MYKOLO ROMERIO UNIVERSITETAS
VIEŠOJO VALDYMO IR VERSLO FAKULTETAS

JULIJA ZORINA
KIBERNETINIO SAUGUMO VALDYMO PROGRAMA

DIRBTINIO INTELEKTO SISTEMŲ KŪRIMO YPATUMAI KIBERNETINIO SAUGUMO UŽTIKRINIMO KONTEKSTE: PASAULINĖS RINKOS ANALIZĖ

Magistro baigiamasis darbas

Darbo vadovas:
prof. dr. Tadas Limba

VILNIUS, 2026

TURINYS

PAVEIKSLŲ SĄRAŠAS

LENTELIŲ SĄRAŠAS

ĮVADAS

Temos aktualumas. Dirbtinio intelekto technologijos jau egzistuoja daugiau nei 50 metų, tačiau būtent pastaruoju metu galima pastebėti ji itin spartų augimą tiek viešajame, tiek privačiajame sektoriuose. Jos gali pasiūlyti sprendimus kibernetinio saugumo gynybai stiprinti, procesų automatizavimą, efektyviai valdyti incidentus bei suteikia daugelį kitų privalumų, todėl jų taikymas tampa vis aktualesnis.

Visgi, kartu su dirbtinio intelekto technologijų privalumais atsiranda ir naujų iššūkių, kadangi šios technologijos gali būti pasitelkiamos ne tik gynybai, bet ir neteisėtiems tikslams. Šiuolaikinės kibernetinės grėsmės tampa vis išmanesnės, kibernetines atakas vykdantys programišiai vis dažniau pasitelkia pažangius dirbtinio intelekto įrankius, galinčius automatizuoti ir sustiprinti kibernetines atakas, apeiti aukšto lygio taikomas technines bei organizacines saugumo priemones, ir tokiu būdu užvaldyti organizacijos duomenis ir (ar) neigiamai ją paveikti kitu būdu. Atsižvelgiant į tai, labai svarbu šiame kontekste užtikrinti ne tik veiksmingą dirbtinio intelekto sistemų kūrimą jos naudos prasme, bet ir pasiūlyti sprendimus, galinčius sustiprinti atsparumą kibernetinėms atakoms nuo pat jų kūrimo pradžios.

Pažymėtina, kad įstatyminiai reikalavimai susiję su kibernetiniu saugumu bei dirbtiniu intelektu gali skirtis priklausomai nuo atitinkamos valstybės įstatyminės bazės. Priklausomai nuo konkrečios šalies, kurioje kuriama atitinkama dirbtinio intelekto sistema, gali būti skiriamas skirtingas dėmesys kibernetiniam saugumui, o tai dar labiau gali apsunkti tinkamą dirbtinio intelekto technologijų kūrimą saugumo prasme. Nepaisant to, jog dirbtinio intelekto galimybės užtikrinant kibernetinį saugumą nagrinėjamos gana plačiai, tačiau jų kūrimo etapui skiriama nepakankamai dėmesio. Atsižvelgiant į tai, svarbu išnagrinėti dirbtinio intelekto kūrimo ypatumus kibernetinio saugumo kontekste pasauliniu mastu.

Temos ištirtumas. Dirbtinio intelekto taikymas kibernetiniame saugume vis plačiau nagrinėjamas mokslinėje literatūroje, tačiau dažniausiai jie analizuojami fragmentiškai, pavyzdžiui, jo taikymo galimybės, etiniai iššūkiai arba techniniai aspektai.

Dirbtinio intelekto naudojimą užtikrinant kibernetinį saugumą yra nagrinėję tokie autoriai, kaip Akhtar ir Rawol (2024), kurie analizavo dirbtinio intelekto galimybes kibernetinėms grėsmėms aptikti. Kiti autoriai labiau akcentuoja rizikas ir pažeidžiamumus. Pavyzdžiui, Brundage ir kt. (2018) analizavo dirbtinio intelekto panaudojimą kenkėjiškiems tikslams. Polito ir Pupillo (2024) nagrinėjo dirbtinio intelekto sistemų pažeidžiamumą manipuliacijoms, o Jabbarova (2023) išryškino žmogaus priežiūros svarbą dirbtiniu intelektu grindžiamose kibernetinio saugumo sistemose.

Atskira tyrimų kryptis siejama su reguliavimo, atsakomybės ir patikimumo klausimais. Buiten (2020) ir Binns (2018) daugiau dėmesio skiria teisiniams, atsakomybės ir reguliavimo klausimams. Šią temą taip pat plėtoja tarptautinės institucijos, tokios kaip OECD, ENISA ir NIST, akcentuodamos patikimo dirbtinio intelekto, rizikų valdymo ir saugumo užtikrinimo principus bei priemones.

Vis dėlto dirbtinio intelekto sistemų kūrimo ypatumams kibernetinio saugumo užtikrinimo kontekste, ypač vertinant juos pasauliniu mastu ir siejant tai su praktinių gairių poreikiu, vis dar skiriama nepakankamai dėmesio, ko pasėkoje trūksta vieningo ir nuoseklaus šaltinio, leidžiančio visapusiškai susipažinti su šia problematika.

Mokslinis naujumas. Nors dirbtinio intelekto sprendimų naudojimas kibernetinio saugumo kontekste šiandien jau yra nagrinėjamas, vis dėl to dar vis stokojama kompleksiškų tyrimų, kur būtų išsamiai analizuojami tokių sistemų kūrimo ypatumai pasauliniu mastu. Šio baigiamojo darbo mokslinis naujumas siejamas su tuo, kad yra analizuojamas dirbtinio intelekto sistemų kūrimas kibernetinio saugumo kontekste ne tik teoriniu, bet ir praktiniu lygmeniu, taip pat apžvelgiamos pasaulinės tendencijos šioje srityje ir praktinis taikymas konkrečiuose sprendimuose.

Tyrimo objektas. Dirbtinio intelekto sistemų kūrimas kibernetinio saugumo užtikrinimo kontekste.

Mokslinė problema. Kokie dirbtinio intelekto sistemų kūrimo ypatumai išryškėja kibernetinio saugumo užtikrinimo kontekste pasauliniu mastu, atsižvelgiant į rinkos tendencijas, reguliavimo skirtumus ir praktinius sprendimus?

Tyrimo tikslas. Atskleisti dirbtinio intelekto sistemų kūrimo ypatumus kibernetinio saugumo užtikrinimo kontekste, remiantis pasaulinės rinkos analize ir praktinių atvejų lyginamuoju vertinimu.

Tyrimo uždaviniai*:*

1. Išanalizuoti dirbtinio intelekto sampratą, raidą bei pagrindinius dirbtinio intelekto metodus.
2. Įvertinti dirbtinio intelekto taikymo galimybes, iššūkius, rizikas ir kibernetinio saugumo reikalavimus.
3. Ištirti pasaulinės rinkos tendencijas, susijusias su dirbtinio intelekto sprendimų taikymu kibernetinio saugumo srityje.
4. Palyginti pasirinktus dirbtinio intelekto sistemų atvejus, siekiant nustatyti jų kūrimo ypatumus kibernetinio saugumo kontekste.
5. Pateikti rekomendacines gaires dėl dirbtinio intelekto sistemų kūrimo kibernetinio saugumo užtikrinimo kontekste.

Darbo praktinė reikšmė. Darbo rezultatai gali būti reikšmingi organizacijoms, kurios kuria, diegia ar planuoja naudoti dirbtinio intelekto sprendimus siekiant užtikrinti kibernetinį saugumą. Atlikta pasaulinės rinkos analizė ir konkrečių praktinių atvejų lyginamasis vertinimas gali leisti labiau suprasti, kokie aspektai yra svarbiausi kuriant tokias sistemas kibernetinio saugumo požiūriu. Darbe pateiktos rekomendacinės gairės gali būti naudingos planuojant tokių sistemų kūrimą, diegimą arba vertinimą jau naudojamų dirbtinio intelekto sprendimų, skirtų kibernetinio saugumo užtikrinimui.

Darbo struktūra: Mokslinį darbą sudaro: įvadas, teorinė dalis, tyrimo metodologija, tiriamoji dalis, išvados, rekomendacijos bei literatūros sąrašas. Pirmojoje darbo dalyje nagrinėjami dirbtinio intelekto sistemų kūrimo teoriniai aspektai, tokie kaip dirbtinio intelekto samprata, raida ir pagrindiniai metodai, atskleidžiamos dirbtinio intelekto taikymo galimybės kibernetinio saugumo srityje, analizuojami

pagrindiniai iššūkiai, rizikos bei įvairių pasaulio regionų kibernetinio saugumo reikalavimai kuriant tokias sistemas. Antrojoje dalyje pateikiama tyrimo metodologija. Trečiojoje dalyje pristatomas dirbtinio intelekto sistemų kūrimo ypatumų tyrimas kibernetinio saugumo užtikrinimo kontekste, kur yra analizuojamos pasaulinės rinkos raidos kryptys, pristatomi tyrimui pasirinkti atvejai, atliekama jų lyginamoji analizė ir pateikiamos rekomendacinės gairės kuriant dirbtinio intelekto sistemas kibernetinio saugumo kontekste. Darbo pabaigoje pateikiamos išvados, pasiūlymai, literatūros sąrašas ir priedai.

1. DIRBTINIO INTELEKTO SISTEMŲ KŪRIMO TEORINIAI ASPEKTAI

1.1. Dirbtinio intelekto sistemų samprata, raida ir pagrindiniai metodai

Prieš analizuojant dirbtinio intelekto sistemų kūrimą bei jų taikymo ypatumus kibernetiniame saugume, pirmiausia reikalinga išgryninti dirbtinio intelekto sistemų sampratą, kadangi ši sąvoka moksliniuose bei norminiuose šaltiniuose apibrėžiama nevienareikšmiškai. Taip pat pažymėtina, kad ji gali būti aiškinama bendriniu, teisiniu ar technologiniu požiūriu. Be to, siekiant tinkamai suprasti dirbtinio intelekto sistemų kūrimo ypatumus, reikalinga apžvelgti jo istorinę transformaciją nuo teorinio atsiradimo iki šių dienų naudojamų technologijų, taip pat išsiaiškinti pagrindinius jo taikymo metodus.

Šiame darbe nagrinėjamas dirbtinio intelekto sistemų terminas yra kompleksinis, susidedantis iš dviejų tarpusavyje susijusių, bet ir tuo pačiu savarankiškų elementų: a) dirbtinis intelektas bei b) sistemos. Atsižvelgiant į tai, šios sąvokos turinys gali būti nagrinėjamas tiek kaip vientisa koncepcija, tiek ir per jos atskirus elementus. Pavyzdžiui, jeigu dirbtinio intelekto sistemos sąvokos ieškotume lietuvių kalbos žodynuose, jos kaip vientiso termino nerastume, tačiau jo atskirų elementų apibrėžimai yra pateikiami. Bendriniame lietuvių kalbos žodyne (2023) dirbtinis intelektas apibrėžiamas kaip „programine įranga grindžiamos virtualiosios technologijos, kurios demonstruoja protingą ir sumanų elgesį, analizuodamos aplinką ir darydamos gana savarankiškus sprendimus nustatytam tikslui pasiekti“. Tuo tarpu sistema apibrėžiama kaip „aibė elementų, kuriuos sieja tam tikri santykiai bei ryšiai, sudarantys vienybę“ (Bendrinis lietuvių kalbos žodynas, 2023). Sujungiant šias dvi sąvokas, galima išvesti tokį dirbtinio intelekto sistemų apibrėžimą: tai yra tarpusavyje susijusių technologinių elementų visuma, kuri remdamasi duomenų analize, savarankiškai priima sprendimus ir veikia siekdama nustatytų tikslų.

Plėtojantis dirbtinio intelekto reguliavimui, šios sąvokos skirtingus apibrėžimus galima rasti ir tarptautiniuose ar nacionaliniuose teisės aktuose, taip pat ir kituose norminiuose dokumentuose. Pavyzdžiui, Europos Parlamentas (2023) dirbtinį intelektą apibūdina kaip mašinos gebėjimą, kuris yra panašus į žmonių galimybes, tokias kaip argumentavimas, mokymasis, planavimas ir kūrybiškumas. Išsamesnis dirbtinio intelekto sistemų apibrėžimas pateikiamas Dirbtinio intelekto akte (2024), kur jis apibūdinamas kaip: „mašina grindžiama sistema, suprojektuota veikti įvairiais autonomijos lygiais, kuri po diegimo gali veikti prisitaikydama ir kuri, siekiant aiškių ar numanomų tikslų, iš gautos įvesties duomenų daro išvadą, kaip generuoti išvedinius, pavyzdžiui, predikcijas, turinį, rekomendacijas ar sprendimus, kurie gali turėti įtakos fizinei ar virtualiai aplinkai“. Vertinant šiuos du apibrėžimus, galima pastebėti, kad juos sieja vienas bendras elementas – „mašinos gebėjimas“ atlikti tam tikrus veiksmus. Pažymėtina, kad vienu atveju jis glaudžiai siejamas su žmogaus galimybėmis, kitu atveju samprata yra kiek platesnė ir nėra siejama išskirtinai su žmogaus galimybėmis, o labiau pabrėžia ir pačios sistemos autonomiškumą bei prisitaikymą.

Kaip buvo minėta aukščiau, dirbtinio intelekto sąvoką galima dar aiškinti ir per technologinį aspektą. Tokius apibrėžimus pateikia tarptautinės organizacijos bei mokslinių tyrimų institucijos, kaip antai, Tarptautinė standartizacijos organizacija *(angl. International Standardization Organization, ISO)*, Nacionalinis standartų ir technologijų institutas *(angl. National Institute of Standards and Technology, NIST)* bei Ekonominio bendradarbiavimo ir plėtros organizacija *(angl. Organisation for Economic Co-operation and Development, OECD)*. ISO (2026) dirbtinio intelekto sąvoką apibrėžia labai plačiai: kaip kompiuterių mokslo šaką, kuri kuria sistemas ir programinę įrangą, galinčią atlikti užduotis, kurios anksčiau buvo laikomos išskirtinai žmogiškomis. Tuo tarpu NIST (2024) pateikia konkrečius tiek dirbtinio intelekto, tiek jo sistemų sąvokas, kur „dirbtinis intelektas tai mašininė sistema, kuri, esant tam tikram žmogaus apibrėžtų tikslų rinkiniui, gali teikti prognozes, rekomendacijas arba priimti sprendimus, darančius įtaką realiai arba virtualiai aplinkai“; o „dirbtinio intelekto sistema – tai bet kokia duomenų sistema, įranga, programa ar įrankis, kuri visiškai arba iš dalies veikia naudodama dirbtinį intelektą“. OECD (2024) dirbtinio intelekto sistemos sąvoką iš esmės atkartoja NIST pateikiamą apibrėžimą, papildomai akcentuodama skirtingų sistemų autonomijos bei prisitaikymo lygį po diegimo. Autorės manymu, nepaisant skirtingos šios sąvokos aiškinimo apimties, visus šiuos apibrėžimus vienija bendras technologinis požiūris pagal kurį dirbtinis intelektas suprantamas ne kaip žmogaus intelekto imitacija, o kaip pažangi sistema, kuri sugeba ne tik generuoti atitinkamus rezultatus, bet ir daryti poveikį aplinkai.

Apibendrinant skirtingus dirbtinio intelekto sistemų apibrėžimus, nagrinėtus įvairiais požiūriais, galima pastebėti, kad nepaisant jų tam tikrų skirtumų, juose kartojasi tam tikri bendri elementai, kuriuos šio darbo autorė išskirtų, kaip esminius požymius, leidžiančius spręsti, jog atitinkama sistema gali būti laikoma dirbtinio intelekto sistema. Iš analizuotų apibrėžčių galima atrinkti šiuos požymius: 1) mašininis ar programinis veikimas, 2) duomenų apdorojimas, 3) numatyto tikslo siekimas, 4) rezultatų / išvadų generavimas, 5) autonomiškumas ir 6) poveikis fizinei ar virtualiai aplinkai. Pažymėtina, kad panašius esminius elementus išskiria ir Europos Komisija aiškindama, kokiais kriterijais remiantis nustatoma, ar konkreti sistema patenka į ES Dirbtinio intelekto akto reglamentavimo taikymo sritį. Europos Komisija (2024) nurodo tokius elementus, kaip mašininė sistema, autonomiškumas, galimas adaptyvumas (kuris nėra privalomas elementas), sistemos tikslai, gebėjimas daryti išvadas, įvairių išvesčių generavimas (prognozė, turinys, rekomendacijos, sprendimai) bei poveikis fizinei ar virtualiai aplinkai. Tai leidžia daryti išvadą, kad iš analizuotų apibrėžimų išskirti požymiai iš esmės atitinka ir Europos Komisijos suformuluotą dirbtinio intelekto sistemos sampratos logiką, ir jais galima remtis kaip teoriniu pagrindu vertinant dirbtinio intelekto sistemas.

Vis dėl to, autorės manymu, vien sąvokos turinio atskleidimo nepakanka tam, kad būtų galima visapusiškai suprasti dirbtinio intelekto sistemų esmę, kadangi dirbtinis intelektas susiformavo kaip nuoseklios technologinės raidos rezultatas. Atsižvelgiant į tai, toliau tikslinga apžvelgti svarbiausius dirbtinio intelekto raidos įvykius ir kaip keitėsi šių sistemų kūrimas palaipsniui.

Pasak Tim Mucci (2026), žmonės nuo senų laikų svajojo sukurti mąstančias mašinas. Šiuolaikinis dirbtinis intelektas iš dalies kildinamas iš XIX amžiuje Charles Babbage sukurto „skirtumų variklio“ *(angl. difference engine)*, t.y. pirmojo pasaulyje sėkmingai veikiančio automatinio skaičiuotuvo (ISO, 2026). Manytina, kad šiai dienai skaičiuotuvas jau nebėra laikomas dirbtinio intelekto sistema, kaip ji suprantama dabar, tačiau tuo metu tai buvo tikrai reikšmingas įvykis, kadangi sudėtingus skaičiavimo procesus jau galėjo atlikti ne tik žmogus, bet ir automatinė mašina, žymiai palengvinant bei pagreitinant žmogaus atliekamą darbą.

Kalbant apie moderniojo dirbtinio intelekto pradžią, ji paprastai siejama su britų logiku ir kompiuterių pradininku Alanu Turingu (Copeland, 2026). 1950 m. Alanas Turingas (1950) paskelbė svarbų straipsnį „Skaičiavimo mašinos ir intelektas“ *(angl. Computing Machinery and Intelligence)*, kur nagrinėjo esminį klausimą „ar mašinos gali mąstyti?“ ir pasiūlė imitacijos žaidimą, vėliau žinomą kaip Turingo testą (Publications Office of the European Union, 2020). Kaip rašo Copeland (2026), šiame teste dalyvauja trys dalyviai: kompiuteris, žmogus tardytojas ir žmogus „maskuotė“. Tardytojas, užduodamas klausimus per klaviatūrą ir ekraną, turi nustatyti, kuris iš dviejų pašnekovų yra kompiuteris. Kompiuteris gali stengtis suklaidinti tardytoją, o žmogus padėti teisingai atpažinti. Tokiu atveju, jeigu pakankamai daug tardytojų nesugeba atskirti kompiuterio nuo žmogaus, laikoma, kad kompiuteris pasižymi intelektu (Copeland, 2026). Turingo testas buvo pirmasis eksperimentas, pasiūlytas mašinų intelektui matuoti (Publications Office of the European Union, 2020). Vis dėl to, tai buvo ne vieninteliai Alano Turingo pasiekimai dirbtinio intelekto srityje. Dar 1945 m. jis numatė, kad kompiuteriai vieną dieną žais labai gerai šachmatais, o kiek daugiau nei po 50 metų IBM *(angl. „International Business Machines Corporation“)* sukurtas šachmatų kompiuteris „Deep Blue“ šešių partijų mače nugalėjo tuometinį pasaulio čempioną Garį Kasparovą (Copeland, 2026). Autorė mano, jog šis laikotarpis rodo, kaip palaipsniui pradėjo formuotis dirbtinio intelekto idėjos įgyvendinimas, taip pat kaip dalis žmogaus atliekamų darbų pradėjo būti perkeliami mašinai atlikti, ko pasėkoje formavosi tolesnė dirbtinio intelekto raida.

