenja za brzu identifikaciju, analizu i neutralizaciju prijetnje kako bi se smanjile štete i vrijeme reakcije. Upozorenja iz više sustava sigurnosnog nadzora agregiraju se u jednu konzolu za upravljanje sigurnošću, putem koje analitičari mogu pregledati

izvore podataka, pogođene strojeve i kontekstualne informacije povezane s incidentom. Obzirom omogućava bržu istragu incidenata, procjenu njihove ozbiljnosti i prioretizaciju, AI podržava i točnije predviđanje ozbiljnih curenja podataka i reakciju na njih. Brzina odgovora izravno ovisi o razini automatizacije koju pružaju AI i strojno učenje (Erokhin, 2020).

U kombinaciji sa Velikim podacima, AI mehanizmi mogu povećati učinkovitost analitike sigurnosnih podataka. Sigurnosni podaci kao što su zapisi sustava, mrežni paketi i aktivnosti krajnjih točaka prikupljaju se iz različitih izvora, indeksiraju, analiziraju i pohranjuju na način pogodan za centralizirano pretraživanje i analizu. Rekonstrukcija korisničkih sesija i aktivnosti vođena inteligencijom otkriva ne samo osnovne pojedinosti kao što su vrijeme ili IP adresa prijenosa paketa podataka, već koje su informacije primljene i odaslane te posljedičnu štetu. To omogućava organizacijama da otkriju sigurnosne anomalije i rekonstruiraju sigurnosne incidente sa sigurnošću i detaljima koji omogućuju brže istrage i sanacije incidenata. Interni sigurnosni podaci i vidljivost dodatno su obogaćeni podacima o prijetnjama iz vanjskih izvora, koji omogućuju organizacijama da uče iz tuđih iskustava kako bi poboljšale vlastite sposobnosti otkrivanja prijetnji (RSA, 2015). Poboljšana i centralizirana analitika sigurnosnih podataka znači i poboljšano izvještavanje o sigurnosnim incidentima.

U kombinaciji sa Velikim podacima, tehnike strojnog učenja mogu biti korisne za procjenu rizika, posebno u djelatnostima u kojima su uporabljivost aplikacije i korisničko iskustvo gotovo jednako važni kao i sigurnost. U djelatnostima poput e-trgovine svaki dodatni korak potreban za dovršetak transakcija može dovesti do odustajanja i negativno utjecati na prihod. Veliki podaci i strojno učenjem mogu prikupljati i analizirati različite korisničke podatke (IP adresa, vrsta uređaja, lokacija uređaja, preglednik, MAC adresa, ISP, korisnička povijest itd.). Ako je temeljem tih karakteristika rizik visok, provest će se dodatne sigurnosne mjere, poput dvofaktorske autentifikacije. Na taj način potreba za sigurnošću neće utjecati na uporabljivost (Krishnappa, 2015).

Obrambena rješenja temeljena na AI koriste se za sigurno upravljanje infrastrukturom u oblaku. Ona pružaju kontinuiranu analizu i praćenje sigurnosnih događaja u oblaku te automatske odgovore na prijetnje, kako bi se osigurala sigurnost podataka i zaštitili resursi u oblaku. Strojno učenje i automatizacija mogu znatno bolje odgovoriti na nemilosrdnu brzinu i razmjere operacija u hibridnom *multicloud* okruženju od čak i najkvalificiranijih stručnjaka za sigurnost. U kombinaciji sa modelom podijeljene odgovornosti koji je svojstven sigurnosti u oblaku i IT integracijom svojstvenom pristupu nultog povjerenja, AI automatizacija će sigurnosne operacije u oblaku dovesti na sasvim novu razinu (Muppidi et al., 2022).

AI se koristi za razvoj rješenja koja prepoznaju i sprječavaju prijevare (engl. *Fraud Detection Systems*) poput krađe identiteta, prijevara s kreditnim karticama ili financijskih prijevara. Ova rješenja analiziraju velike količine podataka o uređajima, korisnicima i njihovim ponašanjima, generiraju korisnički profil te prepoznaju nepravilnosti i generiraju upozorenja na potencijalne prijevare (Muppidi et al., 2022).

AI mehanizmi mogu se koristiti za klasifikaciju e-pošte kao "spam" ili "nije spam". Koriste se tehnike obrade prirodnog jezika i strojnog učenja kako bi se analizirao sadržaj i karakteristike e-pošte te identificirali neželjeni ili zlonamjerni sadržaji.

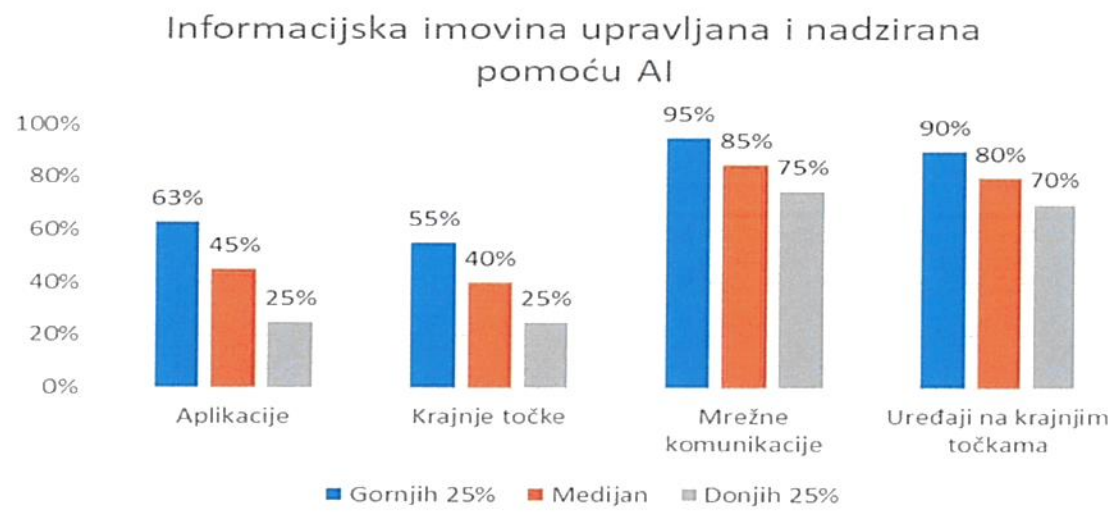
Također ih je moguće koristiti i za blokiranje botova. AI sustavi mogu dešifrirati obrasce organskog prometa i razlikovati legitimne botove, kao što su alati za indeksiranje tražilica, od zlonamjernih botova (Alawadhi et al., 2022).

Strojno učenje može pomoći i u automatizaciji sigurnosnih zadataka odnosno obavljanju ponavljajućih i zamornih zadataka, omogućujući stručnjacima da se koncentriraju na složenije poslove. Prema EMSI-ju, agenciji za analitiku tržišta rada u Sjedinjenim američkim državama, na 100 rola potrebnih u IT sigurnosti postoji samo 68 kvalificiranih kandidata, od kojih su mnogi već zaposleni i dobro plaćeni. Nedavna studija IBV-a pokazala je pak da organizacijama treba 150 dana da popune natječaj kvalificiranim kandidatom (Muppidi et al., 2022).

AI automatizacija može podržati analitičare u procesima upravljanja znanjem, upravljanja slučajevima i operativnoj podršci, npr. chatbotovima na prvoj liniji ili bazama znanja podržanima s NLP (Muppidi et al., 2022). Obzirom na promjenjivu i evoluirajuću prirodu kibernetičkih napada, čak i kada organizacija raspolaže velikim brojem stručnjaka, njima treba pomoć da se nose s velikim mrežnim prometom, novim načinima kršenja privatnosti i poboljšanim vektorima napada koji postaju teški za savladavanje od strane ljudi. Strojno učenje može automatizirati zadatke kao što su pregledavanje mrežnog prometa, sprječavanje prijetnji poput ransomware, eliminacija virusa i analiza mrežnih logova (Alawadhi et al., 2022).

Kao odgovor na izazov manjka osoblja, organizacije koje usvajaju AI automatizaciju navode da time prije svega nastoje poboljšati produktivnost i radno iskustvo za preopterećene resurse. 43% njih navodi povećanje produktivnosti sigurnosnih resursa kao glavni pokretač za korištenje AI. 42% navodi da je cilj smanjenje broja sigurnosnih događaja, incidenata i proboja u sustave, a 38% usredotočeni su na korištenje AI za poboljšanje točnosti rada analitičara. AI i automatizacija mogu imati dramatičan pozitivan utjecaj na sposobnost nošenja s volumenom i tempom sigurnosnih događaja, što je ključan čimbenik poboljšanja radnog okruženja sigurnosnih analitičara. Konačni ishod su dobitak u kapacitetu i specijalizacija radne snage IT sigurnosti (Muppidi et al., 2022).

Također, automatizacija omogućava organizacijama da učinkovitije upravljaju sa više krajnjih točaka i više svojih aplikacija, kao i povećaju kvalitetu nadzora mrežne komunikacije. Na slici 7. vidljivo je da organizacije koje su usvojile AI s najboljim rezultatima koriste automatizirano upravljanje identitetima za 63% aplikacija i 55% krajnjih točaka. Iste organizacije koriste automatizirane AI alate za nadzor 95% mrežnih komunikacija i 90% uređaja na krajnjim točkama (Muppidi et al., 2022).



Slika 7. Postotak informacijske imovine upravljane i nadzirane pomoću AI
Izvor: izrada autora temeljem podataka IBM Institute for Business Value preuzetih iz: Muppidi, S., Fisher, L., Parham, G. (2022), AI and automation for cybersecurity: How leaders succeed by uniting technology and talent

AI rješenja koriste se i kao pomoć pri analizi, reinženjeringu i zamjeni *legacy* sustava. Pomažu organizacijama da uštede novac i vrijeme te izbjegnu pogreške i štete u upravljanju *legacy* sustavima. Postoji direktan negativan utjecaj pogreški pri zamjeni *legacy* sustava na informacijsku sigurnost. AI rješenja mogu samostalno provesti procjenu, analizu, unaprjeđenje i reinženjering ili pak zamijeniti *legacy* sustave, bez očekivanih ili neočekivanih zastoja, grešaka ili troškova za organizacije. Jedina ljudska intervencija potrebna je u nadzoru i kontroli točnosti i kvalitete rada AI rješenja. Primjer ovakvog AI rješenja jest GenProg koji detektira i otklanja greške u *legacy* sustavima. Kada primjena AI i podrazumijeva dodatni trošak, on je manji od potencijalnih troškova curenja podataka ili oporavka od katastrofe *legacy* sustava (Aljindi, 2015).