Reikšmingas postūmis įvyko 1956 m. Dartmuto konferencijoje, kur galima sakyti prasidėjo oficialus dirbtinio intelekto laikotarpis, kur McCarthy sugalvojo terminą „dirbtinis intelektas“, kuris vėliau tapo šio mokslo srities pavadinimu (Publications Office of the European Union, 2020). Vis dėl to tuo metu nebuvo viskas taip sklandžiai. ISO (2026) nurodo, kad tuo metu dirbtinis intelektas išgyveno tikrą iššūkį ir jo plėtra buvo pristabdyta, kadangi kompiuteriai negalėjo saugoti pakankamai daug informacijos arba jos pakankamai greitai apdoroti. Tik vėliau, devintajame dešimtmetyje dirbtinis intelektas išgyveno tikrą renesansą, kurį paskatino algoritmų įrankių rinkinio išplėtimas ir finansavimo padidėjimas (ISO, 2026). Autorė nori atkreipti dėmesį į tai, kad nepaisant to, kad oficialiai dirbtinio intelekto terminas jau buvo apibrėžtas, jo teorijos taip pat jau gyvavo kurį laiką, vis dėl to galima matyti, kad vien teorinių idėjų tuo metu nepakako, ir tam, kad dirbtinio intelekto sistemos būtų sklandžiai kuriamos ir vystomos, buvo reikalingas tinkamas technologinis pažangumas, ko tuo metu trūko.

Siekiant dar labiau atskleisti kaip dirbtinis intelektas formavosi iki šiandien žinomos pažangios technologijos, toliau yra pateikiami apibendrinti pagrindiniai dirbtinio intelekto raidos įvykiai ir trumpai aptariama jų esmė 1 lentelėje.

1 lentelė. Pagrindiniai dirbtinio intelekto raidos įvykiai.

Laikotarpis / metai	**Pagrindinis įvykis**	**Esmė**
1950 m.	Alano Turingo „Turingo testas“	Jis pasiūlė praktinį kriterijų pagal kurį, jeigu mašina geba įtikinamai bendrauti kaip žmogus, ji pasižymi intelektu.
1956 m.	Dartmuto konferencija	Dirbtinis intelektas gavo savo pavadinimą ir tapo aiškiai apibrėžta tyrimų sritimi.
1960–1970 m.	Simbolinis dirbtinis intelektas	Tyrėjai manė, kad jei galėtų užkoduoti pakankamai taisyklių ir faktų, galėtų sukurti intelektualias sistemas, kurios samprotautų kaip žmonės. Vis dėl to paaiškėjo, kad kalba, suvokimas, kontekstas ir neapibrėžtumas yra pernelyg sudėtingi vien logika grindžiamiems metodams.
1970–1980 m.	Pirmoji „dirbtinio intelekto žiema“	Finansavimas ir entuziazmas sumažėjo, nes technologijos nedavė to, ko žmonės tikėjosi.
1980–1990 m.	Mašininio mokymosi iškilimas	Tyrėjai vis dažniau tyrinėjo mašininį mokymąsi, kai sistemos mokosi iš duomenų. Šis metodas galiausiai tapo šiuolaikinio dirbtinio intelekto pagrindu.
2000–2010 m.	Dideli duomenys, geresni skaičiavimai ir gilusis mokymasis	Pažangą paspartino trys pagrindiniai veiksniai: dideli duomenų kiekiai, galingesnė infrastruktūra ir pažangesni giliojo mokymosi algoritmai. Šių veiksnių derinys sudarė sąlygas kurti gerokai efektyvesnius modelius.
2012 m.	Giliojo mokymosi proveržis	Gilusis mokymasis pradėjo lenkti ankstesnius metodus. Tai signalizavo, kad dirbtinis intelektas gali plėstis su didesniu duomenų kiekiu ir skaičiavimo pajėgumu.
2016–2019 m.	Realus dirbtinio intelekto naudojimas	Dirbtinis intelektas pradėtas taikyti sukčiavimo aptikime, medicininių vaizdų analizėje, vertime, kalbos atpažinime ir pramonės optimizavime. Jis tapo nebe eksperimentine, o praktiškai naudojama technologija.

Šaltinis: Sudaryta autorės remiantis Swiss Cyber Institute (2026).

Apibendrinant 1 lentelėje pateikiamus duomenis, galima daryti išvadą, jog dirbtinis intelektas formavosi nuosekliai nuo teorinių idėjų iki šiandien naudojamų pažangių sudėtingų sistemų. Iš lentelės pateiktų duomenų galima matyti, kad dirbtinio intelekto vystymasis iš pradžių buvo ganėtinai lėtas, nes tuo metu buvo ribotos technologinės galimybės, tačiau palaipsniui augant technologinei pažangai, taip pat didėjant duomenų kiekiams, tai tapo reikšmingu varikliu, kurio dėka šiuolaikinis dirbtinis intelektas įgijo visai kitą mąstą ir plėtros tempą. Autorės manymu, dirbtinio intelekto raida yra reikšminga dabartinių sistemų kūrimui, kadangi leidžia suprasti, kurie kūrimo ypatumai tuo metu pasiteisino, o kurie iš jų buvo nesėkmingi.

Kalbant apie šiandieninį dirbtinį intelektą, jis yra galingesnis nei bet kada anksčiau. Jis mato, klausosi, reaguoja, mokosi iš patirties, tobulina savo įgūdžius ir sklandžiai integruojasi į mūsų kasdienį gyvenimą (ISO, 2026). Iš to susidaro išvada, kad šiuo metu jis yra kur kas pažangesnis nei ankstesniais laikotarpiais. Dabartinis dirbtinio intelekto raidos etapas siejamas su generatyviniu dirbtiniu intelektu, kuris ne tik apdoroja duomenis, bet ir kuria muziką, generuoja tikroviškus vaizdus, tekstus neatskiriamus nuo žmogaus darbo (ISO, 2026). Pasak Polito ir Pupillo (2024), generatyvinio dirbtinio intelekto sistemos veikia remdamosi naudotojo pateiktu užklausos *(angl. prompt)* principu ir, analizuodamos didelius duomenų kiekius bei juose esančius modelius, generuoja naują ir originalų turinį. Šios technologijos dažnai grindžiamos dideliais kalbos modeliais *(angl. Large Language Models, LLM)*, kurie naudoja mašininį mokymąsi žmogaus kalbai suprasti ir generuoti. Tokie modeliai mokomi naudojant didelius tekstinių duomenų rinkinius (knygas, straipsnius, interneto turinį), todėl geba kurti tekstą, panašų į žmogaus parašytą ar pasakytą (Polito ir Pupillo, 2024). Darbo autorė pastebi, kad šiuo metu dirbtinio intelekto sistemų kūrimas iš esmės yra pasikeitęs palyginus su ankstesniais laikotarpiais, kai jis buvo grindžiamas iš anksto apibrėžtomis taisyklėmis. Žinoma, dabartinė technologine pažanga bei dideli duomenų kiekiai turi tam didelę įtaką, bet ne mažiau svarbu atkreipti dėmesį ir į tai, kokiais metodais šiomis dienomis grindžiamas tokių sistemų kūrimas.

Analizuojant dirbtinio intelekto metodus, svarbu pažymėti, kad šiame darbe yra aktualūs ne visi egzistuojantys dirbtinio intelekto metodai ar technikos, o tik tie, kurie yra siejami išskirtinai su dirbtinio intelekto sistemų kūrimo etapu. Dėl šios priežasties toliau yra aptariami Europos Komisijos gairėse dėl dirbtinio intelekto sistemų apibrėžimų (2024) pateikiami pagrindiniai dirbtinio intelekto metodai, kurie yra skirti būtent aptariamajam etapui. Jose nurodoma, kad dirbtinio intelekto metodai gali būti skirstomi į dvi pagrindines kategorijas: 1) mašininio mokymosi metodus, kurie leidžia sistemai mokytis iš duomenų siekiant tam tikrų tikslų, ir 2) logika bei žiniomis pagrįstus metodus, kurie grindžiami žmogaus ekspertų užkoduotomis žiniomis arba simboline spręstinos užduoties reprezentacija. Pirmąją grupę sudaro gana nemažai metodų: prižiūrimas mokymasis, neprižiūrimas mokymasis, saviprižiūrimas mokymas, sustiprinamasis mokymasis bei gilusis mokymais. Tuo tarpu antrąją kategoriją sudaro žinių reprezentacija, žinių bazės, išvadų mechanizmai, simbolinis samprotavimas, ekspertinės sistemos bei paieškos ir optimizavimo metodai (Europos Komisija, 2024). Autorė mano, kad toks skirstymas siekia parodyti, jog sistemos gali būti kuriamos taip, kad mokytųsi iš duomenų, arba taip, kad veiktų pagal iš anksto sukonfigūruotas taisykles. Siekiant atskleisti šių metodų reikšmę dirbtinio intelekto sistemų kūrimui, toliau bus aptariamas kiekvienas iš jų išsamiau.

Kaip minėta, nemažai dirbtinio intelekto metodų sudaro mašininio mokymosi, kurių visų bendras bruožas, vien sprendžiant iš pavadinimo, yra mokymasis. Minėtą kategoriją sudaro šie metodai:

- Prižiūrimas mokymasis. Šiuo atveju dirbtinio intelekto sistema mokosi iš pažymėtų duomenų, kai įvesties duomenys yra susieti su teisinga išvestimi. Tai reiškia, kad sistema išmoksta atpažinti ryšį

tarp įvesties ir rezultato, o vėliau gali šį ryšį taikyti naujiems, anksčiau nematytiems duomenims. Šis metodas dažnai taikomas tokiose srityse kaip elektroninio pašto „šlamšto“ aptikimas, pavyzdžiui, kūrimo etape sistema apmokoma naudojant duomenų rinkinį, kuriuose yra el. laiškai, kuriuos žmonės pažymėjo kaip „šlamštas“ arba „ne šlamštas“, kad būtų galima apmokyti modelius iš pažymėtų savybių (Europos Komisija, 2024).

- Neprižiūrimas mokymasis. Šis metodas pasireiškia taip, kad sistema mokosi iš nepažymėtų duomenų ir savarankiškai ieško juose struktūrų, ryšių ar pasikartojančių modelių. Pavyzdžiui, farmacijos įmonėse naudojamos dirbtinio intelekto sistemos, skirtos naujų vaistų atradimui, yra neprižiūrimo mokymosi pavyzdys (Europos Komisija, 2024).
- Saviprižiūrimas mokymasis. Jis laikomas neprižiūrimo mokymosi atmaina, kai sistema mokymosi tikslus susikuria iš pačių duomenų. Tai ypač svarbu šiuolaikinėms kalbos ir vaizdų atpažinimo sistemoms, nes toks metodas leidžia mokytis iš labai didelių duomenų kiekių net ir tada, kai jie nėra pažymėti žmogaus. Pavyzdžiui, vaizdų atpažinimo sistema gali mokytis atpažinti objektus prognozuodama trūkstamus vaizdo pikselius (Europos Komisija, 2024).
- Sustiprinamasis mokymasis. Jo pagrindu sistema mokosi iš savo pačios patirties, remdamasi aplinkos suteikiamu grįžtamuoju ryšiu. Tokiu atveju sistema palaipsniui tobulina savo veiksmus bandymų ir klaidų būdu, siekdama kuo geresnio rezultato. Šis metodas gali būti taikomas robotikoje, autonominio valdymo sistemose ar personalizuotose rekomendacijų sistemose (Europos Komisija, 2024).
- Gilusis mokymasis. Šis metodas grindžiamas daugiasluoksniais neuroniniais tinklais, leidžiančiais sistemai automatiškai išmokti duomenų požymius iš pirminių duomenų, todėl nereikia rankiniu būdu apibrėžti požymių. Dėl didelio sluoksnių ir parametrų skaičiaus tokiems modeliams paprastai reikia didelių duomenų kiekių ir reikšmingų skaičiavimo išteklių, tačiau būtent šis metodas yra plačiai naudojamas ir yra daugelio naujausių dirbtinio intelekto pasiekimų pagrindas (Europos Komisija, 2024).

Autorės manymu, mašininio mokymosi metodų įvairovė rodo, kad šiuolaikinių dirbtinio intelekto sistemų kūrimas vis labiau siejamas ne tik su programavimo logika, bet ir su tinkamu duomenų parinkimu, atitinkamo modelio mokymu bei jo nuolatiniu tobulinimu.

Be mašininio mokymosi metodų, Europos Komisija (2024) išskiria ir logika bei žiniomis pagrįstus metodus. Skirtumas nuo duomenimis pagrįstų metodų yra toks, kad tokios sistemos remiasi žmogaus ekspertų užkoduotomis žiniomis, įskaitant taisykles, faktus ir ryšius. Tokios sistemos gali samprotauti naudodamos dedukcinius ar indukcinius mechanizmus, paieškos, atitikimo ar taisyklių grandinės sudarymo operacijas. Šiai kategorijai priskiriamos žinių bazės, išvadų mechanizmai, simbolinis samprotavimas, ekspertinės sistemos bei paieškos ir optimizavimo metodai (Europos Komisija, 2024). Kaip pažymi pati Europos Komisija (2024), šie metodai buvo ypač svarbūs ankstyvajame dirbtinio intelekto vystymosi etape,

tačiau ir šiandien išlieka reikšmingi tais atvejais, kai svarbus didesnis sprendimų aiškumas ir paaiškinamumas.

Taigi, aptarti pagrindiniai dirbtinio intelekto metodai atskleidžia skirtingą dirbtinio intelekto sistemų veikimo ir kūrimo logiką. Vienais atvejais sistemos kuriamos remiantis duomenimis ir jų pagrindu vykstančiu mokymusi, kitais iš anksto apibrėžtomis taisyklėmis ir žinių struktūromis. Iš to galima daryti išvadą, kad šis skirtumas yra reikšmingas ir šio darbo kontekste, kadangi nuo pasirinkto metodo gali priklausyti ir jos kūrimo procesas, taip pat kokie ištekliai bus reikalingi bei su tuo susiję galimi kibernetinio saugumo iššūkiai.

Apibendrinant tai, kas buvo parašyta, galima teigti, kad nors šiandien dirbtinio intelekto sistemų sąvoka jau yra apibrėžta skirtinguose šaltiniuose, jos turinys nėra suprantamas vienodai, ypatingai pasauliniu mastu. Priklausomai nuo to, kokiame kontekste ši sąvoka vartojama, ji gali būti aiškinama bendriniu, teisiniu ar technologiniu požiūriu. Vis dėlto, kalbant apie dirbtinio intelekto sistemų kūrimą, išsamiausia bei tiksliausia laikytina Europos Komisijos pateikta šios sąvokos samprata ir jos išskirti elementai, kurie leidžia aiškiau suprasti, kas laikytina dirbtinio intelekto sistema. Ne mažiau svarbi yra ir dirbtinio intelekto raidos analizė, nes ji parodo ne tik tai, kaip ši technologija tobulėjo iki šių dienų, bet ir leidžia suprasti, kaip keitėsi jų kūrimo principai, kas pasiteisino ir gali būti naudojama šiandien kuriant tokias sistemas. Taip pat autorė nori pažymėti, jog dirbtinio intelekto sistemų kūrimo etape turėtų būti skiriamas dėmesys ir atitinkamų metodų analizei prieš jų pasirinkimą, kurie, leis nustatyti kryptį, kokiu principu sistema bus kuriama, kaip ji veiks ir kokių išteklių ar valdymo sprendimų prireiks. Šie metodai, kaip buvo nustatyta, skirstomi į mašininio mokymosi metodus ir logika bei žiniomis pagrįstus metodus.

1.2. Dirbtinio intelekto taikymo galimybės kibernetinio saugumo srityje

Nustačius dirbtinio intelekto sistemų sampratą, apžvelgus jo raidą bei išanalizavus jo pagrindinius metodus kūrimo etape, taip pat svarbu peržiūrėti kaip šios sistemos taikomos praktikoje. Turint omenyje, kad šio darbo kontekste yra aktuali tik kibernetinio saugumo erdvė, šioje dalyje bus apžvelgiamos dirbtinio intelekto taikymo galimybės būtent kibernetinio saugumo srityje.

Pasak Shanbhag, Dawson ir Etori (2024), dirbtinio intelekto priemonės gali būti naudojamas įvairiose kibernetinio saugumo srityse, nuo grėsmių aptikimo iki incidentų valdymo, kaip antai:

- Automatizuota rizikos analizė ir poveikio vertinimas. Dirbtinis intelektas gali automatizuoti rizikos balų skaičiavimą, nustatyti saugumo incidentų tikimybę ir pagrindinius pažeidžiamumo rizikos rodiklius, taip pat padėti atlikti rizikos vertinimą ir sprendimų analizę naudojant žurnalų *(angl. log)* duomenis ir grėsmių žvalgybos informaciją (Shanbhag ir kt., 2024 cituojant Sancho ir kt., 2020).
- Prognozinė kibernetinė žvalgyba *(angl. predictive intelligence)*. Remdamasis istoriniais duomenimis, dirbtinis intelektas gali prognozuoti būsimas saugumo grėsmes ir pažeidžiamumus,

taip sudarydamas sąlygas organizacijoms imtis prevencinių priemonių dar prieš įvykstant atakai. Dirbtinio intelekto pagrindu veikiančios prognozinės sistemos gali numatyti galimų įsibrovimų tipą, intensyvumą ir taikinius, taip pat prognozuoti kenkėjiškų programų veikimą (Shanbhag ir kt., 2024 cituojant Kaur ir kt. 2023).

- Grėsmių aptikimas. Mašininio mokymosi modeliai nuolat tobulina savo algoritmus remdamiesi naujais duomenimis, todėl didėja grėsmių aptikimo tikslumas ir efektyvumas. Tai yra daroma tam, kad būtų neatsiliekama nuo sparčiai besikeičiančių kibernetinių grėsmių, nes programišiai nuolat keičia savo taktikas, kad apeitų įdiegtas saugumo priemones (Shanbhag ir kt., 2024).
- Elgsenos analizė. Analizuodamas įprastą vartotojų elgseną, dirbtinis intelektas gali nustatyti nukrypimus, kurie gali rodyti vidines grėsmes arba užkrėstas vartotojų paskyras. Tokia analizė pasižymi savo pranašumu aptinkant sudėtingas kibernetines atakas, kurios gali likti nepastebėtos tradicinių saugumo priemonių (Shanbhag ir kt., 2024 cituojant Calderon ir kt., 2019).