AI tehnologije također se mogu pokazati učinkovitima u zaštiti kritične informacijske infrastrukture. Međutim, u ovom kontekstu, za razliku od modela oblaka, koriste se lokalna rješenja koja ne prenose podatke u vanjski svijet. Prema istraživanju Kaspersky Laba, tipično industrijsko postrojenje generira do 10 000 signala dnevno (podaci senzora, itd.). Činjenica da su svi podaci povezani i temeljeni na zakonima fizike može se koristiti za stvaranje automatizirane obrane. Obzirom napad na jedan element industrijskog sustava utječe na signale koje generiraju drugi dijelovi sustava, AI sustavi mogu naučiti odnose između signala i generirati predviđanja o tome kako će promjena u jednome od njih utjecati na druge (Erokhin, 2020).

Zaštita kritične informacijske infrastrukture je posebno važna jer je danas teško zamisliti kritičnu informacijsku infrasturukturu koja može funkcionirati potpuno neovisno o cyberprostoru. Čak i bez izravne veze s internetom, modernim sustavima upravljanja i drugim objektima kritične infrastrukture pristupaju brojni vanjski izvori. Preko cyberprostora ona je pak povezana s drugim kritičnim infrastrukturama, pa je moguće da ugroza za jednu postane rizik za više njih. Ugroza kritične infrastrukture je također i ugroza nacionalne sigurnosti (Mikhalevich i Trapeznikov, 2019).

Od novijih primjena AI možemo izdvojiti samoorganizirajuće mreže (engl. *Self Organising Network-SON*). Samoorganizirajuće mreže su radio pristupne mreže (engl. *Radio Access Network-RAN*) koje se automatski i samostalno planiraju, konfiguriraju, upravljaju, optimiziraju i same se iscjeljuju. Ovo je u suprotnosti s tradicionalnom implementacijom mobilnih bežičnih mreža, od kojih većina zahtjeva timove tehničara za održavanje, upravljanje i optimizaciju. SON može ponuditi niz različitih funkcija, uključujući samokonfiguraciju, samooptimizaciju, samoozdravljenje i samozaštitu. Ove funkcije omogućene su umjetnom inteligencijom, prediktivnom analitikom i unaprijed optimiziranim softverskim algoritmima. Tri su glavne vrste samoorganizirajućih mreža: distribuirane, centralizirane i hibridne. Iako je danas dostupno mnogo alata za upravljanje infrastrukturom i aplikacijama, mnogi mogu samo automatizirati popravke i rješenja pomoću nefleksibilnih unaprijed pripremljenih skripti, koje trebaju stalne promjene i optimizacije kako bi držale korak s dinamičnim mrežnim okruženjem. Za razliku od skripti, SON-ovi mogu automatski učiti i prilagođavati se promjenama na mreži te uzimati u obzir sve elemente mreže prije primjene promjene ili postavljanja konfiguracije. Ova mogućnost inteligentne procjene mrežne topologije prije unošenja promjena omogućuje skaliranje i implementaciju promjena brže nego ikad prije (Celona, 2021).

Primjene obrambenih mehanizama temeljenih na AI sistematizirane su u tablici 5.

Primjene obrambenih mehanizama temeljenih na AI
Nadzor mreže i krajnih točaka
Automatizirano otkrivanje iskorištavanja ranjivosti
Automatizirana detekcija upada
Identifikacija naprednih prijetnji
Automatizirano prepoznavanje i odgovaranje na sigurnosne incidente
Analitika sigurnosnih podataka
Procjena rizika
Sigurno upravljanje infrastrukturom u oblaku
Otkrivanje prijevara
Klasifikacija e-pošte
Blokiranje botova
Automatizacija sigurnosnih zadataka
Analiza, reinženjering i zamjena *legacy* sustava
Zaštita kritične informacijske infrastrukture
Samoorganizirajuće mreže

Tablica 5. Primjene obrambenih mehanizama temeljenih na AI

Iz navedenih konkretnih primjena moguće je zaključiti da korištenje AI tehnologija može pomoći u rješavanju mnogih problema informacijske sigurnosti: bihevioralna analiza radnji korisnika, otkrivanje mrežnih anomalija, predviđanje incidenata i odgovor na njih, otkrivanje zlonamjernog softvera, otkrivanje znakova kompromitacije sustava u lokalnoj mreži ili u oblaku, otkrivanje prijevara, aplikacije za filtriranje neželjene pošte, otkrivanje i sprječavanje upada u mrežu, otkrivanje botneta, sigurna autentifikacija (biometrijska) korisnika, ocjene kibernetičke sigurnosti itd. (Erokhin, 2020).

Prema istraživanju IBM Institute for Business Value organizacije usvojitelji AI u području sigurnosti naveli su otkrivanje krajnjih točaka i upravljanje imovinom kao vodeći slučaj korištenja AI. Trenutno ih 35% koristi AI u ovu svrhu s planovima za povećanje korištenja na gotovo 50% za 3 godine. Sljedeća najzastupljenija primjena je upravljanje ranjivostima i zakrpama sa 34%. Iste organizacije očekuju da će povećati svoju upotrebu AI za zaštitu i prevenciju za oko 40% u prosjeku u sljedeće 3 godine (Muppidi et al., 2022).

Ovo potpoglavlje zaključit ćemo raspravom o djelotvornosti AI obrambenih mehanizama i isplativosti ulaganja u njih.

Procjena djelotvornosti i efikasnosti AI obrambenih mehanizama djelomično se može provoditi istim metodama kojima se ocjenjuju i tradicionalni obrambeni mehanizmi. To podrazumijeva ocjenjivanje njihove izvedbe u otkrivanju, sprječavanju i ublažavanju cyber prijetnji opće prihvaćenim metrikama i ključnim pokazateljima izvedbe poput stope otkrivanja (engl. *Detection Rate ili True Positive Rate*), stope lažno pozitivnih (engl. *False Positive Rate*) i lažno negativnih rezultata (engl. *False Negative Rate*), preciznosti (engl. *Precision* ili *Positive Predictive Value*), osjetljivosti (engl. *Recall* ili *Sensitivity*), srednjeg vremena otkrivanja (engl. *Mean Time to Detect-MTTD*), srednjeg vremena za rješavanje (engl. *Mean Time to Resolve-MTTR*), srednjeg vremena do oporavka (engl. *Mean Time to Recover- MTTR*) itd. (SecurityScorecard, 2023).

Međutim, u ovome radu smo dosada naglašavali specifičnosti i tipične karakteristike AI tehnologija, pa ćemo se i na mjerenje performansi i efikasnosti osvrnuti sa aspekta specifičnosti u mjerenjima koje proizlaze iz opisanih karakteristika.

Specifične metrike još su u nastajanju i time se bavi više organizacija i istraživačkih timova.

National Institute of Standards and Technology (NIST) ima dugu povijest mjerenja i evaluacije AI sustava. Još u kasnim 1960-ima provodili su evaluaciju automatiziranih sustava za identifikaciju otiska prsta, a otada su proveli procjene tisuća AI sustava. NIST je angažiran u nastojanju da se uspostave zajedničke terminologije, definicije i taksonomije koncepata koji se odnose na različite karakteristike AI tehnologija. Te karakteristike uključuju točnost, objašnjivost i interpretabilnost, privatnost, pouzdanost, robusnost, sigurnost, otpornost i ublažavanje pristranosti. Svaki zahtijeva vlastiti portfelj mjerenja i evaluacija, a način na koji se određena komponenta mjeri i ocjenjuje može se promijeniti ovisno o kontekstu u kojem AI sustav radi. NIST u fokus stavlja kontekst. Za svaku karakteristiku NIST ima za cilj stvoriti, dokumentirati i poboljšati definicije, primjene, zadatke te prednosti i ograničenja metrika i mjernih metoda. NIST je također razvio i održava smislene skupove podataka s obzirom na odabrane atribute od interesa (NIST, 2023).

The National Cybersecurity Center of Excellence (NCCoE) gradi NCCoE AI Software-Testbed, modularnu testnu platformu za organiziranje i izvođenje eksperimenata strojnog učenja, trenutno usredotočenu upravo na sigurnosna pitanja (NIST, 2023).

Michael Rich i suradnici dovode u pitanje tradicionalne metode evaluacije strojnog učenja te predlažu metodu evaluacije usmjerenu na vrijednost koja uključuje pondere specifične za organizaciju evaluatora i odabir praga osjetljivosti specifičnog za vrijednosti povezane s kibernetičkom sigurnošću. Obzirom klasifikatori nikada neće biti savršeni, potrebno je procijeniti

rizik i troškove povezane s lažno pozitivnim i lažno negativnim rezultatima, a procjena bi trebala biti specifična za domenu. Na primjer, trošak Amazona koji ne preporuči proizvod potrošaču je propuštena prodajna prilika, dok lažno negativni rezultati u otkrivanju upada mogu rezultirati kršenjem privatnosti velike količine osobnih podataka. Odluke koje donose algoritmi strojnog učenja trebale bi stoga biti u korelaciji s vrijednostima specifičnim za organizacijsku misiju i imovinu koja se njima štiti. Razmišljanje usmjereno na vrijednost (engl. *Value-Focused Thinking-VFT*) je metoda koja u evaluaciju modela strojnog učenja uključuje vrijednosti specifične za optimizaciju algoritama koje koristimo u kibernetičkoj sigurnosti

Vrijednosti definira donositelj odluka (evaluator), a organizirane su u hijerarhijsku strukturu razina gdje niže grane predstavljaju podvrijednosti nadređenih vrijednosti. Procjenitelj ponderira sve vrijednosti, uz ograničenje da zbroj pondera u granama svake razine mora iznositi jedan. Jednostavna matematička funkcija uključuje navedene težine i metrike za svaku vrijednost za izračunavanje rezultata koji se koristi za ocjenu odluka- najbolja odluka je ona s najvišim rezultatom. Vrijednosti razine 1 relevantne za klasifikatore korištene u hibridnoj IDS konfiguraciji sa sustavom temeljenim na potpisu su stopa otkrivanja poznatih prijetnji, preciznost i stopa otkrivanja nepoznatih prijetnji. Uzimamo u obzir stopu otkrivanja za svaku klasu napada na razini 2, dopuštajući evaluatoru da ponderira vrijednost otkrivanja specifičnih klasa napada. Na primjer, klase napada mogu biti mrežni protokoli za mrežna IDS upozorenja ili vrsta zlonamjernog softvera za antivirusne sustave. Računaju se *figures of merit* (FOM) za svaki klasifikator i svaku vrstu napada uz određeni prag predikcije, a potom FOM za ukupne performanse klasifikatora za sve napade uz određeni prag predikcije. Optimalni pragovi se zatim mogu usporediti za više klasifikatora kako bi se odabrala optimalna kombinacija klasifikatora i praga (Rich et al., 2016).

Program za istraživanje i razvoj mrežnih i informacijskih tehnologija (engl. *Networking and Information Technology Research and Development- NITRD*) ističe koncept nazvan otpornost specifična za misiju (engl. *Mission Related Resilience*) (McDaniel et al., 2020). Svaki AI sustav vođen misijom koji donosi sigurnosne odluke pri odlučivanje bi nužno trebao uzimati u obzir namjeru vođe (bilo da je to vojni zapovjednik ili *chief informations officer*). Ključno istraživačko pitanje je kako je tu namjeru sustavu moguće izraziti. Jedna mogućnost je izražavanje prirodnim jezikom koji AI sustavi mogu koristiti za prijevod "naredbe" ili misije u nešto što je moguće adresirati autonomnim sustavom odlučivanja. Kada sustav namjeru jasno razumije, može njome informirati svoje izbore i prilagodbe otpornosti. Npr. može odlučiti da napadače koji miruju neće iskorijeniti jer to može biti skuplje ili više štetno od mogućeg napada (McDaniel et al., 2020).

AI usmjeren na misiju također može podržati planiranje i izvođenje u inženjeringu sigurnosti. Prvi korak je identificirati informacijsku imovinu ključnu za uspjeh misije ili tzv. ključni *cyber* teren. Ključna imovina se može promijeniti kako se mijenja svrha ili ciljevi misije. Dakle, AI može pomoći u prepoznavanju relevantnih aspekata podataka, klasifikaciji relevantnih informacija i prioretizaciji drugih sigurnosnih čimbenika. Na taj način AI sustavi koje koristimo mogu evoluirati zajedno s organizacijom (McDaniel et al., 2020).

Različite sigurnosne mjere dizajnirane za različite računalne resurse mogu različito djelovati. Jedan autonomni agent može raditi na postavljanju kompleksnog "traga prijevare" kako bi zbunio

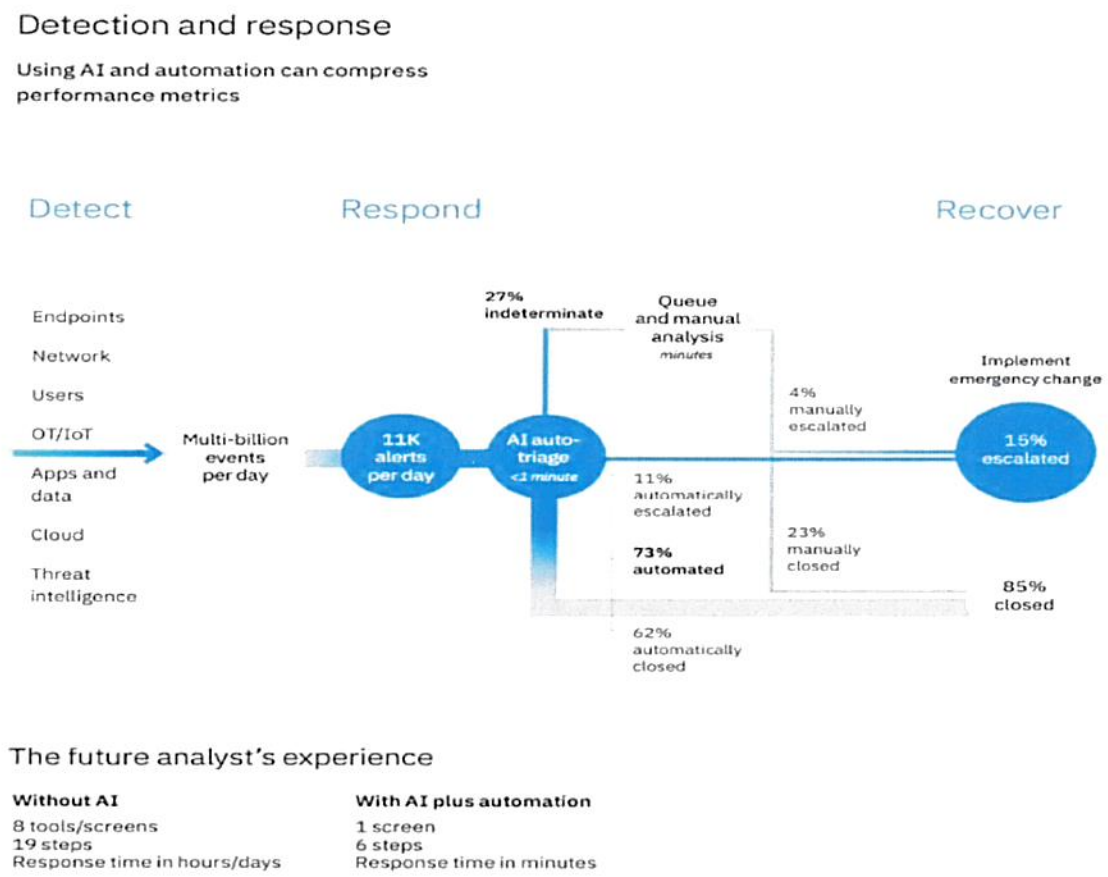
napadača dok će drugi agent pokušati pojednostaviti strukturu mreže kako bi smanjio površinu napada (McDaniel et al., 2020). Prema ovoj teoriji, oba mogu biti djelotvorna.

IBM Security Services objavio je, na temelju analize agregiranih podataka o izvedbi iz 2021. godine, da ulaganje u sigurnosnu AI i automatizaciju dovodi do opipljivih prednosti izvedbe. U usporedbi s korisnicima koji ne koriste AI, oni koji usvajaju AI mogu uštedjeti više od 14 tjedana u otkrivanju prijetnji i odgovoru na njih. Organizacije postižu ovu razinu performansi istovremeno smanjujući troškove i složenost zadataka (Parham, 2022).

Izvješće IBM-a i Ponemon Instituta „The 2022 Cost of a Data Breach report" pokazalo je da su sigurnosna umjetna inteligencija i automatizacija imali najveći pozitivan učinak na smanjenje ukupnih troškova povrede podataka. Za rješavanje prijetnji u nastajanju, IBM X-Force AnnualThreat Intelligence Index predlaže najbolje prakse kao što su usvajanje pristupa nultog povjerenja, automatizacija odgovora na incidente i implementacija proširene detekcije i odgovora (IBM, 2022).

Prema već spomenutom istraživanju IBM Institute for Business Value, organizacije usvojitelji AI koji su uspješno spojili uvide dobivene pomoću AI tehnologije s automatizacijom i stručnošću svojih zaposlenika, navode dodatne korisne učinke korištenja AI na njihove rezultate u sigurnosnim operacijama. 67% navodi da sposobnost učinkovitije trijaže prijetnji razine 1 pomaže u uklanjanju troškova i vremena osnovne detekcije. 65% menadžera navodi da je smanjenje šuma i broja lažno pozitivnih rezultata smanjilo potrebu za ljudskom inspekcijom. Dodatnih 65% navodi da korištenje bihevioralne analitike dovodi do učinkovitijeg predviđanja budućih prijetnji, što je važan korak u tome da organizacije postanu proaktivnije (Muppidi et al., 2022).

Utjecaj korištenja automatizacije na detekciju, odgovor na prijetnje i vrijeme oporavka prikazan je na slici 8.



Slika 8. Utjecaj korištenja AI i automatizacije na detekciju, odgovor na prijetnje i vrijeme oporavka
Izvor: IBM Security Services, preuzeto iz: Parham, G. (2022), 4 Ways AI Capabilities Transform Security

Preopterećeni sigurnosni timovi više se ne mogu usredotočiti isključivo na sprječavanje infiltracije. Prevencija napada gotovo da postaje nemogući cilj, pa ju je potrebno uravnotežiti s otkrivanjem napada i njihovom sanacijom. Činjenica da su današnje organizacije izrazito ranjive na infiltraciju ne znači da nužno mora doći i do krađe podataka ili poslovne štete. Organizacije moraju proaktivno identificirati i neutralizirati prijetnje svome IT okruženju prije nego što napadači postignu svoje ciljeve. Korištenje sigurnosti vođene inteligencijom je strategija za rješavanje najozbiljnijih i najosjetljivijih sigurnosnih izazova današnjice (RSA, 2015).

O djelotvornosti obrambenih mehanizama temeljenih na AI mogli bismo kratko zaključiti- to je jednostavno pitanje učinkovitosti stroja u odnosu na učinkovitost ljudskih napora (Guembe et al., 2022).

No treba napomenuti da je korištenje AI u informacijskoj sigurnosti, u smislu automatizacije i pojednostavljenja većine procesa, kao i dokidanja potrebe za ručnom obradom velikih količina podataka, u poodmakloj fazi, dok je ono u smislu pomoći pri odlučivanju još u početnoj fazi. Tehnologije temeljene na AI na trenutnom stupnju razvoja mogu djelomično zamijeniti stručnjake u tipičnim poslovima održavanja informacijskih sustava, dok kod složenijih zadataka mogu

sudjelovati u početnoj obradi zahtjeva prije nego što ih proslijede stručnjacima. Zasada AI može pomoći sigurnosnim stručnjacima u donošenju odluka, no još ih ne može u tome poslu u cijelosti zamijeniti. Glavni problem povezan je s podjelom odgovornosti i pitanjem u kojoj je mjeri moguće delegirati kritične odluke sustavima umjetne inteligencije te tko će biti odgovoran za odluke koje donosi AI. Problem se može riješiti tako da AI sustavi uvijek rade u suradnji s ljudskim operaterom koji će donijeti najvažnije odluke te snositi odgovornost. U slučaju pune odgovornosti AI, postoje brojni pravni aspekti koje je potrebno riješiti, uključujući naknade štete od mogućih gubitaka zbog pogrešnih odluka (Erokhin, 2020).