Autorės manymu, šių priemonių įvairovė rodo, kad jau šiandien egzistuoja nemažai konkrečių dirbtinio intelekto priemonių, kurios yra taikomos siekiant užtikrinti organizacijos kibernetinį atsparumą. Dar svarbiau atkreipti dėmesį, kad tokios priemonės nuolat tobulinamos ir pačios mokosi iš savo patirčių, kas ypač aktualu nuolat besikeičiančioje kibernetinio saugumo erdvėje. Tai patvirtina ir Akhtar ir Rawol (2024) nurodydami, jog skirtingai nuo tradicinių taisyklėmis pagrįstų sistemų, kurioms reikalingi nuolatiniai atnaujinimai, kad būtų galima kovoti su kylančiomis grėsmėmis, dinamiškas dirbtinio intelekto pobūdis leidžia aktyviai reaguoti į sparčiai besivystančias atakas.

Pasak Shanbhag (2024), tarp įvairių kuriamų technologijų ypač išsiskiria įsibrovimų aptikimo sistemos *(angl. Intrusion Detection Systems)*, kurios pasitelkia dirbtinį intelektą potencialiems saugumo pažeidimams nustatyti. Tradicinės įsibrovimo aptikimo sistemos dažniausiai buvo pagrįstos taisyklėmis ir rėmėsi žinomų kenkėjiškų programų požymiais bei iš anksto užprogramuotais anomalijų aptikimo metodais, tačiau dėl nuolat besikeičiančių kibernetinių grėsmių, naujų kenkėjiškų programų ir sudėtingų atakų strategijų atsiradimo, tapo būtina kurti pažangesnes ir labiau prisitaikančias sistemas (Shanbhag ir kiti, 2024 cituojant Gupta ir kt. 2023). Pažymėtina, kad ir šie autoriai atkreipia dėmesį į tai, kad dirbtinio intelekto priemonių naudojimas lemiamas ne tik pačių priemonių pažangumu, o labiau būtinybe prisitaikyti prie naujų kibernetinių atakų. Šią mintį patvirtina ir Michael, Abbas ir Roussos (2023) nurodydami, kad dirbtinis intelektas yra geresnis stebėtojas ir ieškotojas nei žmogus, nes jis visada ieško išimčių tinkle ir netgi savarankiškai sprendžia dėl geriausio veiksmų plano.

Dar viena sritis kur gali būti taikomas dirbtinis intelektas yra kibernetinio saugumo incidentų valdymas ir prevencija. Remiantis Akhtar ir Rawol (2024), jis padeda greičiau ir tiksliau reaguoti į kibernetines atakas, automatizuodamas saugumo įspėjimų analizę, prioretizavimą bei įvairių duomenų šaltinių apdorojimą, ir tokiu būdu sumažindamas analitikų darbo apkrovą, pagreitėja sprendimų priėmimas ir sumažėja klaidingų perspėjimų tikimybė. Be to, dirbtinis intelektas automatizuoja tokias užduotis kaip

kenkėjiškų programų analizė, žurnalų interpretavimas ir pažeidžiamumų vertinimas, todėl IT specialistai gali daugiau dėmesio skirti sudėtingesnėms strateginėms saugumo užduotims (Akhtar ir Rawol, 2024).

Dirbtinis intelektas taip pat suteikia galimybę stiprinti proaktyvų kibernetinį saugumą. Tai reiškia, kad jis gali būti naudojamas ne tik jau įvykusių ar vykstančių incidentų nustatymui, bet ir būsimų grėsmių prognozavimui, remdamasis istoriniais duomenimis (Shanbhag ir kt., 2024). Manytina, kad tai gali būti svarbu organizacijai kuriant savo kibernetinio saugumo strategiją.

Akhtar ir Rawol (2024) taip pat pažymi, kad dirbtinis intelektas gali gerinti vartotojų autentifikavimą ir prieigos kontrolę, nes analizuoja naudotojų elgseną, įrenginių savybes ir kitus kontekstinius duomenis. Iš to daroma išvada, jog dirbtinio intelekto taikymas šioje srityje leidžia stiprinti ne tik techninę, bet ir organizacinę kibernetinio saugumo pusę.

Praktiniai pavyzdžiai rodo, kad dirbtinio intelekto taikymas kibernetiniame saugume jau seniai nebėra vien teorinis svarstymas. Lazic (2019) pateikia Siemens AG praktinį pavyzdį, kai technologijų bendrovė savo kibernetinės gynybos centre sukūrė dirbtiniu intelektu pagrįstą platformą, veikiančią naudojant Amazon Web Services infrastruktūrą, kuri leidžia automatiškai įvertinti dešimtis tūkstančių galimų grėsmių per sekundę. Tokia sistema leidžia valdyti didelės apimties saugumo procesus naudojant nedidelę specialistų komandą ir nepaveikiant sistemų veikimo. Be to, Lazic (2019) pažymi, kad dirbtinio intelekto sprendimus kibernetinio saugumo srityje jau taiko įvairios technologijų įmonės ir saugumo platformų kūrėjai kaip: Check Point, CrowdStrike, FireEye, Fortinet, LogRhythm, Palo Alto Networks, Sophos Symantec. Autorės manymu, tas faktas, jog didelės žinomos bendrovės jau šiandien taiko dirbtinio intelekto priemones užtikrinant kibernetinį saugumą, patvirtina, jog tokių sprendimų nauda yra ne tik teoriniame lygmenyje, bet ir rodo jos realų praktinį įgyvendinimą.

Vis dėl to požiūris į dirbtinio intelekto priemonių naudojimą kibernetinio saugumo srityje gali būti vienų autorių vertinamas kaip papildomas privalumas efektyvumo prasme, o kitų kaip neatsiejama saugumo priemone. Pavyzdžiui, ISO (n.d.) pažymi, kad vienas pagrindinių dirbtinio intelekto privalumų yra gebėjimas automatizuoti sudėtingus procesus, taip didinant veiklos efektyvumą ir mažinant žmogaus darbo krūvį. Europos Parlamentas (2020) taip pat pažymi, kad dirbtinio intelekto sistemos gali padėti atpažinti kibernetines atakas ir kitas kibernetines grėsmes. O Michael ir kt. (2023) net išskiria keturis pagrindinius etapus, kai dirbtinis intelektas naudojamas kaip konkurencinis pranašumas organizacijos kibernetinio saugumo kontekste: 1) esamo kibernetinio saugumo gyvavimo ciklo, jo pagrindinių principų, užduočių ir procesų supratimas; 2) prieinamų vidinių ir išorinių duomenų rinkimas ir agregavimas iš įvairių šaltinių; 3) dirbtiniu intelektu pagrįstos duomenų analizės taikymas surinktiems duomenims, leidžiantis daryti išvadas ir prognozes apie modelius bei tendencijas, kurioms gali prireikti skubaus dėmesio; 4) žinių naudojimas ir sklaida pagal poreikį, siekiant palaikyti organizacijos kibernetinio saugumo tikslus.

Tuo tarpu, Shanbhag ir kt. pažymi, jog nuolat besikeičiančioje kibernetinio saugumo aplinkoje dirbtinio intelekto technologijų integravimas tampa svarbia strategija, siekiant pagerinti kibernetinių

grėsmių aptikimą ir prevenciją (Shanbhag ir kiti, 2024 cituojant Gupta ir kt. 2023). Panašios nuomonės laikosi Khan ir kt. (2024), pabrėždami, kad dirbtinio intelekto priemonių poreikis kyla dėl to, kad kibernetinės grėsmės tampa vis dažnesnės, sudėtingesnės ir sunkiau prognozuojamos, todėl vien tradicinės saugumo priemonės ne visuomet yra pakankamos. Apibendrinus šias įžvalgas, galima teigti, kad dirbtinio intelekto sprendimų taikymas kibernetinio saugumo srityje be abejonių suteikia tam tikrų privalumų, tačiau autorės manymu, reikėtų ją traktuoti kaip būtiną saugumo priemonę, siekiant tinkamai reaguoti į pažangias kibernetinio saugumo grėsmes.

Papildomai pažymėtina, kad anksčiau kibernetinės grėsmės dažnai likdavo nepastebėtos net geriausiai parengtų saugumo specialistų arba jų nustatymas ir neutralizavimas reikalavo kelių dienų algoritmų kūrimo, siekiant kovoti su nuolat kintančiomis atakomis (Michael ir kt. 2023). Autorės manymu, tai visiškai suprantama, kadangi kibernetinėms atakoms vykdyti vis dažniau naudojami tie patys dirbtinio intelekto įrankiai, atitinkamai norint tinkamai nuo jų apsisaugoti, turi būti taip pat naudojamos inovatyvios saugumo priemonės.

Kalbant apie dirbtinio intelekto taikymą kibernetinio saugumo srityje, darbo autorės vertinimu, reikia turėti omenyje ne tik gynybai nuo kibernetinių atakų skirtas technologijas, bet ir tas atakas sukeliančias priemones. Įvairios kibernetinės grėsmės, tokios kaip išpirkos reikalaujančios programos *(angl. ransomware)*, nesaugūs daiktų interneto įrenginiai, fišingo *(angl. phishing)* atakos ir vidinės grėsmės organizacijose, kelia didelę riziką informacinių sistemų saugumui (Akhtar ir Rawol, 2024). Pasak Lazic (2019), kibernetinės grėsmės gali pasireikšti bet kuriuo metu, todėl vien tik žmogaus vykdoma saugumo stebėsena dažnai nėra pakankama. Jis taip pabrėžia, kad efektyvi kibernetinio saugumo sistema turi gebėti nuolat aptikti, nustatyti ir neutralizuoti galimas grėsmes (Lazic, 2019). Šias įžvalgas patvirtina ir praktiniai pavyzdžiai. Kaip pažymi Akhtar ir Rawol, (2024), tokie incidentai kaip 2017 m. „WannaCry“ ataka prieš Jungtinės Karalystės sveikatos apsaugos sistemą, „Mirai“ kenkėjiškos programos naudojimas „DDoS“ atakoms, plačiai paplitę fišingo bandymai ar vidinių darbuotojų piktnaudžiavimo atvejai parodo, kad kibernetinės grėsmės gali kilti tiek iš išorės, tiek iš organizacijos vidaus. Šie autoriai taip pat nurodo, jog dirbtinis intelektas gali padėti šias rizikas valdyti stebėdamas tinklo srautą, analizuodamas vartotojų bei įrenginių elgseną ir laiku nustatydamas neįprastus veiklos modelius. Tokiu būdu tampa įmanoma anksti aptikti galimas atakas ir imtis kitų prevencinių saugumo priemonių. Vis dėlto autoriai taip pat pabrėžia, kad vien technologinių priemonių nepakanka, todėl išlieka svarbios ir organizacinės saugumo priemonės, kaip antai, darbuotojų mokymas atpažinti fišingo bandymus bei laikytis saugaus elgesio darbe taisyklių (Akhtar ir Rawol, 2024). Autorė mano, kad šiuo atveju susidaro paradoksas, kai įprastos saugumo priemonės yra nepakankamos užtikrinti tinkamą kibernetinį saugumą, nes žmogaus galimybės neleidžia tinkamai reaguoti į kylančias grėsmes, tačiau dirbtinio intelekto technologijų naudojimas be žmogaus dalyvavimo taip pat negali būti laikomas visiškai patikima saugumo priemonė, todėl žmogaus įsitraukimas dažnai tampa privalomas norint pasiekti saugiausią variantą. Vis dėl to, Liubomir Lazic (2019) pažymi, kad dirbtiniu

intelektu pagrįsti sprendimai suteikia reikšmingą pranašumą ir leidžia užtikrinti nuolatinę apsaugą net ir tais laikotarpiais, kai žmogaus įsikišimas yra ribotas ar neįmanomas, taip reikšmingai padidinant organizacijos kibernetinį atsparumą. Atsižvelgiant į tai, galima daryti išvadą, kad norint šiandien užtikrinti tinkamą kibernetinio saugumo lygį organizacijoje, reikalinga įvertinti dirbtinio intelekto pagrindu veikiančių saugumo priemonių įtraukimą į atitinkamos organizacijos procesus.

Atskiras dėmesys skirtinas generatyvinio dirbtinio intelekto taikymui kibernetiniame saugume. Kaip pažymi Polito ir Pupillo (2024), generatyvinis dirbtinis intelektas gali padėti automatizuoti pasikartojančias saugumo užduotis, taip suteikdamas specialistams daugiau laiko spręsti sudėtingesnes problemas, taip pat gali padėti aptikti neįprastus veiklos modelius ar tendencijas, kurias žmogaus analitikai galėtų likti nepastebėti, bei padėti kurti kibernetinių grėsmių scenarijus mokymams bei simuliacijoms ir prognozuoti galimas ateities grėsmes remiantis istoriniais duomenimis. Tai leidžia manyti, kad generatyvinio dirbtinio intelekto priemonės gali reikšmingai prisidėti prie efektyvesnio kibernetinio saugumo užtikrinimo.

Vis dėl to, Jabbarova (2023) mato ir kitą generatyvinio dirbtinio intelekto pusę, ir pateikia jo pagrindu vykdomų kibernetinių atakų pavyzdį, kai jis naudojamas suklastotam skaitmeniniam turiniui kurti. Šio įrankio pagalba galima generuoti realistiškai atrodančius vaizdus ar garso įrašus, kurie gali būti naudojami ir piktavališkiems tikslams, pavyzdžiui, kuriant įtikinamą, bet klaidingą medijų turinį, kuris gali būti naudojamas fišingo atakoms, dezinformacijos kampanijoms ar kitoms kibernetinio nusikalstamumo formoms (Jabbarova, 2023). Autorės manymu, šios dienos dirbtinio intelekto transformaciją jau pasiekė tokį lygį, kad net nebūtina būti IT specialistu, norint panaudoti dirbtinio intelekto priemones kenkėjiškiems tikslams. Vienas iš tokių žinomų pavyzdžių yra vadinamieji „deepfake“ vaizdo įrašai, kurie buvo plačiai aptariami 2019 m. Indijos visuotinių rinkimų metu, kai manipuliuotas vaizdo ir garso turinys buvo naudojamas klaidinančiai informacijai ir politinei propagandai skleisti (Jabbarova, 2023). Kitas pavyzdys yra mokslininkų atliktas eksperimentas, kurio metu buvo sukurtas suklastotas buvusio JAV prezidento Barako Obamos vaizdo įrašas, siekiant parodyti, kaip įtikinamai dirbtinio intelekto technologijos gali sukurti netikrą kalbą (Jabbarova, 2023). Jabbarova (2023) kartu pažymi, kad tokie pavyzdžiai rodo, jog būtina kurti veiksmingas apsaugos priemones bei užtikrinti atsakingą dirbtinio intelekto technologijų kūrimą ir reguliavimą. Šio darbo autorė sutinka su išsakyta mintimi ir svarsto, kad tinkamai nesureguliavus dirbtinio intelekto priemonių naudojimo, ypač viešai jo prieinamų formų, tai gali turėti reikšmingą neigiamą įtaką ne tik organizacijų, bet ir apskritai žmonių saugumui.

Apibendrinant galima teigti, kad dirbtinis intelektas kibernetinio saugumo srityje pasireiškia šiomis praktinėmis galimybėmis, kaip automatizuota rizikos analizė, prognozinė kibernetinė žvalgyba, grėsmių aptikimas, elgsenos analizė, įsibrovimų aptikimo sistemos, kibernetinio saugumo incidentų valdymas ir prevencija, vartotojų autentifikavimo ir prieigos kontrolės stiprinimas. Vis dėl to, egzistuoja ir kita šių priemonių pusė, kai jos gali būti taikomos kibernetinėms atakoms, tokioms kaip fišingas, išpirkos

reikalaujančios programos, ar kitoms kibernetinio nusikalstamumo formoms. Atsižvelgiant į tai, dirbtinio intelekto sistemų kūrimas kibernetinio saugumo kontekste turėtų būti traktuojamas ne tik kaip papildomų privalumų suteikiantis sprendimas, bet kaip neatsiejama organizacijos saugumo priemonė, galinti sukurti atsparumą tuo pačiu technologiniu pagrindu kuriamoms atakoms.

1.3. Iššūkiai ir rizikos kuriant dirbtinio intelekto sistemas

Nepaisant jau aptartų dirbtinio intelekto suteikiamų privalumų kibernetinio saugumo srityje, jo taikymas taip pat gali būti susijęs ir su įvairiais iššūkiais bei rizikomis. Šioje dalyje bus siekiama nustatyti į kokius riziką keliančias aspektus reikalinga atsižvelgti, bei kokius iššūkius reikia išspręsti prieš pradedant kurti ar naudoti dirbtinio intelekto sistemas kibernetinio saugumo kontekste.

Akhtar ir Rawol (2024) nurodo, kad dirbtinis intelektas kibernetinio saugumo srityje susiduria su unikaliais iššūkiais, kuriuos lemia platus atakų paviršius, kvalifikuotų specialistų trūkumas ir labai dideli duomenų kiekiai. Viena vertus, būtent šie veiksniai skatina kurti pažangias, savarankiškai besimokančias sistemas, galinčias rinkti, analizuoti ir susieti duomenis, siekiant sustiprinti organizacijos saugumą. Kita vertus, jie parodo, kad dirbtinio intelekto kūrimas šioje srityje negali būti suprantamas vien kaip technologinės pažangos klausimas, nes kartu atsiranda ir naujų saugumo, kontrolės bei atsakomybės problemų. Atsižvelgiant į tai, manytina, kad rizikų bei iššūkių spektras gali būti ganėtinai platus, tad reikalinga išsamiau pasigilinti į šiuos galimus neigiamus veiksmus.

Kalbant apie pažangias sistemas, pirmiausia reikalinga įvertinti su kokiais technologiniais iššūkiais galima susidurti taikant dirbtinio intelekto sistemas kibernetinio saugumo užtikrinimui. Yampolskiy ir Spellchecker (2016) išsakė mintį, kad visiškai autonominės mašinos niekada negali būti laikomos savaime saugiomis, geriausiu atveju galėtume pasiekti tik tikimybinį įrodymą, kad sistema atitinka tam tikrą fiksuotų apribojimų rinkinį, tačiau tai nereikštų, kad ji yra saugi esant neribotam įvesties duomenų rinkiniui. Shanbhag ir kt. (2024) taip pat pastebi, jog didelis klaidingų teigiamų rezultatų skaičius ir sudėtingumas interpretuojant dirbtinio intelekto sprendimus yra vieni pagrindinių jo naudojimo iššūkių. Kibernetiniai nusikaltėliai gali kurti metodus, skirtus manipuliuoti dirbtinio intelekto sistemomis, dėl ko gali atsirasti klaidingai teigiamų arba klaidingai neigiamų rezultatų (Adewusi, Okoli, Olorunsogo, Adaga, Daraojimba ir Obi, 2023). Klaidingi įspėjimai gali mažinti pasitikėjimą įsibrovimų aptikimo sistemomis ir nukreipti išteklius nuo realių grėsmių sprendimo. Be to, kai kurių dirbtinio intelekto modelių „juodosios dėžės“ pobūdis apsunkina saugumo analitikų galimybes suprasti, kodėl tam tikra veikla buvo nustatyta kaip kenkėjiška (Shanbhag ir kt., 2024). Michael ir kt. (2023) taip pat pažymi, kad dirbtinis intelektas nėra neklystantis. Jis yra linkęs į tas pačias atakas kaip ir kitos kompiuterinės sistemos, o netikėti pokyčiai gali padaryti jas neveiksmingas, dažnai suteikdami saugumo analitikams klaidingą saugumo jausmą. Apibendrinus šias mintis, galima daryti išvadą, jog nepaisant to, kad dirbtinio intelekto sprendimai pasižymi

savo technologiniu pažangumu, taip pat yra kur kas pranašesnės už įprastas technines saugumo priemones, vis dėl to jos negali garantuoti visiško saugumo atsižvelgiant į taip pat sparčiai besivystančias grėsmes.

Polito ir Pupillo (2024) nurodo, jog dirbtinio intelekto sistemų saugumo kontekste ypač svarbu atskirti patikimumo *(angl. reliability)* ir patikėtumo *(angl. trustworthiness)* sąvokas. Šie autoriai taip pat pabrėžia, kad siekiant užtikrinti saugų dirbtinio intelekto veikimą, būtina garantuoti sistemos atsparumą, kad ji veiktų numatytu būdu net ir tada, kai įvesties duomenys ar pats modelis yra paveikiami galimų atakų, tačiau kartu pažymi, jog tokio atsparumo vertinimas yra sudėtingas, nes reikėtų išbandyti sistemą su labai dideliu galimų įvesties sutrikimų skaičiumi, o dėl daugybės galimų trikdžių realiose taikymo situacijose išsamus dirbtinio intelekto sistemų testavimas tampa praktiškai neįmanomas. Atsižvelgiant į tai, pilnas dirbtinio intelekto sistemų patikrinimas ir jų veikimo patvirtinimas visais galimais scenarijais yra labai sudėtingas, o tai riboja galimybę visiškai pasitikėti tokiomis sistemomis (Polito ir Pupillo, 2024). Panašios nuomonės laikosi ir Lazic (2019) pažymėdamas, kad organizacijoms įdiegus dirbtinio intelekto sprendimus kibernetiniam saugumui, gali atsirasti darbuotojų pernelyg didelio pasitikėjimo technologijomis ir sumažėjusio budrumo rizika. Autorės manymu, tokiu atveju bus susiduriama su nuolatiniu iššūkiu, jog dirbtinio intelekto technologija negalima būtų besąlygiškai pasitikėti ir pilnai jai patikėti organizacijos kibernetinį saugumą. Iš to susidaro išvada, kad prieš diegiant tokias technologijas, reikėtų skirti papildomą dėmesį ir organizacinėms saugumo priemonėms, kad minėta rizika būtų valdoma.

Polito ir Pupillo (2024) išskiria žmogaus ir dirbtinio intelekto sąveiką, kaip svarbų iššūkį naudojant dirbtinio intelekto priemones, kurį vadina pusiau automatizavimo problema. Pusiau automatizuotų sistemų kūrime svarbu užtikrinti tinkamą jų projektavimą ir naudotojų mokymą, siekiant išlaikyti nuolatinį žmogaus dėmesį bei aktyvų dalyvavimą sistemos veikime. Pasak juos, kai operatoriai mano, kad dirbtinis intelektas yra labiau pajėgus ar autonomiškas, nei yra iš tikrųjų, jie gali pernelyg juo pasikliauti ir tapti mažiau budrūs. Toks per didelis pasitikėjimas gali sumažinti žmogaus įsitraukimą ir padidinti klaidų tikimybę, ypač situacijose, kuriose vis dar būtinas žmogaus įsikišimas (Polito ir Pupillo, 2024). Šią riziką gerai iliustruoja ir konkretūs praktiniai atvejai. Remiantis viešai skelbta informacija, JAV Kibernetinio saugumo ir infrastruktūros saugumo agentūros vadovas įkėlė tik tarnybiniam naudojimui pažymėtus dokumentus į viešai prieinamą dirbtinio intelekto platformą. Tokių duomenų perdavimas į išorinę sistemą sukėlė saugumo įspėjimus ir inicijavo vidinį tyrimą dėl galimo duomenų nutekėjimo (Politico, 2026). Autorė iš pateiktos informacijos sprendžia, kad atitinkamos organizacijos saugumas, be kita ko, priklauso ir nuo jos žmogiškųjų išteklių bei jų sąmoningumo lygio. Šiuos autorius samprotavimas patvirtina ir Lazic (2019), atkreipdamas dėmesį, jog naudojant pažangias dirbtinio intelekto sistemas, būtina išlaikyti aukštą kibernetinio saugumo sąmoningumo lygį bei nuolat ugdyti darbuotojų saugumo kompetencijas. Vis dėl to, dirbtinio intelekto ir žmogaus sąveika pasireiškia ne tik organizacijos darbuotojų sąmoningumu, kas yra neabejotinai svarbu. Kaip teigia Jabbarova (2023), dirbtiniu intelektu pagrįstos kibernetinio saugumo sistemos gali gerokai padidinti saugumo priemonių efektyvumą ir veiksmingumą, tačiau jos neturėtų veikti

atskirai. Ši autorė pažymi, jog užtikrinant tikslų ir veiksmingą dirbtiniu intelektu pagrįstų kibernetinio saugumo sistemų veikimą, akivaizdus nepakeičiamas žmogaus priežiūros ir įsikišimo vaidmuo (Kamila Jabbarova, 2023). Manytina, kad būtent tokių dviejų veiksnių sąsaja, kaip a) darbuotojų sąmoningumas ir tinkamas apmokymas naudojant dirbtinio intelekto sistemas ir b) darbuotojų įtraukimas prižiūrėti dirbtinio intelekto atliekamus veiksmus, gali padėti valdyti iššūkį dėl pernelyg didelio pasitikėjimo dirbtinio intelekto sistemomis ir šią riziką valdyti.

Atskira rizikų grupė susijusi su tuo, kad pats dirbtinis intelektas gali tapti pažeidžiamas arba būti panaudotas kenkėjiškiems tikslams. Kaip jau buvo nustatyta aukščiau, dirbtinio intelekto priemonės kibernetiniame saugume pasireiškia ne tik per gynybos prizmę, bet ir per atakų. Lazic (2019) pažymi, kad nors dirbtinis intelektas stiprina kibernetinį saugumą, jo taikymas taip pat gali sukurti naujų pažeidžiamumų. Jis šią riziką sieja su tuo, kad nauja kibernetinių atakų banga keičia pasaulinę grėsmių aplinką, nes komerciškai prieinami dirbtinio intelekto įrankiai leidžia net ir santykinai nepatyrusiems įsilaužėliams vykdyti didelio masto įsilaužimus, kurie anksčiau buvo skirti naudoti sudėtingiems kibernetinių nusikaltimų grupuotėms. Tokios rizikos apčiuopiamumą rodo ir praktiniai pavyzdžiai. Remiantis „Amazon Threat Intelligence“ paskelbtomis išvadomis, finansiškai motyvuotas, rusakalbis grėsmių veikėjas, pasinaudodamas keliomis generatyvinio dirbtinio intelekto paslaugomis per daugiau nei mėnesį laiko įsilaužė į daugiau nei 600 „FortiGate“ tinklo įrenginių mažiausiai 55 šalyse (Moses, 2026). Autorės vertinimu šis pavyzdys puikiai iliustruoja tai, kad jeigu nepatyrę asmenys pasinaudodami viešai prieinamomis priemonėmis gali sukelti tokią žalą, rizika dėl pažangių dirbtinių intelekto sistemų pažeidžiamumų yra labai reali, ir tai gali tapti reikšmingu iššūkiu kuriant tokioms grėsmėms atsparias sistemas.

Lazic (2019) išskiria dar vieną atvejį, kai dirbtinis intelektas gali būti panaudojamas kenkėjiškiems tikslams naudojant priešiško dirbtinio intelekto *(angl. Adversarial AI)* priemones. Kaip nurodo pats Lazic (2019), tokiais atvejais užpuolikai sąmoningai modifikuoja į sistemas pateikiamus duomenis ar signalus taip, kad mašininio mokymosi modeliai juos interpretuotų neteisingai, ir dėl tokių manipuliacijų dirbtinio intelekto sistemos gali klaidingai klasifikuoti objektus, neatpažinti svarbių elementų arba priimti sprendimus, kurie yra palankūs užpuolikui. Siekiant labiau suprasti kaip tai atrodo praktikoje, Lazic (2019) kartu pateikia ir tokių galimų tokių atakų pavyzdžius, kai manipuliuojant vaizdiniais duomenimis autonominės transporto priemonės gali neatpažinti eismo ženklų ar šviesoforų signalų, o veidų atpažinimo sistemose specialiai sukurtos vaizdo ar vaizdinės manipuliacijos gali leisti apeiti autentifikavimo mechanizmus. Pasak jį, tokios situacijos rodo, kad dirbtinio intelekto sistemos, nepaisant jų pažangumo, gali būti pažeidžiamos manipuliacijų, jei nėra įdiegtos tinkamos apsaugos priemonės. Polito ir Pupillo (2024) taip pat pabrėžia, kad dirbtinio intelekto sistemas galima manipuliuoti įvairiais būdais, iš kurių dažniausiai minimos įvesties *(angl. input)* ir duomenų nuodijimo *(angl. poisoning)* atakos. Kartu paaiškindami, kad įvesties atakos nukreiptos į duomenis, pateikiamus mašininio mokymosi sistemai, kai

užpuolikas sąmoningai pakeičia ar modifikuoja įvesties duomenis taip, kad sistema juos interpretuotų neteisingai. Tuo tarpu duomenų nuodijimo atakos vyksta sistemos kūrimo ir mokymo etape, kai užpuolikas manipuliuoja mokymo duomenimis ar mokymo procesu, dėl ko sukuriamas iš esmės neteisingai veikiantis modelis (Polito ir Pupillo, 2024). Autorės vertinimu, būtent pastaroji rizika yra ypač aktuali nagrinėjamos temos požiūriu, nes parodo, kad pažeidžiamumai gali atsirasti dar dirbtinio intelekto sistemos kūrimo metu. Atsižvelgiant į tai manytina, jog tai vienas svarbiausių argumentų, kodėl kibernetinio saugumo aspektai turi būti integruojami į dirbtinio intelekto sistemas nuo pat jų kūrimo pradžios.

Be visų techninių ir organizacinių iššūkių, dar vienas aspektas, kuri reikalinga turėti omenyje svarstant apie dirbtinio intelekto sprendimų naudojimą kibernetiniame saugume yra ekonominis aspektas. Ljubomir Lazic (2019), kaip vieną pagrindinių trūkumų išskiria didesnes finansines ir technines sąnaudas palyginti su tradiciniais kibernetinio saugumo sprendimais. Pasak jį, dirbtinio intelekto pagrindu veikiančių sistemų kūrimas ir diegimas reikalauja pažangių technologijų bei didesnių investicijų, todėl tokie sprendimai ilgą laiką buvo sunkiau prieinami mažoms ir vidutinėms įmonėms. Šio darbo autorės manymu, tai yra labai reikšmingas aspektas, nes pažangios saugumo priemonės iš tiesų gali pareikalauti pernelyg didelio biudžeto tokių priemonių diegimui, kurį ne visos organizacijos turi. Akhtar bei Rawol (2024) papildomai pabrėžia, jog atitiktis pasauliniams standartams, tokiems kaip SOC2 ir ISO 27001, didina saugumą ir skatina kibernetinio saugumo prioretizavimo kultūrą. Jie taip pat pažymi, kad nuolatinis darbuotojų švietimas ir mokymas apie geriausią praktiką ir kylančias grėsmes yra gyvybiškai svarbūs atsparios kibernetinio saugumo sistemos komponentai. O tai, taip pat gali prisidėti prie finansinių ir organizacinių iššūkiu, kadangi tam reikia skirti atitinkamą finansavimą bei išteklius. Jų manymu vienas didžiausių sunkumų taikant dirbtinio intelekto informacijos apsaugai yra tai, kad tam paprastai reikia daugiau laiko ir lėšų nei tradiciniams, dirbtinio intelekto nenaudojantiems kompiuterinės apsaugos sprendimams, be to, tokios technologijos nėra pigios (Akhtar ir Rawol, 2024). Mažėjant pažangių technologijų kūrimo ir diegimo kaštams, tokios priemonės tampa vis prieinamesnės, todėl didėja rizika, kad jos bus panaudotos kenkėjiškais tikslais (Lazic, 2019). Vis dėlto šiuo metu atsiranda naujų SaaS tipo, t.y. saugumo kaip paslaugos, technologijų, kurios padeda dirbtinio intelekto pagrįstas kibernetinės gynybos technologijas padaryti ekonomiškai prieinamesnes verslui (Das ir Sandhane, 2021). Be to, investicijos į efektyvų kibernetinį saugumą, pasak Lazic (2019) dažnai yra mažesnės nei nuostoliai, patiriami dėl sėkmingų kibernetinių atakų. Autorė mano, kad didelės finansinės sąnaudos dėl pažangių priemonių diegimo iš tiesų kelia didelį iššūkį, ypač mažoms ir vidutinėms įmonėms, vis dėl to, kartu reikalinga vertinti ir tai, kokios neigiamos ekonominės pasekmės gali kilti tai pačiai organizacijai dėl tokių priemonių nebuvimo. Atkreiptinas dėmesys, kad šias rizikas valdyti rinkoje egzistuoja ir mažiau kainuojančios priemonės, tačiau šiuo atveju, autorės manymu, reikėtų kelti klausimus, ar tokios priemonės iš tiesų būtų pakankamos užtikrinti tinkamą kibernetinio saugumo lygį organizacijoje.

Dar viena svarbi rizikų grupė, susijusi su dirbtinio intelekto sistemų kūrimu bei taikymu, yra etinio bei socialinio pobūdžio klausimai. ISO (n.d.) nurodo, jog dirbtinis intelektas, gebantis per kelias sekundes sintetinti, analizuoti ir reaguoti į milžiniškus duomenų kiekius, yra nepaprastai galingas, tačiau kaip ir bet kurios galingos technologijos atveju, labai svarbu ją įdiegti atsakingai, kad maksimaliai būtų išnaudotas jos potencialas ir būtų sumažintas galimas neigiamas poveikis. Pasak Adewusi ir kt. (2023), dirbtinio intelekto algoritmai gali įtvirtinti esamus šališkumus, o tai gali lemti diskriminacinius rezultatus kibernetinių grėsmių aptikimo ir mažinimo procesuose. Praktiniu atžvilgiu tai atrodo taip: jeigu apmokytas, naudojant nepatikrintus duomenis, dirbtinis intelektas gali atkartoti žalingus išankstinius nusistatymus dėl rasės, religijos, auklėjimo ar kitų žmogaus savybių, gali kilti rimtas iššūkis dėl asmenų diskriminacijos (ISO, n.d.). Autorės manymu, dirbtinio intelekto sistemų kūrimo etape labai svarbu įtraukti papildomus saugiklius dėl galimos asmenų diskriminacijos ir tinkamai apmokyti tokios sistemos modelius, kad vėliau tai netaptų rimtu organizacijos iššūkiu dėl galimos asmenų diskriminacijos įvairiais aspektais.

Kitas svarbus etinis susirūpinimas, susijęs su dirbtiniu intelektu, yra privatumas. Dirbtinio intelekto sistemoms renkant didžiulius duomenų kiekius iš duomenų bazių visame pasaulyje, reikia užtikrinti, kad asmeninė informacija būtų apsaugota ir naudojama atsakingai (ISO, n.d.). Adewusi ir kt. (2023) taip pat pabrėžia, kad dirbtinio intelekto sistemos remiasi didžiuliais duomenų kiekiais, todėl kyla susirūpinimas dėl duomenų privatumo ir saugumo. Pavyzdžiui, veido atpažinimo technologija, dažnai naudojama apsaugos sistemose ar socialinės žiniasklaidos platformose, kelia klausimų dėl sutikimo ir galimo netinkamo naudojimo (ISO, 2026). Manytina, kad šiuo atveju rizika pasireiškia dėl pačių asmenų duomenų naudojimo tokiose pažangiose technologijose, kaip biometrinės ar kitu dirbtinio intelekto pagrindu veikiančios saugumo priemonės. Kaip nurodo Štitilis, Laurinaitis ir Verenius (2023), biometrinių technologijų, kurios padeda unikaliai autentifikuoti vidinius darbuotojus, turinčius prieigą prie ypatingos svarbos infrastruktūros kibernetinio saugumo tikslais, naudojimas turi būti tobulinamas atsižvelgiant į galiojančius reguliavimus, pagrįstus Bendruoju duomenų apsaugos reglamentu (toliau – BDAR). Pasak juos, BDAR numato bendruosius biometrinių duomenų, kaip specialios kategorijos asmens duomenų, tvarkymo pagrindus, tačiau kartu tai reiškia, kad saugant ypatingos svarbos informacinę infrastruktūrą šie duomenys gali būti tvarkomi tik remiantis darbuotojo sutikimu, o dėl darbdavio ir darbuotojo padėties disbalanso, t. y. galios pusiausvyros nebuvimo, toks sutikimas greičiausiai būtų pripažintas nesavanorišku, nebent darbdavys suteiktų alternatyvias sistemas. Tuo tarpu suteikus alternatyvius apsaugos būdus, nebūtų įmanoma užtikrinti vien tik biometrinių technologijų naudojimo infrastruktūros darbuotojų identifikavimui ir autentifikavimui (Štitilis ir kt., 2023). Autorės nuomone, šiuo atveju iššūkiu gali tapti ne pačių pažangių priemonių prieinamumas, bet ir tokių priemonių įteisinimas pagal taikomus teisinius reikalavimus priklausomai nuo atitinkamos valstybės reguliavimo.

Greta su tuo, yra išskirtinas ir atsakomybės klausimas, t.y. gali kilti iššūkis nustatyti, kas yra atsakingas už dirbtinio intelekto valdomo prietaiso ar paslaugos padarytą žalą. Europos Parlamentas (2021)

kelia klausimą, kas bus atsakingas įvykus avarijai, susijusiai su savaime vairuojančiu automobiliu, t.y. ar žalą turėtų padengti savininkas, automobilio gamintojas ar programuotojas? Ji pabrėžia, kad jeigu gamintojas turėtų teisę neprisiimti atsakomybės, tokiu atveju neliktų ambicijų kurti kokybiškus produktus ar teikti kokybiškas paslaugas, taip mažinant pasitikėjimą technologijomis, tačiau per didelis reguliavimas galėtų stabdyti inovacijų plėtrą (Europos Parlamentas, 2021). Autonominės transporto priemonės ir kitos su dirbtiniu intelektu susijusios sistemos gali sudaryti sąlygas saugumo pažeidimams, kibernetinėms atakoms ar netinkamam asmens duomenų naudojimui, ypač kai tai susiję su didelio duomenų kiekio rinkimu ir tvarkymu (Kritikos, 2019). Be to, Europos Parlamentas (2021) atkreipia dėmesį, kad dirbtinio intelekto programos, fiziškai susijusios su žmonėmis arba integruotos į jų kūnus, gali būti blogai parengtos ar neteisėtai naudojamos, taip sukeldamos tiesioginę riziką žmonėms, o blogai reguliuojamas dirbtinis intelektas ginklų naudojime gali tapti apskritai nesuvaldomas. Adewusi ir kt. (2023) taip pat pažymi, jog dirbtinio intelekto modeliai gali būti sudėtingi ir nepermatomi, todėl tampa sunkiau užtikrinti atskaitomybę ir pasitikėjimą jų priimamais sprendimais. Šiuo atveju darbo autorė daro išvadą, kad atsakomybės klausimas taip pat yra labai svarbus ir kuriant tokius dirbtinio intelekto sprendimus, kurie bus taikomi užtikrinant kibernetinį atsparumą, kadangi tokios atsakomybės nebuvimas gali reikšmingai paveikti apskritai tokių priemonių patikimumą.