5.3. Studije slučaja i primjeri

U ovome potpoglavlju ćemo detaljnije opisati 3 poznata komercijalna obrambena rješenja temeljena na AI te navesti financijske i nefinancijske koristi njihove implementacije.

Prvo područje primjene AI u obrambene svrhe koje ćemo potkrijepiti primjerom je sigurnost krajnjih točaka (engl. *Endpoint Security*). Krajnje točke su mjesta gdje su osjetljivi podaci tvrtke najosjetljiviji na napad. Zlonamjerni akteri mogu iskoristiti ranjivosti krajnjih točaka kako bi dobili pristup mreži, ukrali podatke ili proširili zlonamjerni softver, stoga je njihova sigurnost ključna za zaštitu podataka, aplikacija i sustava u cjelini (Goodman, 2023).

U sigurnosti krajnjih točaka, u posljednjih par godina dolazi do promjene pristupa-od stavljanja naglaska na prevenciju do shvaćanja da je vidljivost jednako važna kao i otkrivanje. U tzv. eri EDR-a (engl. *Extended Detection and Response*) u fokusu sigurnosnih stručnjaka su podaci krajnjih točaka i vidljivost (Goodman, 2023).

Kako bi upravljali ovim novim, golemim skupom podataka o krajnjim točkama, sigurnosni stručnjaci okreću se cloud tehnologijama za pohranjivanje i obradu podataka o krajnjim točkama, temeljem koji će identificirati anomalije i raskrinkati aktivne napadače, za koje se sada pretpostavlja da su unutar organizacije. Napadači su se pak prebacili sa pristupa napredne ustrajne prijetnje (engl. *Advanced Persistent Threat- APT*) na "razbij-i-zgrabi" pristup, pa nastoje unovčiti male nedostatke u obrani poduzeća, uglavnom preko ransomware-a. U fokus ponovo dolazi prevencija (Goodman, 2023).

CylanceENDPOINT, proizvođača BlackBerry, je pomoću AI vođen obrambeni mehanizam odnosno platforma za zaštitu krajnjih točaka (engl. *Endpoint Protection Platform*), za koju neovisne analize (Forrester Consulting, 2022) pokazuju da zaustavlja više napada od drugih EPP-ova. To joj uspjeva zahvaljujući sofisticiranim algoritmima koji omogućuju otkrivanje i sprječavanje prijetnji prije nego što se imaju priliku u potpunosti izvršiti, dakle znatno ranije u lancu napada (Forrester Consulting, 2022).

Jedna od ključnih prednosti primjene ML u zaštiti krajnjih točaka jest činjenica da je do 95% "lakši" (računalno manje zahtjevan) od naslijeđenih rješenja koja se oslanjaju na računalno zahtjevne potpise. Zato se može implementirati na širok raspon sustava bez dodavanja značajnih troškova. To ga čini idealnim izborom za organizacije koje žele zaštititi svoju imovinu bez ugrožavanja

performansi sustava (Goodman, 2023). Rješenje pruža bolju vidljivost napada, od prvih znakova upozorenja korištenjem vlastitih podataka o prijetnjama, preko "lanca ubijanja", do sprječavanja stvarnog napada. "Lanac ubijanja" je niz koraka koji prate faze kibernetičkog napada od ranih faza izviđanja do eksfiltracije podataka. Kroz svaki sigurnosni događaj, CylanceENDPOINT konsolidira i povezuje različita upozorenja kako bi stvorio potpunu shemu napada. Pružajući manji broj korisnijih upozorenja podržava organizacijsko odlučivanje i sigurnosne odgovore (Goodman, 2023).

Zaštita krajnjih točaka temeljena na AI pruža superiorno otkrivanje u usporedbi s rješenjima temeljenim na potpisu. Strojevi koji nisu povezani na internet ne moraju povlačiti bilo kakva ažuriranja, što se kod svih antivirusnih programa pokazalo kao izvor rizika i greški pri izvođenju (Forrester Consulting, 2022).

Dodatne prednosti uključuju: sposobnost učinkovitog upravljanja sigurnošću u sve većem broju uređaja i operativnih sustava, uključujući osobne uređaje i USB; podrška naslijeđenim sustavima omogućuje organizacijama da pojednostave svoja sigurnosna okruženja, eliminirajući alate koji se preklapaju; povećanje zadovoljstva zaposlenika i svijesti o kibernetičkoj sigurnosti, uz popratni program obuke (Forrester Consulting, 2022).

Za kompozitnu organizaciju modeliranu u studiji, Forrester je zaključio da je sigurnosno rješenje za krajnje točke temeljeno na umjetnoj inteligenciji isplatilo početno ulaganje za manje od šest mjeseci te pružilo druge financijske prednosti tijekom trogodišnjeg razdoblja: više od 1,2 milijuna dolara smanjenja ukupnih troškova sigurnosnih upada; smanjenje troškova pretplate za više od 800.000 dolara za naslijeđeni softver koji je njime zamijenjen; ušteda više od 8000 radnih sati osoblja zbog nepostojanja potrebe tradicionalnih antivirusnih ažuriranja; ušteda od gotovo 2000 radnih sati osoblja za istraživanja sigurnosnih incidenata i ponovne postavke zaraženih uređaja (Forrester Consulting, 2022).

Sljedeće područje primjene AI u obrambene svrhe koje ćemo potkrijepiti primjerom jest napredna analitika/inteligencija prijetnji odnosno kognitivna sigurnost (engl. *Cognitive Security*).

Kako kibernetički napadi rastu u opsegu i složenosti, AI pomaže analitičarima sigurnosnih operacija da se nose s nedostatkom resursa, a istovremeno budu ispred prijetnji. IBM Security nudi SIEM (engl. *Security Information and Event Management*) platformu pod nazivom QRadar SIEM te QRadar Advisor with Watson, koji sigurnosnim analitičarima na raspolaganje stavlja mogućnosti IBM Watsona za automatiziciju rutinskih zadataka sigurnosnog operativnog centra (engl. *Security Operations Center-SOC*). Ova rješenja za cilj imaju poboljšati učinkovitost prioritizacije prijetnji, istrage i eventualne eskalacije (Forrester Consulting, 2019).

IBM Watson iskorištava prednosti različitih oblika AI, uključujući algoritme strojnog učenja i neuronske mreže dubokog učenja, dok IBM Security QRadar Advisor s Watsonom pomaže u procjeni incidenata kako bi organizacije smanjile kibernetičke rizike. Naime, nedostatak dosljednih, visokokvalitetnih i kontekstom bogatih istraga može dovesti do velike vjerojatnosti propuštanja ključnih uvida, što organizacije izlaže rizicima. Dokazano je da analitičari mogu pratiti samo osam posto informacija potrebnih za obavljanje svog posla. AI može automatski pronaći

uzorke u incidentima korištenjem kognitivnog zaključivanja i pružiti djelotvorne povratne informacije obogaćene kontekstom (IBM Security, 2022).

Zahvaljujući ovim rješenjima organizacije mogu usmjeriti napore svojih sigurnosnih timova na proaktivne sigurnosne zadatke, koji poboljšavaju razinu organizacijske sigurnosti, dok Advisor rješava ponavljajuće prijetnje te zadatke njihove istrage i prioretizacije. Osnovna prednost implementacije je upravo poboljšana produktivnost sigurnosnih timova i njihovih analitičara (Forrester Consulting, 2019). Na taj način kognitivna sigurnost kombinira prednosti umjetne inteligencije i ljudske inteligencije.

Prije implementacije AI rješenja, analitičari SOC-a provodili su 65% svog radnog dana istražujući prijetnje,od kojih su većina bile prijetnje prve razine. Za svaku istragu prijetnje analitičarima je u prosjeku trebalo 4 sata, pri čemu su neke istrage trajale nekoliko tjedana. Dugi ciklusi istrage odgodili su napore za otklanjanje prijetnji, riskirajući rastuća vremena zadržavanja prije eskalacije. Nakon implementacije QRadar Advisora s Watsonom, prosjek trajanja istrage smanjio se na manje od 10 minuta odnosno 15% radnog vremena (Forrester Consulting, 2019). Jedna od najznačajnijih prednosti bila je fleksibilnost uključivanja približno 38% jeftinijeg osoblja NOC-a (engl. *Network Operations Center*) za pomoć u istrazi prijetnji prve razine. Povećane sposobnosti mrežaša omogućuju organizacijama da izbjegnu zapošljavanje dodatnog sigurnosnog osoblja, kao i da analitičare s formalnom sigurnosnom obukom oslobode za fokusiranje na prijetnje više razine i proaktivne sigurnosne mjere. Bez implementacije ovih rješenja, potencijalno kritične sigurnosne prijetnje ostale bi neistražene i neriješene, što je za organizacije velik rizik. Uz implementaciju, kapacitet organizacije za istraživanje prijetnji znatno se povećao (sa 1800 godišnje na više od 7000 prijetnji u istom vremenskom rasponu). Rizik organizacije od značajne povrede sigurnosti podataka koja proizlazi iz neistraženih prijetnji smanjena je za 8% (Forrester Consulting, 2019).

Intervjuirana organizacija ostvarila je sljedeće kvantificirane koristi: ušteda produktivnosti SOC analitičara od 1,8 milijuna USD., izbjegnute naknade za outsourcing za istrage od 126.829 USD, poboljšana organizacijska sigurnost za 651.936 dolara (Forrester Consulting, 2019).

Nekvantificirana korist je činjenica da je osoblje NOC-a počelo graditi vještine prepoznavanja prijetnji, čime postiže razvoj karijere, dok je osoblje SOC-a postalo vještije kod istraga prijetnji na višoj razini (Forrester Consulting, 2019).

Preostalo područje primjene AI u obrambene svrhe koje ćemo potkrijepiti primjerom jest sigurnost mreže temeljena na namjeri (engl. *Intent Based Networking- IBN*).