Vis dėl to, kaip jau buvo nustatyta aukščiau, dirbtinio intelekto priemonių taikymas kibernetinio saugumo kontekste pamažu tampa būtina saugumo priemonė. Atitinkamai, organizacijai, kuri kurtų ar diegtų tokias priemones neabejotinai teks susidurti su aukščiau aprašytais iššūkiais bei rizikomis. Šią autorės mintį patvirtina ir Akhtar bei Rawol (2024) nurodydami, kad mūsų kasdienis gyvenimas vis labiau susipina su mašinų valdomais procesais ir programuojamomis instrukcijomis, kurias skatina duomenų srautas, todėl nors ši pažanga suteikia lengvą prieigą ir komfortą, jei ji nebus kontroliuojama, ji gali tapti grėsme mūsų egzistencijai. Atsižvelgiant į tai, tinkamos gairės ir ekspertų priežiūra yra būtini norint išlaikyti kontrolę ir pusiausvyrą.

Apibendrinant galima teigti, kad dirbtinio intelekto sistemų kūrimas kibernetinio saugumo srityje yra susijęs su plačiu iššūkių ir rizikų spektru. Šios rizikos apima technologinius iššūkius, kurie pasireiškia, tuo, jog dirbtinio intelekto sistemos saugumu negalima besąlygšikai pasitikėti dėl sparčiai besivystančių kibernetinių atakų ir tai, kad tos pačios sistemos gali būti naudojamos kenkėjiškiems tikslams. Taip pat rizikos pasireiškia dėl pernelyg didelio pasitikėjimo technologijomis ir iššūkiu sukurti tinkamą žmogaus ir dirbtinio intelekto sąveiką. Naudojant šias technologijas kyla ir etiniai ir socialiniai klausimai, kaip šališkumas, privatumas ir reguliacinės problemos, susijusios su atsakomybe ir atitinkamų procesų įteisinimu. Taip pat ne mažiau svarbus yra ir praktinis organizacijų iššūkis, susijęs su finansiniais kaštais, infrastruktūra ir darbuotojų tinkamu apmokymu tiek naudojant tokias sistemas, tiek jas prižiūrint. Galima daryti išvadą, jog visa tai rodo, kad dirbtinio intelekto sistemų kūrimas negali būti grindžiamas vien technologiniu požiūriu sukurti maksimaliai kibernetinėms grėsmėms atsparią sistemą. Kartu būtina

užtikrinti, kad šios sistemos būtų kuriamos atsakingai, saugiai, skaidriai ir taip, kad jų veikimas išliktų valdomas tiek techniniu, tiek organizaciniu, tiek teisiniu požiūriu.

1.4. Kibernetinio saugumo reikalavimai kuriant dirbtinio intelekto sistemas

Praeituose poskyriuose aprašytos dirbtinio intelekto taikymo galimybės ir su tuo ateinančios rizikos bei iššūkiai paskatino šią sritį sureguliuoti. Skirtingose valstybėse bei regionuose šie reikalavimai reguliuojami nevienodai, t.y. vienur egzistuoja griežtos privalomos taisyklės, kitur tai gali būti tik rekomendacinio pobūdžio gairės. Siekdami apžvelgti šiuos skirtumus, toliau bus nagrinėjami atskirų regionų bendrieji teisiniai reikalavimai ir (ar) rekomendacijos taikomi kuriant dirbtinio intelekto sistemas bei kartu užtikrinant kibernetinį saugumą.

Kalbant apie dirbtinio intelekto reguliavimą Europos Sąjungoje, autorės manymu, šiuo atveju aktualiausias teisės aktas būtų Reglamentas (ES) 2024/1689, kitaip vadinamas Dirbtinio intelekto aktas (2024). Kaip nurodo Europos Komisija (2026), šis aktas yra „pirmoji dirbtinio intelekto teisinė sistema, kuria sprendžiama dirbtinio intelekto keliama rizika ir užtikrinama, kad Europa atliktų vadovaujamą vaidmenį pasaulyje“. Tuo tarpu kalbant bendrai apie kibernetinio saugumo reguliavimą Europos lygmeniu, jį šiuo metu apibrėžia Direktyva (ES) 2022/2555, dar žinoma kaip TIS2 (2022), kuria siekiama užtikrinti aukštą bendrą kibernetinio saugumo lygį svarbiuose sektoriuose visoje Europos Sąjungoje (Europos Komisija, 2026). Autorė papildomai atkreipia dėmesį, kad direktyvos reikalavimai, skirtingai nuo reglamento, nėra tiesiogiai taikomi, o jų įgyvendinimas turi būti perkeliamas į kiekvienos valstybės narės nacionalinius teisės aktus. Pavyzdžiui, Lietuvoje TIS2 direktyvos reikalavimai yra perkelti į LR Kibernetinio saugumo įstatymą (2014). Šio darbo autorės manymu, vertinant dirbtinio intelekto reguliavimo kibernetinio saugumo kontekste Europos lygmeniu, pakanka remtis Dirbtinio intelekto akto bei TIS2 direktyvos nuostatomis, nes šie aktai atspindi bendrus ir vienodai taikomus reikalavimus visame Europos Sąjungos kontekste.

Kaip nurodo Europos Komisija (2026) „Dirbtinio intelekto aktu užtikrinama, kad europiečiai galėtų pasitikėti tuo, ką gali pasiūlyti dirbtinis intelektas“ ir pabrėžia, jog Dirbtinio intelekto aktu nustatytas reguliavimas Europos Sąjungoje remiasi rizika grindžiamu požiūriu. Ryšių reguliavimo tarnyba (toliau – RRT) (2026) taip pat pabrėžia, jog „Dirbtinio intelekto aktu numatytas reguliavimas pagrįstas rizikos hierarchijos principu, kuris nustato skirtingus reikalavimus priklausomai nuo dirbtinio intelekto sistemos keliamos rizikos visuomenei“ (žr. 1 pav.).

1 pav. Rizika grįstas požiūris pagal Dirbtinio intelekto aktą.

Rizikos lygis		Apibūdinimas
	Nepriimtina rizika	DI sistemos, laikomos prieštaraujančiomis pagrindinėms teisėms (pvz., manipuliacinės ar socialinio reitingavimo sistemos) – **draudžiamos**
	Didelė rizika	DI sistemos, naudojamos svarbiose srityse (pvz., švietimas, įdarbinimas, teisėsauga, sveikata) – **taikomi griežti atitikties reikalavimai**
	Ribota rizika	Reikalaujamos **skaidrumo pareigos** (pvz., informuoti naudotojus apie tai, kad jie sąveikauja su DI sistema)
	Minimali rizika	Laisvai naudojamos sistemos (pvz., rekomendacijų algoritmai), joms **taikoma savireguliacija** arba gerosios praktikos gairės

Šaltinis: Ryšių reguliavimo tarnyba (n.d.).

Remiantis RRT (n.d.) 1 pav. pateiktu rizika pagrįstu požiūriu, galima matyti, kad Europos Sąjungoje atitinkami dirbtinio intelekto reikalavimai priklauso nuo to, kokią riziką sukuria atitinkama dirbtinio intelekto sistema. Pavyzdžiui, dirbtinio intelekto sistemos, kurios yra laikomos prieštaraujančiomis pagrindinėms teisėms ir laisvėms apskritai yra draudžiamos Europos Sąjungoje. Tuo tarpu dirbtinio intelekto sistemoms, kurios naudojamos tokiose svarbiose srityse, kaip švietimas, įdarbinimas, teisėsauga bei sveikata, turi būti taikomi griežti atitikties reikalavimai. Ribotos rizikos dirbtinio intelekto priemonėms reguliavimas yra kiek švelnesnis ir numatomas reikalavimas dėl skaidrumo, t.y. tokiu atveju yra būtinas aiškus naudotojų informavimas apie tai, kad jie sąveikauja su tokia sistema. O minimalios rizikos sistemoms, kaip antai rekomendacijų algoritmai, nėra taikomi griežti reikalavimai, o apsiribojama gerosiomis praktikos gairėmis (RRT, n.d.). Šio darbo autorės manymu, toks požiūris yra ganėtinai lankstus dirbtinio intelekto taikymo prasme, visgi, tai gali sukelti ir papildomų klausimų nustatant atitinkamos sistemos taikomą riziką bei pritaikant pagal ją tinkamus teisinius reikalavimus.

Dirbtinio intelekto akte (2024) konstatuojamoje dalyje akcentuojama, jog „kibernetinis saugumas atlieka esminį vaidmenį užtikrinant, kad dirbtinio intelekto sistemos būtų atsparios bandymams pakeisti jų naudojimą, elgseną, veikimo efektyvumą arba pažeisti jų saugumo funkcijas, kai trečiosios šalys imasi kenkėjiškų veiksmų sistemos pažeidžiamumui išnaudoti”, taip pat papildomai nurodoma, jog turi būti imtasi tinkamų priemonių užtikrinti tokį kibernetinio saugumo lygį, kuris atitiktų riziką. Vis dėl to, darbo autorė pastebi, kad konkrečių techninių ar organizacinių priemonių būtent dirbtinio intelekto sistemų kūrimui Dirbtinio intelekto aktas nepateikia, o labiau nustato bendrąją pareigą užtikrinti tokį kibernetinio saugumo lygį, kuris atitiktų sistemos keliamą riziką, o techninių ir organizacinių priemonių konkretizavimą palieka standartizacijai, praktinėms gairėms ir rizikos vertinimui konkrečiu atveju.

Kaip minėta, konkrečių reikalavimų dirbtinio intelekto sistemų kūrimui kibernetinio saugumo kontekste reikėtų ieškoti gerosios praktikos šaltiniuose. Pavyzdžiui, Prancūzijos duomenų apsaugos institucija CNIL pabrėžia, kad kuriant dirbtinio intelekto sistemą, turi būti užtikrintas: a) duomenų konfidencialumas: labai svarbu atsižvelgti į duomenų privatumo riziką; b) sistemos veikimas ir

vientisumas: reikia skirti daugiau dėmesio našumo palaikymui laikui bėgant ir jų poveikio žmonėms stebėjimui; c) visos informacinės sistemos saugumas, įskaitant jos komponentus, pavyzdžiui atsarginės kopijos (CNIL, 2026). Prancūzijos duomenų apsaugos institucija pateikia ir labai aiškias gaires, kokių priemonių, susijusių su sistemos kūrimu, turi būti laikomasi: a) asmens duomenų apsaugos įvertinimas renkantis sistemos projektavimo sprendimus (pvz., BDAR principų užtikrinimas ir kt.), b) gerųjų saugumo praktikų laikymasis atitinkamoje srityje, c) nuolatinio kūrimo ir integravimo procedūros įdiegimas, d) išsamios dokumentacijos rengimas, e) saugumo auditų atlikimas (CNIL, 2026). Autorės manymu, toks požiūris leidžia nustatyti aiškų veiksmų planą į kokius akcentus turi būti atkreipiamas dėmesys siekiant užtikrinti asmens duomenų apsaugos bei kibernetinio saugumo reikalavimus, kad kuriama dirbtinio intelekto sistema atitiktų teisinį reglamentavimą.

Europos Komisijos Jungtinio tyrimų centro parengtoje ataskaitoje paaiškinama, kad kibernetinio saugumo reikalavimai turi būti taikomi ne vien atitinkamam dirbtinio intelekto modeliui, bet visai dirbtinio intelekto sistemai, nes ją sudaro ir kiti komponentai, tokie kaip sąsajos, duomenų bazės, tinklo ryšio elementai ar stebėsenos priemonės ir kt. (Junklewitz ir kt., 2023). Toje pačioje ataskaitoje išskiriami keturi pagrindiniai elementai kaip turi būti kuriamos dirbtinio intelekto sistemos: 1) dirbtinio intelekto sistemos turi būti suprojektuotos taip, kad būtų atsparios kenkėjiškų trečių šalių grėsmėms; 2) turi būti įgyvendinti organizaciniai ir techniniai saugumo sprendimai; 3) didelės rizikos dirbtinio intelekto sistemoms turi būti atliktas kibernetinio saugumo rizikos vertinimas; 4) techniniai sprendimai turi būti parinkti atsižvelgiant į konkrečias rizikas ir gerąsias praktikas (Junklewitz ir kt., 2023). Toks požiūris taip pat leidžia aiškiai suprasti kokie techniniai bei organizaciniai komponentai turi būti įtraukti kuriant atitinkamą dirbtinio intelekto sistemą.

Kalbant konkrečiai apie sistemų kūrimą kibernetinio saugumo kontekste, prieš tai aptartas bendras dirbtinio intelekto kūrimo gaires reikėtų kartu derinti su techniniais kibernetinio saugumo reikalavimais. Kaip minėta, kibernetinio saugumo reikalavimus šiuo metu numato TIS2 Direktyva, kuri kiekvienos Europos Sąjungos valstybėje yra įgyvendinama nacionaliniu teisės aktu. Atsižvelgiant į tai, kad TIS2 Direktyva iš esmės nurodo vieningą reguliacinę kryptį visose Europos Sąjungos šalyse, kaip pavyzdį galima paimti Lietuvos kibernetinio saugumo reguliavimą. Lietuvos nacionaliniu lygmeniu jis reglamentuojamas LR kibernetinio saugumo įstatymu bei jį įgyvendinančiu Nutarimu Nr. 818, kur pateikiamos konkrečios techninės bei organizacinės saugumo priemonės, tačiau būtent dirbtinio intelekto sistemų kūrimui skirtų priemonių šiuose teisės aktuose nėra detalizuojami. Tuo tarpu Nacionalinis kibernetinio saugumo centras (toliau – NKSC) savo rekomendacijose saugiam dirbtinio intelekto sprendimų naudojimui organizacijoje (2025) numato konkrečias saugumo kontrolės priemones, kurias atitinkama organizacija turėtų įdiegti, kaip antai: a) „taikyti griežtą prieigos kontrolę ir įjungti kelių veiksnių autentikavimą *(angl. Multi Factor Authentication arba MFA)*; b) užtikrinti, kad visos dirbtinio intelekto sistemos programavimo sąsajos *(angl. API)* būtų žinomos, valdomos ir apsaugotos; c) naudoti duomenų nutekėjimo prevencijos *(angl. Data Loss*

Prevention arba DLP) priemones, kurios geba identifikuoti į dirbtinio intelekto sistemą keliamą informaciją, įskaitant įvedamo turinio filtravimo ir auditavimo mechanizmus; išjungti modelio tobulinimo jūsų organizacijos įvedamais duomenimis funkciją; d) esant galimybei dirbtinio intelekto sistemą įtraukti į reguliarius saugumo auditus; e) užtikrinti reguliarų įdiegtų dirbtinio intelekto sistemų atnaujinimą ir saugumo pataisų diegimą; f) numatyti saugias dirbtinio intelekto sistemos pašalinimo galimybes; g) įsitikinti, kad duomenys yra šifruojami, tiek saugojimo, tiek perdavimo į dirbtinio intelekto sistemą metu (pvz., naudojant TLS 1.3 protokolą)". Papildomai NKSC (2025) nurodo ir darbuotojų apmokymą, kaip dirbtinio intelekto sprendimų naudojimui taikoma kibernetinio saugumo kontrolės priemonė. Autorės vertinimu, šiomis praktinėmis gairėmis jau galima aiškiai remtis kuriant dirbtinio intelekto sistemas, norint atitikti kibernetinio saugumo reikalavimams.

Žvelgiant plačiau ir nagrinėjant tokius reikalavimus pasauliniu mastu, kaip pavyzdį galima paimti tarptautines gaires. Jungtinės Karalystės Nacionalinis kibernetinio saugumo centras *(angl. NCSC)* ir JAV Kibernetinio saugumo ir infrastruktūros saugumo agentūra *(angl. CISA)* kartu su tarptautiniais partneriais iš įvairių pasaulio valstybių nacionalinių kibernetinio saugumo centrų, saugumo agentūrų ir kt. (įskaitant, bet neapsiribojant Kanados, Čilės, Baltijos šalių, Skandinavijos, Korėjos, Singapūro, Nigerijos ir kt.) paskelbė Saugaus dirbtinio intelekto kūrimo gaires (2023). Šiose gairėse pateikiamos apibendrintos rekomendacijos, taikomos dirbtinio intelekto sistemos kūrimo etapui, įskaitant tiekimo grandinės saugumą, dokumentaciją, taip pat išteklių ir techninės skolos valdymą. Rekomendacijas apima 4 žingsniai: 1) apsaugoti savo tiekimo grandinę: jeigu kuriamos sistemos komponentai nėra kuriami organizacijos viduje, reikia reikalauti, kad pasitelkti tiekėjai laikytųsi tokių pačių saugumo standartų, kokių laikosi ir pati organizacija; 2) nustatyti, sekti ir apsaugoti savo organizacijos išteklius: reikalinga turėti procesus ir kontroles priemones, skirtas valdyti, prie kokių duomenų gali prieiti dirbtinio intelekto sistemos, ir valdyti dirbtinio intelekto sugeneruotą turinį pagal jo jautrumą ir pagal įvesties duomenų, naudotų jam sugeneruoti, jautrumą; 3) dokumentuoti savo duomenis, modelius bei užklausas: tai apima dokumentavimą visų modelių, duomenų rinkinių ir meta arba sisteminių užklausų kūrimą, veikimą ir gyvavimo ciklo valdymą; ir galiausiai 4) valdyti savo technines spragas: sprendimai, kurie priimami siekiant trumpalaikių rezultatų, bet neatitinkantys gerųjų praktikų, laikomos techninės spragos, jas reikalinga nustatyti, sekti bei valdyti per visą dirbtinio intelekto sistemos gyvavimo ciklą (Saugaus dirbtinio intelekto kūrimo gaires, 2023). Autorės manymu, šios techninės ir organizacinės priemonės iš esmės papildo prieš tai aptartas Lietuvos NKSC pateiktas gaires dėl saugių dirbtinio intelekto priemonių kūrimo ir naudojimo kibernetinio saugumo atžvilgiu, todėl manytina, kad gali būti taikomos kartu.

White & Case (2025) duomenimis šiuo metu JAV nėra išsamių federalinių teisės aktų ar reglamentų, kurie reguliuotų dirbtinio intelekto kūrimą. Vis dėl to, kaip buvo išsiaiškinta aukščiau, JAV laikomasi jau anksčiau minėtų CISA gairių, taip pat NIST Dirbtinio intelekto rizikos valdymo sistemos (AI RMF 1.0) gairių, kuriose skiriamas dėmesys būdams, kaip organizacijos gali apsaugoti savo dirbtinio intelekto

sistemas, gintis nuo kibernetinių atakų naudojant dirbtinio intelekto kibernetinio saugumo operacijoms stiprinti ir proaktyviai užkirsti kelią keliamoms grėsmėms (NIST, 2023). Tai leidžia teigti, kad JAV reguliacinis modelis užtikrinant kibernetinį saugumą dirbtinio intelekto kūrime labiau remiasi organizacijų pareiga pačioms diegti rizikos vertinimo ir kontrolės priemones, o ne detaliu horizontaliu reguliavimu.

Manytina, kad Azijos regionas ypač pasižymi dirbtinio intelekto pažangumu šiandien, tad verta apžvelgti kokie reikalavimai yra taikomi šiame regione. Šiuo atveju galima būtų išskirti Singapūrą, kuris pasižymi gana aiškiu gairėmis grindžiamu požiūriu. Kaip nurodo Singapūro kibernetinio saugumo agentūra *(angl. CSA)* (2024), nepaisant dirbtinio intelekto teikiamos reikšmingos naudos ekonomikai ir visuomenei, kuris didina efektyvumą ir skatina inovacijas įvairiuose sektoriuose, tokių sistemų kūrimas ir diegimas yra susijęs su kibernetinio saugumo rizikomis. Singapūro kibernetinio saugumo agentūra (2024) dirbtinio intelekto sistemos kūrimo etape taip pat išskiria 4 pagrindinius žingsnius, kurie iš esmės atsikartoja ir prieš tai aptartose gairėse: 1) apsaugoti savo tiekimo grandinę (pvz., mokymo duomenis, modelius, API ir programinės įrangos bibliotekas ir kt.); 2) renkantis atitinkamą modelį, įvertinti jo saugumo naudą ir galimus pažeidžiamumus: reikalinga atsižvelgti į veiksnius, kurie gali turėti įtakos atitinkamo modelio saugumui, tokius kaip sudėtingumas, paaiškinamumas, interpretuojamumas ir mokymo duomenų jautrumas; 3) nustatyti, sekti ir apsaugoti susijusius išteklius: reikalinga turėti procesus, skirtus ištekliams sekti, autentifikuoti, valdyti jų versijas ir juos apsaugoti; 4) apsaugoti pačios dirbtinio intelekto sistemos kūrimo aplinką: reikalinga taikyti standartinius infrastruktūros saugumo principus, tokius kaip tinkamų prieigos kontrolės priemonių ir žurnalizavimo bei stebėsenos įgyvendinimas, aplinkų atskyrimas ir saugios pagal numatytuosius nustatymus konfigūracijos. Autorė pastebi, kad šie reikalavimai iš esmės sutampa su prieš tai aptartomis CISA gairėmis, ir toks sutapimas yra suprantamas, kadangi Singapūras taipogi dalyvavo CISA gairių rengime. Tokios aplinkybės leidžia teigti, kad Singapūro reguliavimas dėl kibernetinio saugumo reikalavimų dirbtiniam intelektui atitinka tarptautinį reguliacinį lygį.

Kiek kitokios taisyklės yra taikomos Australijoje. Ten galioja dirbtinio intelekto saugos standartas, kuris nėra privalomas (savanoriškas) (Australijos Vyriausybė Pramonės, mokslo ir išteklių departamentas, 2025). Šis standartas nėra skirtas išskirtinai kibernetiniam saugumui, vis dėl to jame taip pat akcentuojami panašūs techniniai ir organizaciniai reikalavimai kaip buvo aptarti prieš tai, taip pat ir skiriamas dėmesys kibernetiniam saugumui. Australijos Savanoriškame dirbtinio intelekto saugumo standarte (2025) be kita ko, nurodoma, kad turi būti įdiegtos tinkamos kibernetinio saugumo priemonės, taip pat atsižvelgta į dirbtinio intelekto sistemos kibernetinį pažeidžiamumą. Tuo tarpu kibernetinio saugumo reikalavimai Australijoje nustatomi atskirai rizikos mažinimo strategijoje „Esminis aštuonetas“ *(angl. „Essential Eight“)* (Australijos signalų direktoratas, 2023). Autorės manymu, Australijos pavyzdys rodo, kad net ir tais atvejais, kai Saugumo standartas nėra privalomas, vis tiek formuojama aiški kryptis, jog dirbtinio intelekto sistemos turi būti kuriamos ir diegiamos atsižvelgiant į jų kibernetinio pažeidžiamumo aspektus.

Indijos atveju matomas dar kiek kitoks, labiau principais grindžiamas modelis. 2025 m. paskelbtose Indijos dirbtinio intelekto valdymo gairėse *(angl. India AI Governance Guidelines)* nurodoma, kad nacionalinis dirbtinio intelekto valdymo pagrindas turėtų būti grindžiamas septyniais pagrindiniais principais, tarp kurių išskiriami pasitikėjimas, žmogaus prioritetas, teisingumas ir lygybė, atskaitomybė, suprantamumas pagal projektavimą bei sauga, atsparumas ir tvarumas (India AI Governance Guidelines, 2026). Autorės manymu, Indijos pavyzdys rodo, kad dirbtinio intelekto sistemų saugumas gali būti kuriamas ne tik per griežtą vieną horizontalų teisės aktą, bet ir per nuosekliai formuojamą rizikos valdymo bei institucinio koordinavimo sistemą. Vis dėl to, manytina, kad tokių sistemų patikimumas tarptautinėje rinkoje nesulauktų didelio pasitikėjimo, ypač regionuose, kur yra apibrėžti aiškūs ir griežti kibernetinio saugumo reikalavimai kuriant ar diegiant dirbtinio intelekto sistemas.

Apibendrinant tai, kas buvo išdėstoma, galima daryti išvadą, kad pasaulyje dirbtinio intelekto sistemų kūrimo kibernetinio saugumo kontekste reikalavimai reglamentuojami nevienodai: Europos Sąjungoje labiau vyrauja privalomomis teisės normomis grindžiamas reguliavimas, o tokiuose regionuose kaip JAV, Jungtinė Karalystė, Singapūras, Australija ar Indija didesnė reikšmė dažniau teikiama gairėms, principams ir gerajai praktikai. Vis dėlto, nepaisant šių skirtumų, galima išskirti bendrus reikalavimus, kuriuos būtų galima laikyti atspirties tašku kuriant tokias sistemas, kaip antai, a) rizika grindžiamą požiūrį, b) kibernetinio saugumo užtikrinimą nuo pat kūrimo pradžios, c) visos sistemos, o ne tik modelio, apsaugą, d) duomenų konfidencialumo ir vientisumo užtikrinimą, e) prieigos kontrolę, f) dokumentavimą, e) auditą, g) reguliarius atnaujinimus, h) tiekimo grandinės saugumą ir i) žmogaus priežiūrą. Papildomai galima pažymėti, kad pasauliniu mastu kuriant dirbtinio intelekto sistemas svarbu atsižvelgti ne tik į bendrus saugumo reikalavimus, bet ir į tai, į kokią rinką konkretus sprendimas bus orientuotas, nes nuo to gali priklausyti tiek taikomi teisiniai reikalavimai, tiek praktiniai saugumo užtikrinimo sprendimai.

2. TYRIMO „DIRBTINIO INTELEKTO SISTEMŲ KŪRIMAS KIBERNETINIO SAUGUMO KONTEKSTE“ METODOLOGIJA

Apžvelgus teorinius aspektus, taip pat tikslinga šiuos dirbtinio intelekto sistemų kūrimo ypatumus įvertinti ir praktiniu lygmeniu. Toliau atliktu tyrimu siekiama įvertinti, kokie dirbtinio intelekto sistemų kūrimo ypatumai išryškėja kibernetinio saugumo kontekste, analizuojant pasaulinės rinkos tendencijas ir konkrečius praktinius atvejus. Taip pat siekiama nustatyti, kokie veiksniai yra svarbiausi kuriant tokias sistemas, kaip jose sprendžiami duomenų tvarkymo, žmogaus priežiūros, autonomiškumo ir saugumo užtikrinimo klausimai bei kaip šie aspektai atsispindi praktikoje.

Tyrimo teorinis pagrindimas.

Tyrimo teorinis pagrindas buvo formuojamas remiantis mokslinės literatūros, tarptautinių institucijų dokumentų ir analitinių ataskaitų analize. Nagrinėti šaltiniai parodė, kad dirbtinio intelekto sistemų kūrimas kibernetinio saugumo srityje nėra vien techninis procesas. Jis apima daug daugiau aspektų: duomenų valdymą, žmogaus dalyvavimą, technines ir organizacines saugumo priemones, taip pat teisinius bei reguliacinius reikalavimus. Mokslinėje literatūroje dažnai pabrėžiama, kad dirbtinis intelektas gali būti naudingas aptinkant grėsmes, nustatant įsibrovimus, analizuojant vartotojų elgseną, tiriant incidentus ir reaguojant į kibernetinius išpuolius. Vis dėlto kartu atkreipiamas dėmesys ir į tai, kad tokios sistemos sukuria naujų rizikų. Jos gali būti pažeidžiamos, ne visada pakankamai skaidrios, priimti klaidingus sprendimus arba net būti panaudotos kenkėjiškais tikslais (Akhtar ir Rawol, 2024; Shanbhag ir kt., 2024; Lazic, 2019).

Analizuoti moksliniai šaltiniai taip pat leidžia teigti, kad dirbtinio intelekto sistemos kibernetinio saugumo srityje neturėtų būti suprantamos tik kaip neutralūs technologiniai įrankiai. Polito ir Pupillo (2024) pažymi, kad tokios sistemos gali būti paveikiamos manipuliuojant įvesties duomenimis arba taikant duomenų nuodijimo metodus. Be to, jų patikimumą apsunkina tai, kad praktiškai neįmanoma iš anksto patikrinti visų galimų veikimo situacijų. Panašią mintį iškelia ir Jabbarova (2023), teigdama, kad dirbtiniu intelektu paremtos kibernetinio saugumo sistemos gali gerokai padidinti saugumo priemonių veiksmingumą, tačiau jos negali visiškai pakeisti žmogaus priežiūros.

Teorinį tyrimo pagrindą papildo ir norminiai bei strateginiai dokumentai. Europos Komisijos Jungtinio tyrimų centro ataskaitoje akcentuojama, kad kibernetinio saugumo reikalavimai turėtų būti taikomi ne tik pačiam dirbtinio intelekto modeliui. Jie turi apimti visą sistemą: naudojamus duomenis, sąsajas, infrastruktūrą ir kitus susijusius komponentus (Junklewitz et al., 2023). Panašios pozicijos laikosi ir ENISA, kuri dirbtinio intelekto kibernetinį saugumą sieja su platesniais sistemos patikimumo, atsparumo ir standartizavimo klausimais (ENISA, 2023). NIST ir Saugaus dirbtinio intelekto sistemų kūrimo gairėse taip pat pabrėžiama, kad saugumas turi būti užtikrinamas viso sistemos gyvavimo ciklo metu. Šiuose

dokumentuose išskiriamas rizika grindžiamas požiūris, aiškus atsakomybių paskirstymas ir būtinybė vertinti ne tik technologinius, bet ir organizacinius veiksnius (NIST, 2023; NCSC 2023).

Svarbus šio tyrimo teorinio pagrindimo elementas yra ir pasaulinės rinkos analizė. Pasaulio ekonomikos forumo 2025 ir 2026 m. ataskaitose nurodoma, kad dirbtinis intelektas tampa vienu iš svarbiausių veiksnių, keičiančių kibernetinio saugumo sritį. Vis dėl to šios ataskaitos taip pat rodo, kad organizacijos susiduria su nemažais praktiniais sunkumais. Tarp jų minimas specialistų kompetencijų trūkumas, ne iki galo aiškios rizikos, poreikis išlaikyti žmogaus atliekamą priežiūrą ir finansiniai apribojimai (World Economic Forum, 2025; World Economic Forum, 2026). Tai leidžia manyti, kad dirbtinio intelekto sistemų kūrimo ypatumų nepakanka nagrinėti vien tik teoriniu ar teisiniu lygmeniu. Šiuos klausimus būtina vertinti ir praktiniame organizacijų veiklos kontekste, nes būtent ten išryškėja realios tokių sistemų taikymo galimybės ir ribos.

Remiantis šiomis teorinėmis prielaidomis, tyrime pasirinkta derinti kelias tarpusavyje susijusias analizės kryptis: teorinių šaltinių analizę, pasaulinės rinkos vertinimą ir konkrečių atvejų lyginamąją analizę. Toks tyrimo modelis leidžia nuosekliai pereiti nuo bendresnio teorinio pagrindo prie praktinio klausimo, kaip realiai yra kuriamos dirbtinio intelekto sistemos ir kaip yra taikomos kibernetinio saugumo srityje.

Tyrimo tikslas ir uždaviniai.

Tyrimo tikslas – atskleisti dirbtinio intelekto sistemų kūrimo ypatumus kibernetinio saugumo užtikrinimo kontekste, remiantis pasaulinės rinkos analize ir pasirinktų atvejų lyginamuoju vertinimu.

Tyrimo uždaviniai:

1. Išskirti pagrindines pasaulines tendencijas, susijusias su dirbtinio intelekto sprendimų taikymu kibernetinio saugumo srityje.
2. Ištirti pasaulinės rinkos dirbtinio intelekto sistemų kūrimo praktiką kibernetinio saugumo kontekste.
3. Palyginti pasirinktus dirbtinio intelekto sistemų atvejus pagal vienodus analizės kriterijus, siekiant nustatyti bendrus ir skirtingus jų kūrimo požymius.
4. Pateikti rekomendacines gaires dėl dirbtinio intelekto sistemų kūrimo kibernetinio saugumo kontekste.

Tyrimo metodai ir analizės įrankiai.

Duomenų rinkimo metodai. Dirbtinio intelekto sistemų kūrimo kibernetinio saugumo kontekste tyrimui atlikti pasirinkta kokybinio pobūdžio tyrimo strategija, papildyta antrinių statistinių duomenų analize. Toks metodų derinys leidžia išsamiai nagrinėti tiek teorinius, tiek praktinius dirbtinio intelekto sistemų kūrimo aspektus bei jų sąsajas su kibernetiniu saugumu pasauliniame kontekste. Tyrimas yra ieškomojo ir aprašomojo pobūdžio: pirmiausia siekiama nustatyti ir susisteminti aktualias teorines įžvalgas,

o vėliau, remiantis pasaulinės rinkos duomenimis ir konkrečiais atvejais, įvertinti, kaip šie aspektai pasireiškia praktikoje. Tyrime naudoti šie duomenų rinkimo metodai:

1. Mokslinės literatūros analizė: teorinių šaltinių, tokių kaip moksliniai straipsniai, monografijos, tyrimų ataskaitos ir institucijų parengti dokumentai, analizė, siekiant išsiaiškinti dirbtinio intelekto sistemų kūrimo principus, jų taikymą kibernetinio saugumo srityje, pagrindinius iššūkius ir rizikas.
2. Pasaulinės rinkos analizė: antrinių statistinių duomenų, tarptautinių organizacijų ir rinkos ataskaitų duomenų rinkimas bei sisteminimas, siekiant nustatyti pagrindines dirbtinio intelekto sprendimų taikymo kibernetinio saugumo srityje tendencijas.
3. Atvejų analizė: viešai prieinamos informacijos apie pasirinktus dirbtinio intelekto sprendimus rinkimas ir sisteminimas, siekiant išsamiau įvertinti jų kūrimo logiką, veikimo principus, duomenų naudojimą, žmogaus vaidmenį ir taikomas saugumo priemones.

Duomenų analizės metodai. Tyrime surinkti duomenys analizuojami taikant šiuos metodus:

1. Mokslinės literatūros turinio analizė: teorinių šaltinių analizė, siekiant pagrįsti tyrimo problematiką ir išskirti pagrindinius dirbtinio intelekto sistemų kūrimo aspektus kibernetinio saugumo kontekste.
2. Antrinių statistinių duomenų analizė: pasaulinės rinkos ataskaitų ir kitų antrinių duomenų analizė, siekiant nustatyti dirbtinio intelekto taikymo kibernetinio saugumo srityje mastą, kryptis ir pagrindines tendencijas.
3. Lyginamoji atvejų analizė: pasirinktų dirbtinio intelekto sistemų palyginimas pagal vienodus kriterijus, siekiant nustatyti bendrus ir skirtingus jų kūrimo požymius kibernetinio saugumo kontekste.

Tyrimo instrumentas ir empirinis modelis.

Šiame tyrime empirinė dalis grindžiama dviem tarpusavyje susijusiais etapais: antrinių statistinių duomenų analize ir pasirinktų atvejų analize. Toks tyrimo modelis pasirinktas todėl, kad jis leidžia nuosekliai pereiti nuo bendrų pasaulinės rinkos tendencijų vertinimo prie detalesnio konkrečių dirbtinio intelekto sistemų nagrinėjimo kibernetinio saugumo kontekste.

Pirmajame etape atliekama antrinių statistinių duomenų analizė. Šio etapo tikslas: įvertinti pasaulinės rinkos tendencijas, susijusias su dirbtinio intelekto sprendimų taikymu kibernetinio saugumo srityje. Šiuo tikslu analizuojami tarptautinių organizacijų ir rinkos ataskaitų duomenys apie dirbtinio intelekto naudojimo mastą, pagrindinius sprendimų tipus, jų taikymo kryptis, diegimo motyvus ir su tuo susijusius iššūkius. Ši tyrimo dalis leido nustatyti bendras rinkos raidos kryptis ir išskirti tas sritis, kurios yra svarbiausios tolesnei analizei.

Antrajame etape atliekama pasirinktų atvejų analizė. Šio etapo tikslas: detaliau įvertinti, kaip praktikoje pasireiškia dirbtinio intelekto sistemų kūrimo ypatumai kibernetinio saugumo kontekste. Šiam etapui pasirinkti trys konkretūs dirbtinio intelekto sprendimai: Microsoft Security Copilot, Darktrace

ActiveAI Security Platform ir Google Security Operations su Gemini. Šie atvejai analizuojami pagal vienodus kriterijus, siekiant nustatyti bendrus ir skirtingus jų kūrimo požymius.

Atvejų analizėje taikomi trys pagrindiniai kriterijų blokai: 1) sistemų paskirtis ir funkcinė logika; 2) naudojamų dirbtinio intelekto sprendimų, duomenų tvarkymo ir integracijos ypatumai; 3) žmogaus priežiūros, autonomiškumo ir saugumo priemonių santykis. Tokie kriterijai pasirinkti todėl, kad jie leidžia nuosekliai palyginti nagrinėjamus atvejus ir geriau atskleisti, kaip praktiškai pasireiškia dirbtinio intelekto sistemų kūrimo logika kibernetinio saugumo srityje.

Atliekant tyrimą remtasi tik viešai prieinamais, oficialiais ir patikimais produktų aprašymais. Tyrimo rezultatai buvo sisteminami pagal darbo tikslą ir uždavinius, o jų pagrindu suformuluotos išvados ir rekomendacinės gairės dėl dirbtinio intelekto sistemų kūrimo kibernetinio saugumo kontekste.

Tyrimo imtis ir atrankos pagrindimas.

Šiame tyrime imtis formuojama analizuojamų šaltinių ir pasirinktų praktinių atvejų pagrindu. Atsižvelgiant į tai, kad tyrimas grindžiamas antrinių statistinių duomenų analize ir konkrečių dirbtinio intelekto sistemų atvejų analize, imties pagrindimą tikslinga sieti su duomenų šaltinių parinkimu ir atrinktų atvejų tinkamumu tyrimo tikslui pasiekti.

Pirmąją tyrimo imties dalį sudaro antriniai statistiniai ir analitiniai duomenys, naudoti pasaulinės rinkos analizei. Į šią imtį įtraukti viešai prieinami tarptautinių organizacijų, tyrimų centrų ir rinkos analizės bendrovių duomenys, kurie leidžia įvertinti dirbtinio intelekto sprendimų taikymo mastą kibernetinio saugumo srityje, pagrindines rinkos tendencijas, taikymo kryptis ir su tuo susijusius iššūkius. Tokie šaltiniai pasirinkti todėl, kad jie leidžia nagrinėti tyrimo objektą platesniu, pasauliniu mastu ir suteikia pakankamą pagrindą išskirti bendras rinkos raidos kryptis.

Antrąją tyrimo imties dalį sudaro trys pasirinktų dirbtinio intelekto sistemų atvejai. Šie atvejai atrinkti taikant tikslinę atranką. Toks atrankos būdas pasirinktas todėl, kad buvo siekiama nagrinėti ne bet kokius rinkoje esančius sprendimus, o tuos, kurie geriausiai leidžia atskleisti dirbtinio intelekto sistemų kūrimo ypatumus kibernetinio saugumo užtikrinimo kontekste.