Kada moraju pustiti u rad novu aplikaciju, napraviti migraciju nekog servisa ili učiniti bilo koju drugu promjenu, mrežni timovi moraju paziti da ne ometaju veze koje podržavaju aplikacije i servise organizacije te da pritome ne stvore sigurnosne propuste ili kršenja usklađenosti (Wool, 2019). Cloud Security Alliance (CSA), vodeća svjetska organizacija posvećena definiranju i podizanju svijesti o najboljim praksama za sigurno okruženje računalstva u oblaku, i AlgoSec, vodeći pružatelj poslovnih rješenja za upravljanje mrežom i sigurnošću u oblaku, objavili su rezultate studije pod nazivom "Složenost sigurnosti u oblaku: izazovi u upravljanju sigurnošću u izvornim, hibridnim i multi-cloud okruženjima". Anketirano je 700 IT i sigurnosnih stručnjaka te je zaključeno da je više od 42% organizacija doživjelo prekide rada aplikacija ili mreže uzrokovane

jednostavnim ljudskim pogreškama ili pogrešnim konfiguracijama (Wool, 2019). S obzirom na složenost današnjih mreža, to nas ne treba čuditi. Upravljanje mrežom i procesi automatizacije poduzeća moraju uzeti u obzir sve sigurnosne uređaje i pravila (bilo u podatkovnom centru, na njegovom perimetru, unutar *on-premise* mreža ili u oblaku) kako bi se omogućilo istinska agilnost bez ugrožavanja zaštite. Usto, sigurnosne politike većih organizacija mogu uključivati tisuće ili čak milijune pravila postavljenih na njihovim vatrozidima i usmjerivačima. IBN ima potencijal osigurati takvim organizacijama brzu prilagodbu mreže promjenjivim potrebama poslovanja te povećanu agilnost bez stvaranja rizika (Wool, 2019).

Glavno načelo IBN-a jest da mreža preuzima poslovnu namjeru i automatski je pretvara u mrežne konfiguracije za sve uređaje, svodeći ručni rad na minimum. Obzirom kontinuirano prati i prilagođava konfiguracije, osigurava potrebnu brzinu i agilnost uz zajamčenu usklađenost i smanjene rizike (Wool, 2019).

Prema predsjedniku Gartner Research, Andrewu Lerneru, svaki mrežni sustav temeljen na namjeri (engl. *Intent Based Networking System-IBNS*) trebao bi imati četiri komponente. Prva komponenta je prijevod i provjera valjanosti. Poslovnu politiku visoke razine, definiranu od strane upravitelja mreže, IBNS prevodi u radnje koje softver izvodi i provjerava da se politika može izvršiti. Druga komponenta je automatizirana implementacija, kada IBNS softver manipulira mrežnim resursima kako bi stvorio željeno stanje i nametnuo politike definirane od upravitelja mreže. Treća komponenta je svijest o stanju, kroz prikupljanje podataka za stalno praćenje stanja mreže. Četvrta komponenta je osiguravanje i dinamička optimizacija/sanacija odnosno održavanje željenog stanja mreže. IBNS koristi strojno učenje za odabir najboljeg načina za implementaciju željenog stanja i po potrebi može poduzeti automatizirane korektivne radnje za njegovo održavanje (Lerner, 2017).

Cisco Software Defined Access (Cisco SD-Access) prvo je IBNS rješenje za poduzeća na tržištu. Izgrađen je na principima Ciscove digitalne mrežne arhitekture (Cisco DNA) (Hill et al., 2019).

Cisco SD-Access je programabilna mrežna arhitektura koja pruža politike/pravila temeljenu na softveru i segmentaciju od ruba mreže do aplikacija. SD-Access implementiran je putem Cisco Digital Network Architecture Center (Cisco DNA Center) koji pruža mogućnosti postavki dizajna, definiranja pravila i automatiziranog pružanja mrežnih elemenata, kao i analitiku usklađenosti sa sigurnosnim pravilima za inteligentne žične i bežične mreže. Arhitekturi poduzeća koja uključuje mnoštvo lokacija, uređaja, servisa i sigurnosnih pravila, ovo rješenje nudi sveobuhvatnu arhitekturu mreže dosljednu u smislu povezanosti, segmentacije i usklađenosti politika/pravila (Hill et al., 2019) .

Sastoji se od dva glavna sloja: SD-Access Fabric sadrži fizičku i logičku infrastrukturu za prosljeđivanje mreže, dok Cisco DNA Center sadrži sve komponente potrebne za automatizacije, politike/pravila, osiguranje njihove usklađenosti i integracije (Hill et al., 2019).

SD-Access Fabric podsjeća na dvoslojnu tkaninu, gdje je jedan sloj (poznat kao podloga) zadužen za fizičke uređaje i prosljeđivanje prometa, a drugi potpuno virtualni sloj (poznat kao nadsloj) je mjesto gdje su žičani i bežični korisnici i uređaji logično povezani te gdje se primjenjuju i izvršavaju usluge/servisi i pravila. Sloj podloge sačinjen je od fizičkih mrežnih uređaja, kao što su usmjerivači, preklopnici i bežični LAN kontroleri (WLC), uz dodatak tradicionalnog Layer 3

usmjerivačkog protokola. Radi se o jednostavnoj, skalabilnoj i otpornoj osnovi za komunikaciju između mrežnih uređaja. Nadsloj je logična, virtualizirana topologija izgrađena na opisanoj fizičkoj podlozi. Ovakva arhitektura omogućuje optimiziranu konvergenciju na više staza te pojednostavljuje implementaciju, rješavanje problema i upravljanje mrežom (Hill et al., 2019).

Cisco DNA Center je centralizirana operativna platforma za *end-to-end* automatizaciju i osiguranje usklađenosti LAN, WLAN i WAN okruženja organizacije, kao i orkestraciju s vanjskim rješenjima i domenama. Omogućuje mrežnom administratoru korištenje jedne nadzorne ploča za upravljanje i automatizaciju mreže. Odgovoran je za konfiguraciju prosljeđivanja prometa i distribuciju pravila, kao i upravljanje uređajima i analitiku, a dvije su mu glavne funkcije automatizacija i osiguranje usklađenosti (Hill et al., 2019).

Automatizacija Cisco DNA centra pruža definiciju i upravljanje politikama temeljenim na grupama, zajedno s automatizacijom svih konfiguracija povezanih s pravilima. Cisco DNA Center koristi automatizaciju temeljenu na kontroleru kao primarni model konfiguracije i orkestracije, za dizajn, implementaciju, provjeru i optimizaciju žičanih i bežičnih mrežnih komponenti. Uz njegovo potpuno upravljanje infrastrukturom, IT timovi ne moraju brinuti o detaljima implementacije, što rezultira pojednostavljivanjem operacija, minimiziranjem šanse za ljudske pogreške i standardizacijom cjelokupne mreže (Hill et al., 2019).

Osim općeg upravljanja mrežom, mrežno osiguranje usklađenosti kvantificira rizike iz perspektive mrežnih timova te mjeri utjecaj promjene mreže na sigurnost, dostupnost i usklađenost. Ključni čimbenik Cisco DNA Assurance-a je analitika: sposobnost kontinuiranog prikupljanja podataka s mreže i njihove transformacije u korisne uvide po kojima je moguće djelovati. Prikuplja različite mrežne telemetrije, u tradicionalnim oblicima (npr. *netflow, syslogs* itd.), ali i novim oblicima (npr. streaming telemetrija). Zatim provodi njihovu naprednu obradu kako bi procijenio i povezao događaje te kontinuirano pratio rad uređaja, korisnika i aplikacija. Drugim rješenjima za upravljanje mrežom često nedostaje ova razina korelacije i posljedična vidljivost problema u podsloju mreže koji mogu utjecati na izvedbu nadsloja (Hill et al., 2019).

Među benefitima Cisco SD-Access potrebno je još spomenuti one vezane uz upravljanje sigurnosnim politikama/pravilima (engl. *Policy management).* Osnovna unaprjeđenja u ovome području su: odvajanje politika od dizajna infrastrukture, pojednostavljena definicija politika, automatizacija politika te orkestracija poduzeća temeljena na politikama. Odvajanje politika od mrežne topologije omogućuje učinkovitije operacije njihova provođenja, besprijekornu mrežnu mobilnost, smanjenje napora u svakodnevnoj administraciji mreže, ali i poslovne benefite poput omogućavanja novih poslovnih usluga u kraćem vremenu (Hill et al., 2019).

Obzirom tehnologija nije sasvim "zrela", nema drugih značajnih komercijalnih rješenja s kojima bi ovo rješenje moglo biti uspoređeno niti dostupnih studija o financijskim učincima njegove implementacije. Proizvođač kao osnovne benefite ističe: automatizirani *deployment* na veliki broj lokacija i uređaja, integriranu žičnu i bežičnu infrastrukturu, omogućavanje sigurnog pristupa korisnicima i uređajima te povezane uvide i analitiku. Za organizacije to znači povećanu agilnost i efikasnost te smanjene rizike (Hill et al., 2019).

Uz tri detaljno opisana primjera obrambenih rješenja temeljenih na AI, u nastavku ćemo ukratko navesti još neka komercijalna AI obrambena rješenja, spomenuta u analiziranoj literaturi.

Apple koristi AI za poboljšanje kvalitete i eliminaciju nedostataka biometrijske autentifikacije (npr. nepouzdanost zbog promjene izgleda). Apple tehnologija obrađuje crte lica korisnika pomoću ugrađenih infracrvenih senzora i mehanizama neuronske mreže. Softver stvara složeni model korisnikova lica, identificiranjem glavnih korelacija i uzoraka, uspješno radi i uz različite uvjete osvjetljenja te kompenzira promjene poput promjene frizure, puštanja brkova ili brade, nošenja šešira itd. Apple tvrdi da je vjerojatnost da napadač prevari AI i dobije pristup uređaju čiji nije vlasnik jedan naprema milijun (Erokhin, 2020).

Dataguiseov DgSecure Monitor sustav dizajniran je za otkrivanje curenja podataka. Koristeći ML i analize ponašanja, sustav generira upozorenja kada radnje korisnika odstupaju od tipičnih. Na taj način DgSecure Monitor olakšava i razvoj politike upravljanja podacima koje zahtijevaju različite razine zaštite (Erokhin, 2020).

Distil Networks nudi tehnologiju koja štiti web aplikacije od zlonamjernih botova i napada temeljenih na zlouporabi programskih sučelja. Distilovi klijenti dobivaju pristup globalnoj platformi koja pomoću strojnog učenja analizira obrasce napada u stvarnom vremenu. Distil prepoznaje botove povezivanjem velikog broja promjenjivih parametara i otkrivanjem anomalija u ponašanju na temelju obrasca prometa specifičnih za određeno web mjesto (Erokhin, 2020).

Qrator Labos nudi Anti-DDoS rješenj koje koristi metode strojnog učenja za analiziranje ponašanja posjetitelja zaštićenog mjesta. Temeljem analiza, gradi se model koji predviđa moguće postupke legitimnog korisnik. Ovaj model omogućuje filtriranje prometa DDoS napada s visokom točnošću i relativno niskim postotkom lažno pozitivnih rezultata. Ova metoda je posebno učinkovita za tzv. *Full Browser Stack* napade, koje je najteže filtrirati (Erokhin, 2020).

Specijalizirana platforma za otkrivanje ciljanih napada, uključujući kampanje cyber špijunaže, Kaspersky Anti Targeted Attack Platform, otkriva ciljane napade i sve sumnje na zlonamjerne aktivnosti u mreži organizacije, čak i prije nego što napadači poduzmu bilo kakve ozbiljne korake. Rješenje ih detektira zahvaljujući skupu senzora koji neprestano prate situaciju unutar zaštićene IT infrastrukture, potom analizira podatke primljene iz različitih čvorova mreže. Kako bi se procijenilo koliko je sumnjiva aktivnost opasna, prikupljeni podaci se preusmjeravaju u izolirano virtualno okruženje, gdje se proučava ponašanje potencijalno zlonamjernih objekata. Konačna odluka o tome radi li se o ciljanom napadu donosi se korištenjem posebnog analitičkog alata koji koristi AI tehnologije i sposoban je usporediti različite podatke (Erokhin, 2020).

6. ZAKLJUČAK

Nedvojbeno je da će primjena AI transformirati organizacije na kvalitativno različite načine od drugih tehnologija, stoga je ključno razviti sposobnosti organizacija da se suoče s ovim izazovima (tj. njihovu spremnost za AI). Njoj je potrebno pristupiti holistički. Nije dovoljno promatrati samo korištenje AI tehnologija, već i koliki broj aktivnosti ključnih za stvaranje vrijednosti one podupiru, u kojoj mjeri utječu na širenje granica organizacije i njezine odnose sa okruženjem te u kojoj mjeri pridonose identitetu organizacije i ostvarenju njezinih strateških ciljeva.

Stvaranje i održavanje vrijednosti, odnosi sa okruženjem i ostvarenje strateških ciljeva u ovisnosti su o informacijskoj sigurnosti odnosno sigurnosti podataka organizacije i njezinih korisnika.

Predmet ovoga rada je razmatranje dvojake uloge AI u informacijskoj sigurnosti. Radom se nastojala dokazati teza da je AI tehnologija čije karakteristike podjednako predstavljaju izvor prijetnji, kao i odgovor na prijetnje u dinamičnome poslovnom okruženju.

U nastavku su navedeni ciljevi rada, način na koji su ciljevi rada ostvareni te odgovori na glavna istraživačka pitanja koja proizlaze iz ciljeva.

C1. Identificirati karakteristike umjetne inteligencije koje je čine značajnim izvorom prijetnje i uspješnim odgovorom na prijetnju informacijskoj sigurnosti

Analizom dostupne znanstvene i stručne literature zaključeno je da svi kibernetički napadi koji koriste AI dijele specifične karakteristike, koje ih čine sofisticiranijima i težima za otkrivanje. Te karakteristike su: prilagodljiva i evoluirajuća priroda, automatizacija i skalabilnost, poboljšana manipulacija podacima, nedostatak objašnjivosti i interpretabilnosti, pristupačnost i demokratizacija, nedostatak etičkog kodeksa.

Analizom dostupne znanstvene i stručne literature zaključeno je da obrambeni mehanizmi koji su temeljeni na umjetnoj inteligenciji dijele specifične karakteristike, koje ih čine sofisticiranijima i učinkovitijima od tradicionalnih obrambenih mehanizama. Te karakteristike su: poboljšana manipulacija podacima, poboljšana detekcija i klasifikacija anomalija, proaktivnost, automatizacija i autonomnost, prilagodljiva i evoluirajuća priroda, fleksibilnost i intregracija.

Uočena su preklapanja između karakteristika koje predstavljaju izvor prijetnji i karakteristika koje čine AI učinkovitim odgovorom na prijetnje. Iste karakteristike koje AI čine potencijalnom prijetnjom mogu se iskoristiti u obrambene svrhe- obrambenim mehanizmima temeljenima na AI organizacije se uspješno brane od napada razvijenih korištenjem AI. Sve veća automatizacija napada i obrane dovodi do stalne "utrke u naoružanju" na polju kibernetičke sigurnosti. Upravo stoga je postizanje ravnoteže između iskorištavanja sposobnosti AI za obranu i upravljanja rizicima povezanim sa zlonamjernom upotrebom AI stalni izazov za sigurnosne stručnjake u organizacijama i istraživače AI.

C2. Ispitati mogućnosti umjetne inteligencije da donese unaprjeđenja u točnosti pri detekciji prijetnji, skraćenju vremena istraživanja prijetnji, automatizaciji odgovora i implementaciji proaktivnih mehanizama zaštite

Korištenjem naprednih algoritama strojnog učenja, kao što su neuronske mreže ili gensko programiranje te kombinirajući ih s analizom anomalija i heurističkim analizama, moguće je identificirati prijetnje koje se ne bi mogle prepoznati konvencionalnim sigurnosnim sustavima. Bihevioralna analitika omogućuje prevenciju štete od napada koji ne bi bili primijećeni standardnim alatima za upravljanje prijetnjama, uključujući one temeljene na zlouporabi legitimnih validacijskih podataka. AI mehanizmi mogu identificirati čak i napredne prijetnje i zlonamjerni softver koji koriste sofisticirane tehnike skrivanja ili promjene svojih oblika.

AI omogućava bržu analizu incidenata i na taj način bolje razumijevanje onoga što se događa u mreži organizacije, omogućava točnije predviđanje ozbiljnih curenja podataka, brže otkrivanje incidenata i reakciju na njih, kako bi se smanjila moguća šteta. Brzina odgovora izravno ovisi o razini automatizacije koju pružaju AI i strojno učenje.

Zaštita krajnjih točaka temeljena na AI pruža superiorno otkrivanje u usporedbi s rješenjima temeljenim na potpisu. Poboljšane analitičke sposobnosti i napredna inteligencija prijetnji mogu unaprijediti upravljanje informacijskom sigurnošću i na druge načine. To uključuje generiranje izvješća o sigurnosnim događajima, statističku analizu i vizualizaciju podataka te pružanje preporuka za jačanje sigurnosti.

U odnosu na tradicionalne obrambene mehanizme, AI ima pozitivan utjecaj na sposobnosti organizacija da djeluju proaktivno u slučaju neočekivanih sigurnosnih događaja. Platforme za obavještavanje o prijetnjama koje pokreće AI mogu pomoći sigurnosnim timovima da budu informirani o najnovijim prijetnjama, proaktivno poboljšaju svoju obranu i ostanu korak ispred napadača.

C3. Ocijeniti mogućnosti umjetne inteligencije da se nosi s izazovima s kojima su suočeni stručnjaci iz područja informacijske sigurnosti (previše zadataka i podataka, ograničeno vrijeme i vještine)

Obzirom na promjenjivu i evoluirajuću prirodu kibernetičkih napada, čak i kada organizacija raspolaže velikim brojem stručnjaka, njima treba pomoć da se nose s velikim mrežnim prometom, novim načinima kršenja privatnosti i poboljšanim vektorima napada koji postaju teški za savladavanje od strane ljudi. Organizacije moraju poduzeti korake kako bi povećale svoju učinkovitost te učinile više s manje resursa u sve kompliciranijem okruženju prijetnji.

Svaki zadatak koji zahtijeva stručnu intervenciju može se modelirati korištenjem AI tehnika, ako se značajke povezane sa zadatkom mogu identificirati i mogu se prikupiti podaci koji predstavljaju te značajke.

Uporaba AI olakšava otkrivanje zlonamjernog softvera korištenjem podataka iz prethodnih napada i raznih metoda, uključujući analizu ponašanja, procjenu rizika, blokiranje botova, zaštitu krajnjih točaka i automatizaciju sigurnosnih zadataka. Osnovna prednost implementacije je upravo poboljšana produktivnost sigurnosnih timova i njihovih analitičara. Implementacijom AI rješenja stručnjaci mogu napustiti zamorne i ponavljajuće ručne provjere te se početi baviti stručnijim poslom nadzora i upravljanja AI rješenjima te kontrolom kvalitete njihove izvedbe. AI i automatizacija mogu imati dramatičan pozitivan utjecaj na sposobnost nošenja s volumenom i

tempom sigurnosnih događaja, što je ključan čimbenik poboljšanja radnog okruženja sigurnosnih analitičara. Konačni ishod su dobitak u kapacitetu i specijalizacija radne snage IT sigurnosti.

Ne treba zanemariti ni sposobnost AI da pomogne u stvaranju i rastu organizacijskog znanja o rizicima informacijske sigurnosti. Znanje već dostupno u organizaciji, AI može vremenski i troškovno efikasnije uvećavati i stavljati na raspolaganje stručnjacima. To je važno jer, kako se tehnologija razvija, nove sigurnosne prijetnje pojavljuju se i kao rezultat neažurnosti znanja i informacija profesionalaca u kibernetičkoj sigurnosti.

Zasada AI može pomoći sigurnosnim stručnjacima u donošenju odluka, no još ih ne može u tome poslu u cijelosti zamijeniti. Postoje brojne odluke koje još uvijek može donijeti samo čovjek, a spomenuta ravnoteža između rizika i koristi je možda i najvažnija od njih. Glavni problem povezan je s podjelom odgovornosti i pitanjem u kojoj je mjeri moguće delegirati kritične odluke sustavima umjetne inteligencije te tko će biti odgovoran za odluke koje donosi AI. Problem se može riješiti tako da AI sustavi uvijek rade u suradnji s ljudskim operaterom koji će donijeti najvažnije odluke te snositi odgovornost. U slučaju pune odgovornosti AI, postoje brojni pravni aspekti koje je potrebno riješiti, uključujući naknade štete od mogućih gubitaka zbog pogrešnih odluka.