Konkrečių atvejų atrankai taikyti šie kriterijai:

1. sistema turi būti taikoma kibernetinio saugumo srityje pasauliniu mastu;
2. apie ją turi būti pakankamai viešai prieinamos informacijos;
3. remiantis turimais duomenimis turi būti galima analizuoti sistemos paskirtį, veikimo logiką, duomenų naudojimą, žmogaus įsitraukimo lygį ir saugumo priemones;
4. pasirinkti atvejai turi būti tarpusavyje palyginami pagal vienodus kriterijus;
5. pasirinktos sistemos turi leisti nagrinėti skirtingus dirbtinio intelekto sistemų kūrimo modelius.

Pažymėtina, kad šio tyrimo tikslas nėra statistiškai reprezentuoti visą pasaulinę dirbtinio intelekto kibernetinio saugumo rinką ar apimti visus joje esančius sprendimus. Tyrimu siekiama analitiškai įvertinti

pagrindines rinkos tendencijas ir, remiantis tikslingai atrinktais atvejais, nustatyti svarbiausius dirbtinio intelekto sistemų kūrimo ypatumus.

Tyrimo eiga ir struktūra.

Tyrimo eigą sudaro šie pagrindiniai veiksmai:

1. antrinių statistinių ir analitinių duomenų šaltinių atranka;
2. pasaulinės rinkos duomenų rinkimas ir sisteminimas;
3. pasaulinės rinkos tendencijų analizė;
4. pagrindinių dirbtinio intelekto sprendimų tipų nustatymas kibernetinio saugumo srityje;
5. tyrimui tinkamų atvejų atrankos kriterijų nustatymas;
6. konkrečių atvejų atranka;
7. pasirinktų atvejų lyginamoji analizė pagal vienodus kriterijus;
8. gautų rezultatų apibendrinimas ir interpretavimas;
9. išvadų ir rekomendacinių gairių formulavimas.

Tyrimo struktūrą sudaro keturi pagrindiniai etapai.

Pirmuoju etapu buvo atliekama pasaulinės rinkos analizė. Šio etapo metu buvo renkami ir sisteminami antriniai statistiniai duomenys bei tarptautinių organizacijų ir rinkos ataskaitų informacija. Siekta įvertinti pagrindines dirbtinio intelekto sprendimų taikymo kibernetinio saugumo srityje tendencijas, jų augimo mastą, paklausą lemiančius veiksnius ir su jų diegimu susijusius iššūkius.

Antruoju etapu buvo nustatomi pagrindiniai dirbtinio intelekto sprendimų tipai, šiuo metu vyraujantys kibernetinio saugumo srityje. Šio etapo metu buvo vertinama, kokioms funkcijoms dirbtinis intelektas dažniausiai naudojamas, kokiose srityse jo taikymas yra labiausiai paplitęs ir kokios praktinės priežastys lemia vienų ar kitų sprendimų pasirinkimą.

Trečiuoju etapu buvo atrinkti konkretūs praktiniai atvejai ir atlikta jų lyginamoji analizė. Atvejai pasirinkti taikant tikslinę atranką, atsižvelgiant į tai, ar sistema naudojama kibernetinio saugumo srityje, ar apie ją yra pakankamai viešai prieinamos informacijos ir ar ją galima palyginti su kitais pasirinktais atvejais pagal vienodus kriterijus. Šiame etape pasirinkti sprendimai buvo analizuojami pagal jų paskirtį ir veikimo logiką, naudojamų dirbtinio intelekto sprendimų, duomenų tvarkymo ir integracijos ypatumus, taip pat žmogaus priežiūros, autonomiškumo ir saugumo priemonių santykį.

Ketvirtuoju etapu buvo apibendrinti tyrimo rezultatai, suformuluotos pagrindinės išvados ir pateiktos rekomendacinės gairės dėl dirbtinio intelekto sistemų kūrimo kibernetinio saugumo kontekste. Šiame etape siekta ne tik susisteminti nustatytus bendrus ir skirtingus požymius, bet ir parodyti jų praktinę reikšmę.

Tyrimo instrumentarijus.

2 lentelė. Tyrimo instrumentarijus.

Tyrimo kriterijai	Tyrimo indikatoriai	Tyrimo metodai
Pasaulinės rinkos raidos kryptys	1. Dirbtinio intelekto sprendimų taikymo mastas kibernetinio saugumo srityje.	Antrinių statistinių duomenų analizė.
	2. Rinkos augimo tendencijos ir plėtros kryptys.	Tarptautinių ataskaitų ir rinkos duomenų analizė.
	3. Pagrindiniai dirbtinio intelekto sprendimų diegimą lemiantys veiksniai ir iššūkiai.	Antrinių duomenų sisteminimas ir analizė.
Pagrindiniai dirbtinio intelekto sprendimų tipai kibernetinio saugumo srityje	1. Dažniausiai taikomi dirbtinio intelekto sprendimų tipai.	Pasaulinės rinkos analizė.
	2. Pagrindinės jų taikymo sritys ir paskirtis.	Antrinių statistinių duomenų analizė.
	3. Skirtingų sektorių prioritetai taikant dirbtinio intelekto sprendimus.	Tarptautinių ataskaitų duomenų analizė.
Tyrimui pasirinktų atvejų analizė	1. Sistemos paskirtis ir funkcinė logika.	Atvejų analizė.
	2. Naudojamų dirbtinio intelekto sprendimų, duomenų tvarkymo ir integracijos ypatumai.	Atvejų analizė ir dokumentų analizė.
	3. Žmogaus priežiūros, autonomiškumo ir saugumo priemonių santykis.	Atvejų analizė ir lyginamoji analizė.
Dirbtinio intelekto sistemų kūrimo ypatumų lyginamasis vertinimas	1. Bendri analizuotų sistemų kūrimo požymiai.	Lyginamoji analizė.
	2. Skirtingi analizuotų sistemų kūrimo požymiai.	Lyginamoji analizė.
Rekomendacinių gairių formulavimas	1. Pagrindinių tyrimo rezultatų apibendrinimas.	Tyrimo rezultatų interpretavimas.
	2. Rekomendacinių gairių dėl dirbtinio intelekto sistemų kūrimo pateikimas.	Tyrimo rezultatų apibendrinimas ir interpretacija.

Šaltinis: sudaryta darbo autorės.

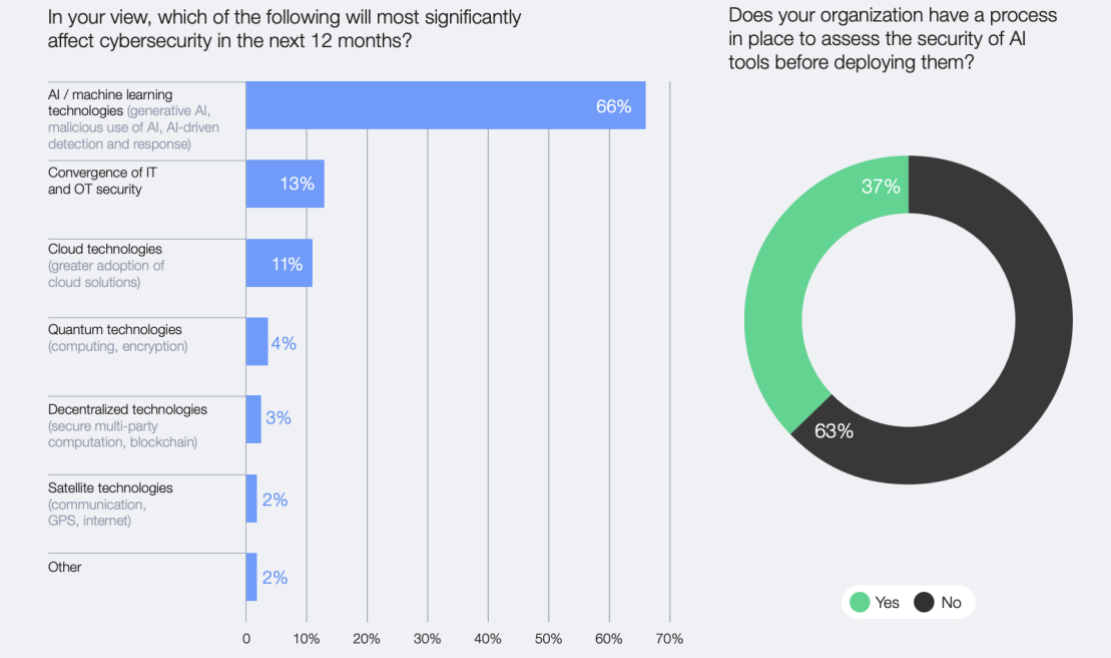
3. DIRBTINIO INTELEKTO SISTEMŲ KŪRIMO YPATUMŲ TYRIMAS KIBERNETINIO SAUGUMO UŽTIKRINIMO KONTEKSTE

3.1. Dirbtinio intelekto taikymo kibernetinio saugumo srityje pasaulinis kontekstas ir rinkos tendencijos

Ankstesniuose šio darbo skyriuose buvo aptarti dirbtinio intelekto sistemų kūrimo teoriniai aspektai, tokie kaip samprata, taikymo galimybės, susiję iššūkiai bei rizikos, taip pat įvairių pasaulio valstybių reguliavimas dirbtinio intelekto bei kibernetinio saugumo srityje. Šioje dalyje tikslinga pereiti prie praktinės dalies ir įvertinti kokios dirbtinio intelekto sistemos vyrauja šuo metu pasaulio rinkoje kibernetinio saugume kontekste. Atsižvelgiant į tai, toliau bus apžvelgiamos dirbtinio intelekto sprendimų, taikomų užtikrinant kibernetinį saugumą, pasaulinės rinkos tendencijos bei jų pasirinkimų motyvacijos.

Kaip jau buvo nustatyta anksčiau, kibernetinės grėsmės pasaulyje tampa vis sudėtingesnės ir dažnesnės, o kartu su tuo kyla ir susijusius saugumo rizikos. Dirbtinio intelekto priemonių įtaką kibernetiniam saugumui patvirtina ir Pasaulio ekonomikos forumo *(angl. World Economic Forum)* 2025 m. ataskaita, kurioje nurodoma, jog jų atliekamo *„Global Cybersecurity Outlook“* tyrimo duomenimis 72% respondentų nurodė, kad kibernetinė rizika išsaugo. Tas pats tyrimas taip pat išryškino kibernetinių nusikaltimą dažnumą bei sudėtingumą, prie kurių, be kita ko, prisidėjo ir dirbtinio intelekto priemonių paremtos atakos. Net 66% respondentų 2025 m. metais manė, kad dirbtinio intelekto technologijų taikymas turės didžiausią įtaką kibernetiniam saugumui per ateinančius 12 mėnesių (žr. 2 pav.). Kartu pažymėtina, kad tik 37% respondentų nurodė, jog jų organizacija turi procesus skirtus įvertinti dirbtinio intelekto priemonių saugumą prieš jų diegimą (World Economic Forum, 2025). Šie duomenys leidžia daryti išvadą, kad dirbtinio intelekto vaidmuo iš tiesų turi reikšmingą poveikį kibernetiniam saugumui.

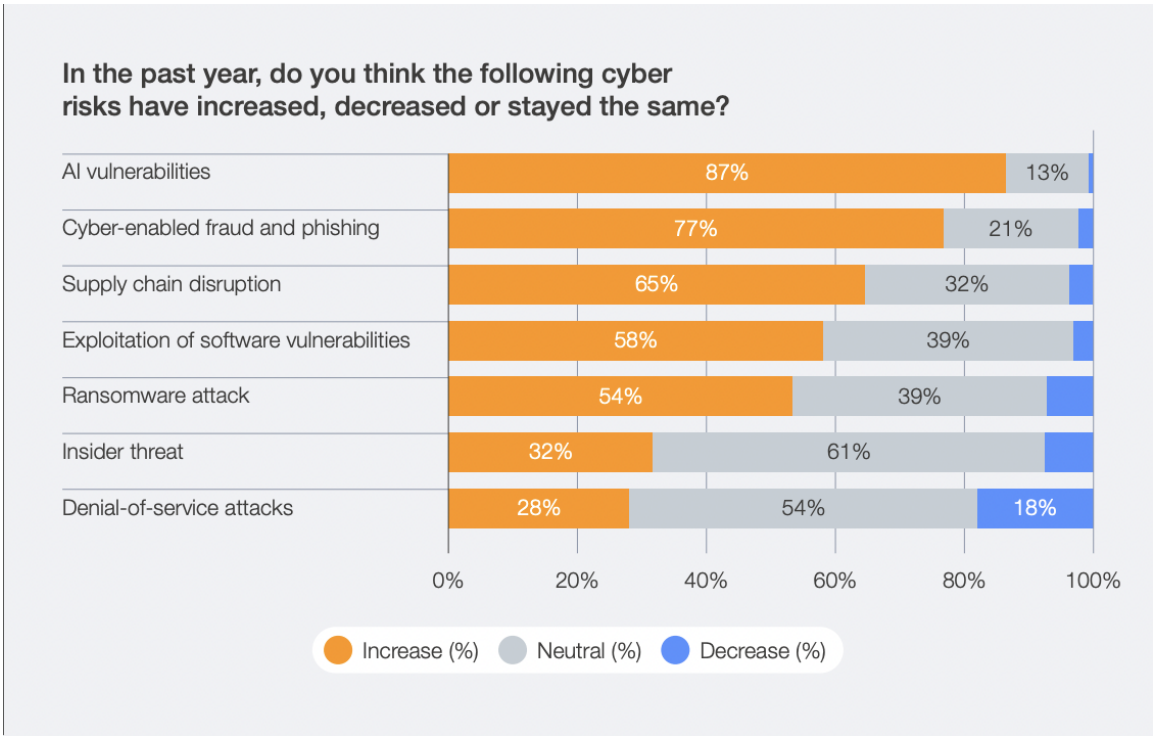
2 pav. Dirbtinio intelekto priemonių įtaka kibernetiniam saugumui.



Šaltinis: World Economic Forum, 2026

Dar iškalbingesni yra 2026 m. duomenys. Pasaulio ekonomikos forumo 2026 m. ataskaitos (World Economic Forum, 2026) duomenimis 94% respondentų mano, jog dirbtinis intelektas bus svarbiausias kibernetinio saugumo pokyčių veiksnys. Toje pačioje ataskaitoje taip pat atspindima, jog 87% respondentų dirbtinio intelekto priemonių sukeliamus pažeidžiamumus įvardijo kaip sparčiausiai augančią kibernetinę riziką (žr. 3 pav.). Toliau išskiriamos kitos kibernetinio saugumo grėsmės tokios, kaip sukčiavimas, fišingas, programinės įrangos pažeidžiamumai ir kiti (World Economic Forum, 2026).

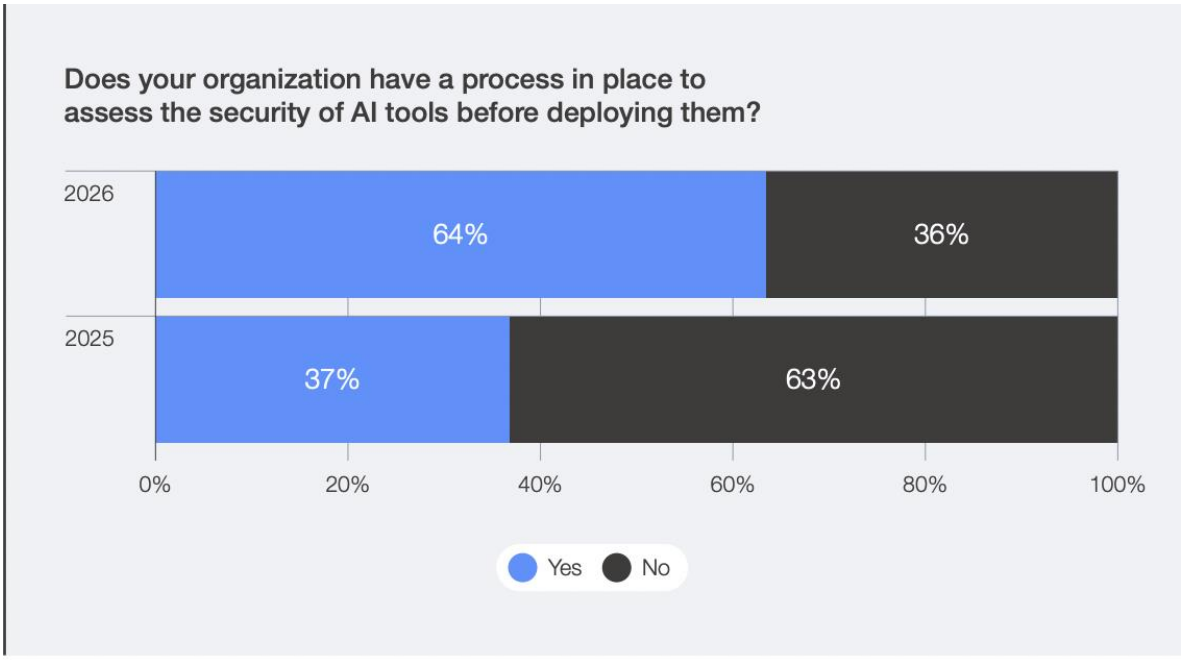
3 pav. Kibernetinio saugumo rizikos faktoriai. Šaltinis: World Economic Forum, 2026



Šaltinis: World Economic Forum, 2026

Iš ataskaitos duomenų taip pat galima matyti, kad ryšium su didėjančia grėsme, auga ir organizacijų sąmoningumas ir jos vis dažniau pradeda vertinti dirbtinio intelekto įrankių saugumą prieš jų diegimą: 2025 m. tokius vertinimus atliko 37% organizacijų, o 2026 m. šis procentas dvigubai išaugo iki 64% (žr. 4 pav.) (World Economic Forum, 2026).

4 pav. Dirbtinio intelekto priemonių saugumo vertinimo procesas organizacijose.



Šaltinis: World Economic Forum, 2026

Kaip buvo nagrinėta aukščiau, kartu su dirbtinio intelekto sprendimų naudojimo privalumais, atsiranda ir tam tikri iššūkiai bei rizikos. Pasaulio ekonomikos forumo 2026 m. ataskaitos duomenimis organizacijos susiduria su šiais sunkumais įgyvendinant dirbtinį intelektą kibernetiniam saugumui užtikrinti: 54% pakankamų žinių ir (ar) įgūdžių turėjimas, 41% aplinkybė, jog šie sprendimai turi būti peržiūrimi žmogaus prieš jų įgyvendinimą, 39% nurodo neapibrėžtumą dėl rizikų, 36% išskiria problemą dėl lėšų trūkumo įgyvendinant šiuos sprendimus ir 33% dėl neapibrėžtos naudos (žr. 5 pav.) (World Economic Forum, 2026). Šie duomenys leidžia manyti, kad siekiant sėkmingai įgyvendinti dirbtinio intelekto įrankius kibernetiniam saugumui užtikrinti, reikalinga atsižvelgti ir į galimus su tuo susijusius kylančius iššūkius.