C4. Donijeti zaključke o opravdanosti ulaganja u zaštitne mehanizme temeljene na umjetnoj inteligenciji te rizicima koji mogu proisteći iz neiskorištavanja potencijala umjetne inteligencije u ovome području.

Kao rezultat analize komercijalnih AI rješenja i njihova ekonomskog učinka, u poglavlju 5 izneseni su kvantitativni podaci o opravdanosti ulaganja u zaštitne mehanizme temeljene na AI. Dokazano je da ulaganja utječu na smanjenje trajanja istrage prijetnji, skraćenje vremena do eskalacije i otklanjanja prijetnji te povećanje kapaciteta (godišnjeg broja istraga). Organizacijama daju nekvantificirane koristi poput fleksibilnosti uključivanja jeftinijeg IT osoblja u istrage prijetnji niske razine te im omogućuju da izbjegnu zapošljavanje dodatnog sigurnosnog osoblja, a analitičare s formalnom sigurnosnom obukom oslobode za fokusiranje na prijetnje više razine i proaktivne sigurnosne mjere. Dodatna prednost korištenja strojnog učenja u informacijskoj sigurnosti je da, uvodeći i trenirajući ga, organizacija zauzima prediktivni pristup sigurnosti. Intervjuirane organizacija ostvarile su i znatne kvantificirane koristi u obliku ušteda proizašlih iz povećane produktivnosti, izbjegnutih naknada za outsourcing i ažuriranja, u obliku poboljšane razine organizacijske sigurnosti i smanjenog rizika od značajne povrede sigurnosti podataka.

Bez implementacije ovih rješenja, potencijalno kritične sigurnosne prijetnje ostale bi neistražene i neriješene, što je za organizacije velik rizik od ozbiljnih poremećaja poslovanja i gubitka prihoda.

Zaključimo, korištenje umjetne inteligencije u kibernetičkoj obrani ima brojne prednosti, ali nije rješenje koje odgovara svim organizacijama. AI ima potencijal biti od koristi za kibernetičku sigurnost, kao i potencijal da joj naškodi. AI se koristi i kao mač (za podršku zlonamjernim napadima) i kao štit (za suzbijanje sigurnosnih rizika). U kojoj mjeri AI unaprjeđuje informacijsku sigurnost ovisi o tome koliko efikasno organizacije upravljaju svojim AI rješenjima.

Ako se u prevenciji sigurnosnih incidenata ponavljaju ljudske greške ili se množe aktivnosti održavanja, nedostatak odluke da se implementiraju sigurnosna rješenja temeljena na AI može se

smatrati nemarom, posebno u organizacijama koje raspolažu velikim količinama osjetljivih podataka pojedinaca ili su dio ključne infrastrukture.

POPIS IZVORA

1. Abeshu, A. i Chilamkurti, N. (2018), Deep learning: the frontier for distributed attack detection in fog-to-things computing, *IEEE Communications Magazine*, 56(2), 169-175. doi: 10.1109/MCOM.2018.1700332.
2. Afsah, E. (2022). Artificial Intelligence, Law, and National Security, u: Voeneky, S., Kellmeyer, P., Mueller, O i Burgard, W. (ur.), *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives,* (str. 445-446), Cambridge: Cambridge University Press
3. Agarwal, S. i Farid, H. (2019), Protecting World Leaders Against Deep Fakes, preuzeto 13. lipnja 2023. s https://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf
4. Agrawal, A., Gans, J. i Goldfarb, A, (2022, 2.prosinac), ChatGPT and How AI Disrupts Industries, *Harvard Business Review,* preuzeto s: https://hbr.org/2022/12/chatgpt-and-how-ai-disrupts-industries
5. Alawadhi, S. A., Zowayed, A., Abdulla, H., Khder, M. A. i Ali, B. J. A. (2022), Impact of Artificial Intelligence on Information Security in Business, u: *ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, (str. 437-442), Manama: IEEE
6. Aljindi, A. (2015), Information security, artificial intelligence and legacy information systems, doktorski rad, Northcentral University, Graduate Faculty of the School of Business and Technology Management, Prescott Valey, Arizona , preuzeto 26. veljače 2023. s: https://www.proquest.com/dissertations-theses/information-security-artificial-intelligence/docview/1752253295/se-2

7. Bird. E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E. i Winfield, A. (2020), *The ethics of artificial intelligence: Issues and initiatives* [e-publikacija], preuzeto s https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf
8. Bostrom, N. (2017), *Superintelligence: Paths, dangers, strategies*, Oxford: Oxford University Press
9. Brooks, T.N. (2019), Survey of Automated Vulnerability Detection and Exploit Generation Techniques in Cyber Reasoning Systems, u: Arai, K., Kapoor, S. i Bhatia, R. (ur.), *Intelligent Computing. SAI 2018. Advances in Intelligent Systems and Computing* (str. 1083–1102.), Cham:Springer
10. Burgard, W. (2022). Artificial Intelligence: Key Technologies and Opportunities, u: Voeneky, S., Kellmeyer, P., Mueller, O i Burgard, W. (ur.), *The Cambridge Handbook of Responsible Artificial Intelligence: Interdisciplinary Perspectives* (str. 11-18), Cambridge: Cambridge University Press
11. Burström, T., Parida, V. , Lahti, T. i Wincent, J. (2021), AI-enabled business-model innovation and transformation in industrial ecosystems: A framework, model and outline for further research, *Journal of Business Research,* 127, 85-95. https://doi.org/10.1016/j.jbusres.2021.01.016
12. Celona (2021), What Is a Self-Organizing Network? SON Overview & Explainer, preuzeto 17. svibnja 2023. s https://www.celona.io/network-architecture/self-organizing-network
13. Dhanrajani, S. (2020, 14. siječanj.), AI-Driven Disruption And Transformation: New Business Segments To Novel Market Opportunities, *Forbes*, preuzeto s: https://www.forbes.com/sites/cognitiveworld/2020/01/14/ai-driven-disruption-and-transformation-new-business-segments-to-novel-market-opportunities/?sh=1c407116e841
14. Dixon, W. i Eagan, N. (2019), 3 ways AI will change the nature of cyber attacks, preuzeto 9. lipnja 2023. s https://www.weforum.org/agenda/2019/06/ai-is-powering-a-new-generation-of-cyberattack-its-also-our-best-defence/
15. Erdődi, L., Sommervoll, A.A. i Zennaro, F.M. (2021), Simulating SQL injection vulnerability exploitation using Q-learning reinforcement learning agents, *Journal of Information Security and Applications*, 61, 1-10. https://doi.org/10.1016/j.jisa.2021.102903.

16. Erokhin, S.D. (2020), Artificial Intelligence for Information Security, u: *2020 Systems of Signals Generating and Processing in the Field of on Board Communications*, (str. 1-4), Moskva: IEEE
17. Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C.,…Song, D. (2018), Robust Physical-World Attacks on Deep Learning Visual Classification, u: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (str. 1625-1634.), Salt Lake City, UT: IEEE
18. Fares, O.H., Butt, I. i Lee, S.H.M. (2022), Utilization of artificial intelligence in the banking sector: a systematic literature review, *Journal of Financial Services Marketing*, https://doi.org/10.1057/s41264-022-00176-7
19. Forrester Consulting (2019), *The Total Economic Impact™ Of IBM QRadar Advisor With Watson: Cost Savings And Business Benefits Enabled By AI For The Enterprise Security Team* [e-publikacija], preuzeto s https://www.ibm.com/downloads/cas/YEZ6W5EL
20. Forrester Consulting (2022), *The Total Economic Impact of CylancePROTECT from BlackBerry* [e-publikacija], preuzeto s https://www.blackberry.com/us/en/pdfviewer?file=/content/dam/resources/blackberry-com/resource-library/en/cyber/2022/standard/rp/rp-forrester-total-economic-impact-study-of-cylance-protect.pdf
21. Goodman, J.S. (2023), Endpoint Security Evolution: Protection and the Rise of Prevention, preuzeto 13. lipnja 2023. s https://blogs.blackberry.com/en/2023/04/endpoint-security-evolution-protection-and-prevention
22. Gregory, J. (2021), AI Security Threats: The Real Risk Behind Science Fiction Scenarios, preuzeto 03. ožujka 2023. s https://securityintelligence.com/articles/ai-security-threats-risk/
23. Guembe, B., Azeta, A., Misra, S., Osamor, V.C., Fernandez-Sanz L. i Pospelova, V. (2022), The Emerging Threat of Ai-driven Cyber Attacks: A Review, *Applied Artificial Intelligence*, 36 (1), 2376-2409, doi: 10.1080/08839514.2022.2037254
24. Haleem, A., Javaid, M., Qadri, M.A., Singh, R.P. i Suman, R. (2022), Artificial intelligence (AI) applications for marketing: A literature-based study, *International Journal of Intelligent Networks*, 3, 119-132. https://doi.org/10.1016/j.ijin.2022.08.005
25. Hill, C., Miller, D., Zacks, D., Suhr, J., Thatikonda, K.K., Karmarkar, K.…Pendharkar, V. (2019), *Cisco Software-Defined Access Enabling intent-based networking 2nd*

edition,preuzeto s https://www.cisco.com/c/dam/en/us/products/se/2018/1/Collateral/nb-06-software-defined-access-ebook-en.pdf

26. Hintze, A. (2016), Understanding the four types of AI, from reactive robots to self-aware beings, preuzeto 28. svibnja 2023. s https://theconversation.com/understanding-the-four-types-of-ai-from-reactive-robots-to-self-aware-beings-67616
27. Holmström, J. (2022), From AI to digital transformation: The AI readiness framework, *Business Horizons*, 65(3), 329-339. https://doi.org/10.1016/j.bushor.2021.03.006
28. Hu, W. i Tan, Y. (2022), Generating Adversarial Malware Examples for Black-Box Attacks Based on GAN, u: Tan, Y. i Shi, Y. (ur.), *Data Mining and Big Data. DMBD 2022. Communications in Computer and Information Science*, (str. 409-423.), Singapur: Springer
29. IBM (2022), *The 2022 Cost of a Data Breach report* [e-publikacija], preuzeto s https://www.ibm.com/downloads/cas/3R8N1DZJ
30. IBM Security (2022), *IBM QRadar Advisor with Watson* [e-publikacija], preuzeto s https://www.ibm.com/downloads/cas/52GBXLK8
31. ISO (2020), *ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence* [e-publikacija], preuzeto s https://www.iso.org/standard/77608.html
32. ISO (2021), *ISO/IEC TR 24030:2021 Information technology — Artificial intelligence (AI) — Use cases* [e-publikacija], preuzeto s https://www.iso.org/standard/77610.html
33. ISO (2023), *ISO/IEC TR 27563:2023 Security and privacy in artificial intelligence use cases — Best practices* [e-publikacija], preuzeto s https://www.iso.org/standard/80396.html
34. Jagannathan, J. i Parvees, M. (2022), Cognitive Intelligence for Interrogation and Inflation of Information Security: A Survey, *Journal of Applied Research and Technology*, *20(5)*, 570-575. https://doi.org/10.22201/icat.24486736e.2022.20.5.1310
35. Joshi, N. (2022), 7 Types Of Artificial Intelligence, preuzeto 28. svibnja 2023. s https://cognitiveworld.com/articles/2021/9/15/7-types-of-artificial-intelligence
36. Kane, G. C., Palmer, D. Phillips, A. N., Kiron, D. i Buckley, N. (2017), Achieving Digital Maturity: Adapting Your Company to a Changing World, u: *MIT Sloan Management Review*, Cambridge, MA: MIT and Deloitte University Press
37. Kant, D. i Johannsen, A. (2022), Evaluation of AI-based use cases for enhancing the cyber security defense of small and medium-sized companies (SMEs), *Journal of Electronic Imaging*, 34, 1-8. https://doi:10.2352/EI.2022.34.3.MOBMU-387

38. Kehayov, M., Holder, L. i Koch, V. (2022), Application of artificial intelligence technology in the manufacturing process and purchasing and supply management, *Procedia Computer Science*, 200, 1209-1217. https://doi.org/10.1016/j.procs.2022.01.321
39. Krishnappa, B. (2015), *Big data analytics for cyber security* [e-publikacija], preuzeto s https://education.dell.com/content/dam/dell-emc/documents/en-us/2015KS_Krishnappa-Big_Data_Analytics_for_Cyber_Security.pdf
40. Lee, D. i Yoon, S.N. (2021), Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges. *International Journal of Environmental Research and Public Health*, 18, 271. https:// doi.org/10.3390/ijerph18010271
41. Lerner, A. (2017), Intent-based Networking, preuzeto s: https://blogs.gartner.com/andrew-lerner/2017/02/07/intent-based-networking/
42. Littman, M.L., Ajunwa, I., Berger, G., Boutilier, C., Currie, M. Doshi-Velez, F. …Walsh, T. (2021), *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report* [e-publikacija], preuzeto s https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/AI100Report_MT_10.pdf
43. Liu, T., Liu, Z., Liu, Q., Wen, W., Xu, W. i Li, M. (2020), Stegonet: Turn deep neural network into a stegomalware, u: *Proceedings of the Annual Computer Security Applications Conference* (str. 928–938.), New York, NY: Association for Computing Machinery
44. Mack, H. (2017), IBM shares data on how Watson augments cancer treatment decision-making, preuzeto 9. lipnja 2023. s https://www.mobihealthnews.com/content/ibm-shares-data-how-watson-augments-cancer-treatment-decision-making
45. McCarthy, J., Minsky, M. L., Rochester, N. i Shannon, C. E. (2006). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955., *AI Magazine*, 27(4), 12. https://doi.org/10.1609/aimag.v27i4.1904
46. Mikhalevich, I.F. i Trapeznikov, V.A. (2019), Critical Infrastructure Security: Alignment of Views, u: *Proceedings of 2019 Systems of Signals Generating and Processing in the Field of on Board Communications* (str.1-5), Moskva: IEEE
47. Muppidi, S., Fisher, L., Parham, G. (2022), *AI and automation for cybersecurity: How leaders succeed by uniting technology and talent* [e-publikacija], preuzeto s https://www.ibm.com/downloads/cas/9NGZA7GK

48. Nilsson, Nils J. (2010) *The Quest for Artificial Intelligence: A History of Ideas and Achievements*, Cambridge: Cambridge University Press
49. NIST (2023), AI measurement and evaluation, preuzeto 12. lipnja 2023. s https://www.nist.gov/ai-measurement-and-evaluation
50. McDaniel, P., Launchbury, J., Martin, B., Wang, C. i Kautz, H. (2020), *Artificial Intelligence and Cybersecurity: A Detailed Technical Workshop Report* [e-publikacija], preuzeto s https://www.nitrd.gov/pubs/AI-CS-Detailed-Technical-Workshop-Report-2020.pdf
51. Parham, G. (2022), 4 Ways AI Capabilities Transform Security, preuzeto 03. ožujka 2023. s https://securityintelligence.com/posts/ai-capabilities-transform-security/
52. Quach, M. (2021), Infrastructure requirements for AI and machine learning, preuzeto 23. svibnja 2023. s https://irendering.net/infrastructure-requirements-for-ai-and-machine-learning/
53. Raza, M. (2021), AI Cyberattacks & How They Work, Explained, preuzeto 27. lipnja 2023. s https://www.bmc.com/blogs/artificial-intelligence-cyberattacks/
54. Rich, M.D., Mills, R.F., Dube, T.E. i Rogers, S.K. (2016), Evaluating Machine Learning Classifiers for Defensive Cyber Operations, *Military Cyber Affairs*, 2 (1), 1-18. http://doi.org/10.5038/2378-0789.2.1.1005
55. Rios Insua, D., Naveiro, R., Gallego, V i Poulos, J. (2020), Adversarial Machine Learning: Perspectives from Adversarial Risk Analysis, *arXiv*, https://doi.org/10.48550/arXiv.2003.03546
56. Rios, D. (2021), AI and Machine Learning: Defense Mechanisms That Need to Be Defended, preuzeto 03. ožujka 2023. s https://axa.foleon.com/axa-research-fund/building-cyber-resilience/2-3-ai-and-machine-learning-defence-mechanisms-that-need-to-be-defended
57. Sarker, I.H., Furhad, H. i Nowrozy, R. (2021), AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions, *SN Computer Science*, 2 (173). https://doi.org/10.1007/s42979-021-00557-0
58. Schrage, M., Kiron, D., Candelon, F., Khodabandeh, S. i Chu, M. (2023), AI Is Helping Companies Redefine, Not Just Improve, Performance, *MITSloan Management Review*,

preuzeto s: https://sloanreview.mit.edu/article/ai-is-helping-companies-redefine-not-just-improve-performance/

59. SecurityScorecard (2023), 21 Cybersecurity Metrics & KPIs to Track in 2023, preuzeto 15. srpnja 2023. s https://securityscorecard.com/blog/9-cybersecurity-metrics-kpis-to-track/
60. Sewpersadh, N.S. (2023), Disruptive business value models in the digital era, *Journal of Innovation and Entrepreneurship*, 12 (2), https://doi.org/10.1186/s13731-022-00252-1
61. Sparapani, J. i Ruma, L. (2021), *Preparing for AI-enabled cyberattacks* [e-publikacija], preuzeto s https://wp.technologyreview.com/wp-content/uploads/2021/04/Preparing-for-AI-enabled-attacks_final.pdf?_ga=2.50604521.1347593536.1686514032-356469198.1686514032
62. Spremić M. (2017), *Digitalna transformacija poslovanja*, Zagreb: Ekonomski fakultet Sveučilišta u Zagrebu
63. Spremić M. (2017), *Sigurnost i revizija informacijskih sustava u okruženju digitalne ekonomije,* Zagreb: Ekonomski fakultet Sveučilišta u Zagrebu
64. Spremić M., Šimunić A. (2018), Cyber Security Challenges in Digital Economy, u: *Proceedings of the World Congress on Engineering 2018* (str. 341-346.), London: International Association of Engineers
65. Spremić, M., Ivancic, L. i Bosilj Vukšić, V. (2020), Fostering Innovation and Value Creation Through Ecosystems: Case of Digital Business Models and Digital Platforms, u: Sandhu, K. (ur.), *Leadership, Management, and Adoption Techniques for Digital Service Innovation,* (str. 25-44), Hershey, PA: IGI Global
66. Stone, P., Brooks, R., Brynjolfsson, E., Calo, R. Etzioni, O., Hager, G. …Teller, A. (2016), *Artificial intelligence and life in 2030:The One Hundred Year Study on Artificial Intelligence (AI100) 2015 Study Panel Report* [e-publikacija], preuzeto s https://ai100.stanford.edu/sites/g/files/sbiybj18871/files/media/file/ai100report10032016fnl_singles.pdf
67. Taylor, J.E.T. i Taylor, G.W. (2021), Artificial cognition: How experimental psychology can help generate explainable artificial intelligence, *Psychon Bull Rev, 28*, 454–475. https://doi.org/10.3758/s13423-020-01825-5
68. Toffler, A. (2022), Powershift: Knowledge, Wealth, and Power at the Edge of the 21st Century, New York, NY: Random House Publishing Group

69. Vassilopoulos, A.P. i Georgopoulos, E.F. (2010), 5 - Novel computational methods for fatig life modeling of composite materials, u: Vassilopoulos, A.P. (ur.), *Fatigue Life Prediction of Composites and Composite Structures* (str. 139-173.), Cambridge: Woodhead Publishing
70. Visvizi, A. i Bodziany, M. (2022), *Artificial Intelligence and Its Contexts: Security, Business and Governance, 1,* Cham: Springer
71. Wang, Z., Liu, C. Cui, X., Yin, J. i Wang, X. (2022), EvilModel 2.0: Bringing Neural Network Models into Malware Attacks, *Computers & Security*, 120, 78-90. https://doi.org/10.1016/j.cose.2022.102807.
72. Wolff, J. (2020), How to improve cybersecurity for artificial intelligence, preuzeto 27. veljače 2023. s https://www.brookings.edu/research/how-to-improve-cybersecurity-for-artificial-intelligence/
73. Wool. A. (2019), Make it So: How Intent-Based Network Security Accelerates the Enterprise, preuzeto 03. ožujka 2023. s https://www.infosecurity-magazine.com/opinions/intent-network-security/
74. Yamin, M.M., Ullah, M.,Ullah, H. i Katt, B. (2021) Weaponized AI for cyber attacks, Journal of Information Security and Applications, 57, 69-79. https://doi.org/10.1016/j.jisa.2020.102722

POPIS SLIKA

POPIS TABLICA