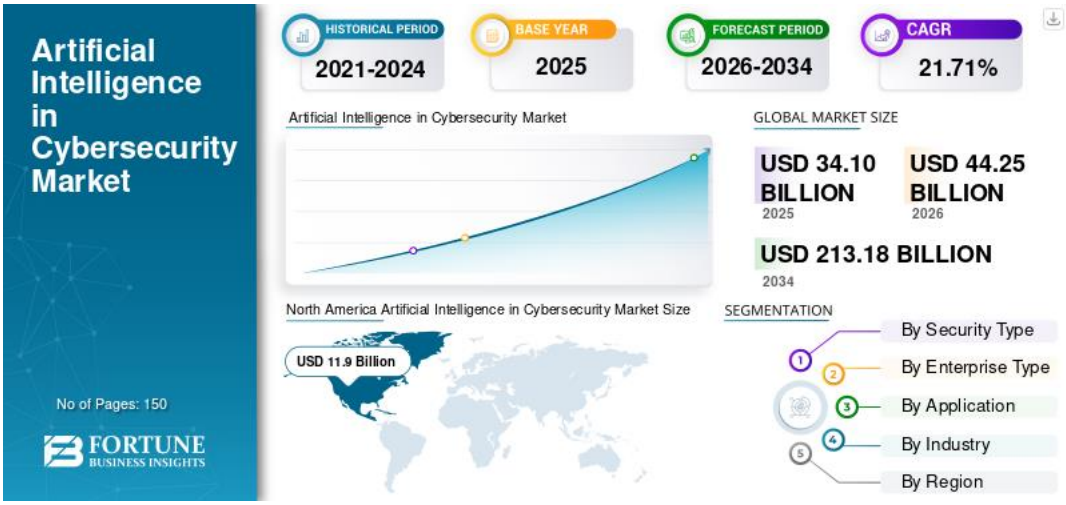
5 pav. Iššūkiai diegiant dirbtinio intelekto sprendimus.

Šaltinis: World Economic Forum, 2026

Kaip jau buvo aptarta teorinėje šio darbo dalyje, kalbant apie inovatyvių sprendimų įgyvendinimą, labai svarbu atsižvelgti ir jų diegimo kainą. Autorės manymu, siekiant visapusiškai įvertinti tokių sprendimų finansinę pusę, reikalinga apžvelgti ne tik kiek atitinkamas toks sprendimas kainuoja, bet ir kiek organizacijoms gali kainuoti tokių priemonių nebuvimas. IBM ir Ponemon Institute (2025) ataskaitoje nurodoma, kad vidutinė asmens duomenų saugumo pažeidimų sukelti nuostoliai pasaulyje sumažėjo iki 4,44 mln. JAV dolerių, palyginus su 2024 m. duomenimis, kur jie siekė 4,88 mln. JAV dolerių. Remiantis tos pačios ataskaitos duomenimis, įmonės, naudojančios dirbtinio intelekto sprendimus net 80 dienų sutrumpino savo vidutinę asmens duomenų saugumo pažeidimų nustatymų trukmę ir sutaupė 1,9 mln. JAV dolerių, palyginus su įmonėmis, kurios tokių sprendimų nenaudoja (IBM, 2025). Šie duomenys aiškiai rodo, kad vertinant atitinkamos dirbtinio intelekto technologijos diegimo kainą, reikalinga įvertinti ne tik tai, kiek kainuos jos diegimas, bet ir kiek finansinės bei kitos naudos ji gali atnešti.

Fortune Business Insights (2025) duomenimis pasaulinės dirbtinio intelekto kibernetinio saugumo rinkoje vertė siekė 34,09 mlrd. JAV dolerių. Analizuojant prognozes, numatoma, kad vertė augs nuo 44,24 mlrd. JAV dolerių 2026 m. iki 213,17 mlrd. JAV dolerių 2034 m., o prognozuojamu laikotarpiu metinis sudėtinis augimo tempas sudarys 21,71%. Taip pat iš šių duomenų matyti, kad Šiaurės Amerikos regionas užima dominuojančią padėtį šiame kontekste, sudarydama 34,90% dalies, šalia jos yra Europa, kuri taip pat užima reikšmingą rinkos dalį su 33,30%, Azijos ir Ramiojo vandenyno reginio rinka sudarė 21,41%, toliau išskiriama Artimųjų rytų bei Afrikos rinka su 10,65%, ir Pietų Amerika su 6,21% rinkos dydžiu (žr. 6 pav.) (Fortune Business Insights, 2025). Atsižvelgiant į tai, galima spręsti, jog dirbtinio intelekto priemonių naudojimas kibernetinio saugumo srityje turėtų ateityje dar labiau augti ir tapti neatsiejamu saugumo užtikrinimo sprendimu.

6 pav. Dirbtinis intelektas kibernetinio saugumo rinkoje.



Šaltinis: Fortune Business Insights, 2025.

Išanalizavus statistinius pasaulinės rinkos duomenis, galima daryti išvadą, kad dirbtinio intelekto naudojimas kibernetinio saugumo srityje jau šiandien užima reikšmingą vietą, o žvelgiant į prognozes, manytina, jog jų taikymas turėtų dar sparčiau augti. Dirbtinio intelekto priemonių taikymas siekiant užtikrinti kibernetinį saugumą yra lemiamas įvairių veiksnių, kaip antai, sudėtingesnės kibernetinio saugumo grėsmės, procesų automatizavimas bei galimos neigiamos ekonominės pasekmės dėl tokių priemonių nebuvimo. Vis dėl to, šių priemonių taikymą reikalinga kartu derinti su galimais iššūkiais, tokiais kaip darbuotojų kompetencijų kėlimas, techninės spragos ir su tuo kylančias rizikas bei skirti atitinkamą finansavimą tokioms saugumo priemonėms.

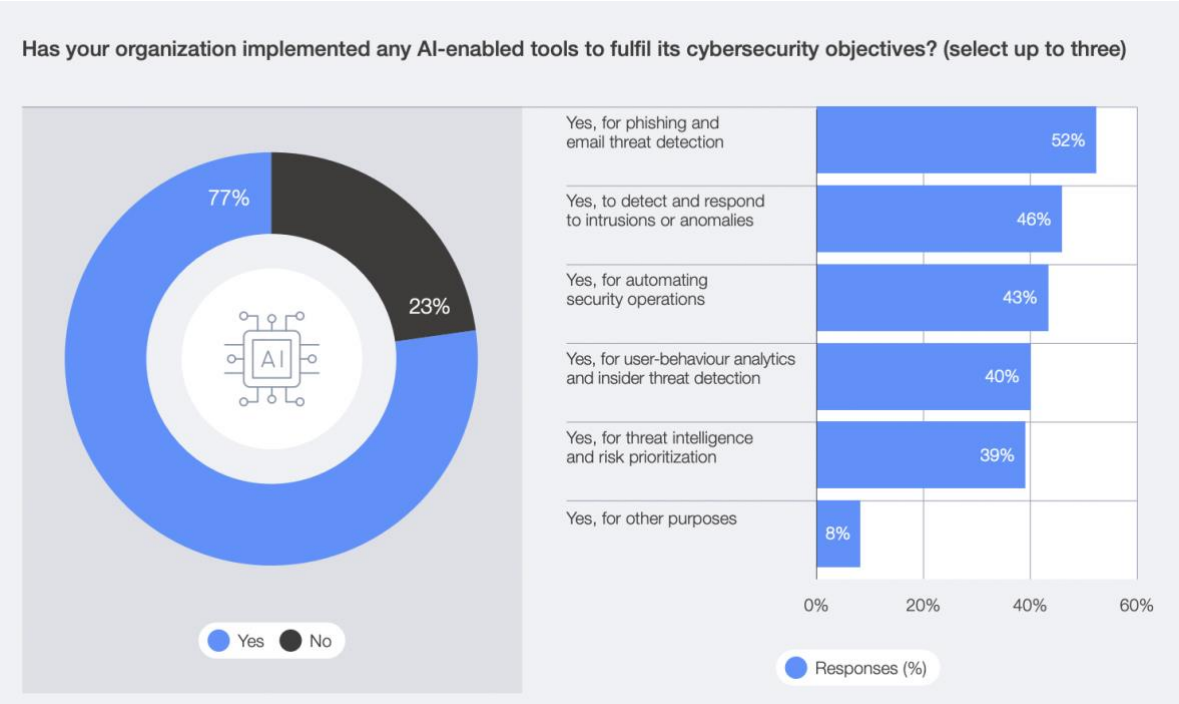
Pasaulinės rinkos vertinimas parodė, kad dirbtinio intelekto sprendimai kibernetinio saugumo srityje yra plačiai taikomi, todėl šioje dalyje taip pat bus siekiama nustatyti kokie iš jų šiuo metu vyrauja pasaulinėje rinkoje ir dėl kokių priežasčių.

Ucheji, Ekeneme ir Ezekwem (2025) atlikta sistemine ir bibliometrine analize leidžia pamatyti, kokios kryptys šiuo metu šioje srityje vyrauja ir dėl kokių priežasčių jos stiprėja. Tarp ryškiausių tendencijų

autoriai išskiria: a) įsibrovimų aptikimo sistemas, b) automatizuotą kenkėjiškų programų analizę, c) fišingo ir socialinės inžinerijos prevenciją bei d) saugumo operacijų automatizavimą. Šios tendencijos rodo, kad dirbtinio intelekto taikymas kibernetinio saugumo srityje plečiasi pirmiausia dėl to, kad tradicinės saugumo priemonės vis dažniau nebėra pakankamos reaguoti į sparčiai kintančias ir sudėtingėjančias grėsmes, todėl yra žvelgiamasi į inovatyvias priemones.

Kaip jau buvo nustatyta anksčiau, egzistuoja įvairių dirbtinio intelekto sprendimų, kurie yra skirti apsaugoti organizaciją nuo kibernetinių grėsmių. Pasaulio ekonomikos forumo 2026 m. duomenimis net 77% organizacijų jau yra įsidiegusios dirbtinio intelekto pagrindu veikiančius įrankius, kad galėtų pasiekti savo kibernetinio saugumo tikslus. Didžiąją dalį jų, 52%, yra naudojami fišingo ir el. pašto grėsmių aptikimui, 46% įsibrovimų ar anomalijų nustatymui, 43% automatizuotoms saugumo operacijoms, 40% vartotojų elgesio analitikai bei vidinių grėsmių aptikimui, 39% rizikų nustatymui ir dar 8% kitiems tikslams (žr. 7 pav.) (World Economic Forum, 2026). Šie duomenys rodo ne tik tai, kad dirbtinio intelekto technologijos jau yra aktyviai taikomos kaip kibernetinio saugumo sprendimų dalis, bet ir tai, kad egzistuoja tokių sprendimų įvairovė.

7 pav. Dirbtinio intelekto priemonių užtikrinant kibernetinį saugumą tendencijos.

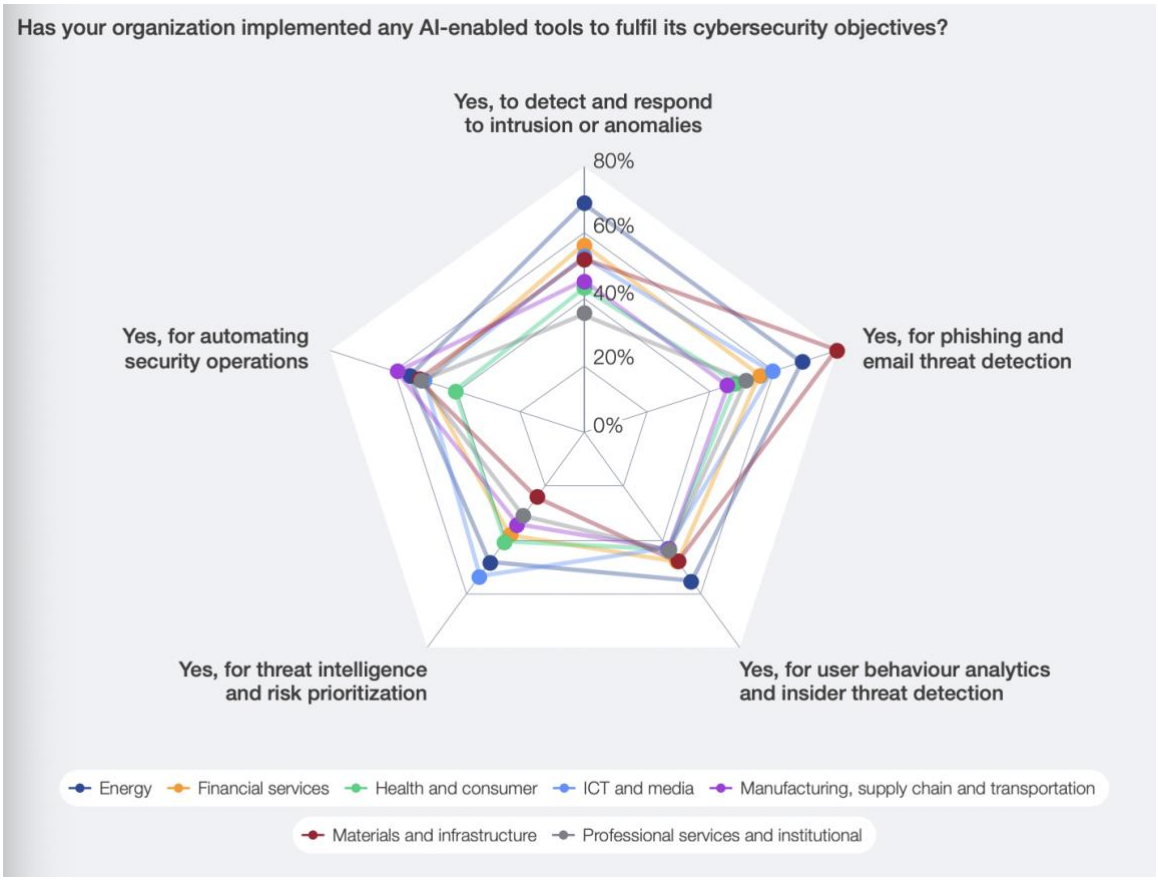


Šaltinis: World Economic Forum, 2026

Vis dėl to, dirbtinio intelekto priemonių taikymo prioretizavimą labiau atspindi pramoninis pasiskirstymas, atspindėdamas konkrečių sektorių rizikos profilius ir operacinius poreikius (World Economic Forum, 2026). Pasaulio ekonomikos forumo 2026 m. tyrimo duomenys rodo, jog fišingo atsparumui labiausiai taiko prioritetą medžiagų ir infrastruktūros sektorius (80%), kiek mažiau jį taiko

energetikos sektorius, taip pat ir kiti sektoriai. Energetikos sektorius didžiausią naudą mato taikant dirbtinį intelektą įsilaužimų ir anomalijų aptikimui (69%), taip pat vartotojų elgesio analizei ir vidinių spragų nustatymui (mažiau nei 60%). Grėsmių analizei ir rizikos prioritizavimui dirbtinio intelekto įrankių pagalba labiausiai taiko informacinių ir komunikacinių technologijų įmonės (daugiau nei 50%). Tuo tarpu automatizuotų saugumo sprendimų taikymui labiausiai taiko prioritetą gamybos ir transporto sektorius (59%) (žr. 8 pav.) (World Economic Forum, 2026). Šie duomenys leidžia daryti išvadą, kad nėra vienos universalios priemonės, kuri užtikrintų kibernetinį saugumą visose įmonėse. Taip pat manytina, kad šis pasiskirstymas rodo, jog skirtingų pramonių organizacijos, atsižvelgiant į jų veiklos pobūdį, susiduria su skirtingomis kibernetinėmis grėsmėmis, atitinkamai tokių priemonių įvairovė yra būtina ir konkretus sprendimas turi būti pritaikomas pagal aktualiausią tai organizacijai taikomą grėsmę.

8 pav. Dirbtinio intelekto priemonių naudojimas, pramoninis pasiskirstymas.



Šaltinis: World Economic Forum, 2026

Apibendrinus išanalizuotus duomenis, daroma išvada, jog dirbtinio intelekto įrankiai, skirti užtikrinti kibernetinį saugumą šiuo metu jau yra plačiai naudojami, tarp jų galima išskirti tokias priemones kaip apsauga nuo fišingo atakų, el. pašto grėsmių aptikimo sistemos, įsibrovimų ir anomalijų nustatymo įrankiai, automatizuoti saugumo sprendimai, taip pat grėsmių žvalgybos bei rizikos prioretizavimo technologijos. Vis dėl to, aptarti tyrimo duomenis taip pat leidžia daryti išvadą, kad priemonės turėtų būti

parenkamos ne pagal jų populiarumą rinkoje, o atsižvelgiant į konkrečios organizacijos ar pramonės galima kibernetinio saugumo grėsmes.

Šio poskyrio statistinė analizė parodė, kad dirbtinio intelekto taikymas kibernetinio saugumo srityje pasauliniu mastu jau dabar užima svarbią vietą, kasmet poreikis tokiems sprendimams vis auga, ir tikėtina, artimiausiais metais dar labiau didės. Tam įtakos turi ne tik sparčiai besivystančių technologijų pažanga ir susiję veiksniai, bet ir vis sudėtingesnės kibernetinės atakos, taip pat poreikis greičiau reaguoti į incidentus, automatizuoti saugumo procesus ir mažinti galimus finansinius nuostolius. Kartu taip pat paaiškėjo, kad dirbtinio intelekto sprendimų taikymas šioje srityje nėra vienodas, nes jų pasirinkimą lemia konkrečios pramonės / sektoriaus veiklos pobūdis, jai būdingos rizikos ir praktiniai saugumo poreikiai. Atsižvelgiant į tai galima teigti, kad dirbtinio intelekto priemonės kibernetinio saugumo srityje neturėtų būti vertinamos kaip universalus ar vien dėl inovatyvumo pasirenkamas sprendimas. Daug svarbiau įvertinti, kiek jos iš tikrųjų atitinka konkrečios organizacijos poreikius ir galimybes, rizikas ir bendrą kibernetinio saugumo aplinką.

3.2. Tyrimui pasirinktų atvejų pristatymas

Prieš tai aptarta apibendrinta pasaulinės rinkos analizė leido suprasti, jog dirbtinio intelekto sistemų kūrimui kibernetinio saugumo kontekste gali turėti įtakos įvairūs aspektai, kaip antai, atitinkamų regionų ar valstybių technologinė pažanga, šiems sprendimams skiriamas finansavimas, taip pat jų prioretizavimas pagal aktualiausias kibernetines grėsmes ir tokių technologijų konkrečios diegimo bei taikymo galimybės. Toliau šiame tyrime bus siekiama išsiaiškinti, kokios ir kokiu principu yra kuriamos atitinkamos dirbtinio intelekto sistemos skirtos kibernetinio saugumo užtikrinimui, ir kokius tokių sprendimų kūrimo ypatumus galima įžvelgti. Siekiant nustatyti minėtus ypatumus, toliau bus analizuojami konkretūs atvejai, kad būtų galima išnagrinėti kokiu principu buvo sukurtos tokios sistemos ir kokius kibernetinio saugumo komponentus jos turi.

Konkrečių atvejų analizei buvo pasirinkti trys atvejai, remiantis tikslinės atrankos kriterijais. Visų pirma buvo siekta, kad analizuojama dirbtinio intelekto sistema būtų taikoma kibernetinio saugumo užtikrinimo srityje, pavyzdžiui, reagavimui į incidentus, tinklo stebėsenai, išsilaužymų aptikimui ir pan., ir būtų plačiai taikoma pasaulyje. Ne mažiau svarbu buvo, kad apie pasirinktą sistemą būtų pakankamai viešai prieinamos informacijos tam, kad būtų galima įvertinti jos paskirtį, veikimo logiką, architektūrą, duomenų tvarkymo mechanizmus, žmogaus įsitraukimo lygį bei taikomas technines ir organizacines saugumo priemones. Galiausiai, atrinkti atvejai turėjo sudaryti galimybę analizuoti skirtingus dirbtinio intelekto sistemų kūrimo modelius, taip pat ketvirta, būti tarpusavyje palyginami pagal vienodus kriterijus.

Atsižvelgiant į prieš tai minėtus kriterijus, šio darbo tyrimui buvo pasirinkti trys dirbtinio intelekto sistemų analizės atvejai: Microsoft Security Copilot, Darktrace ActiveAI Security Platform ir Google Security Operations su Gemini.

Pirmasis pasirinktas atvejis yra Microsoft Security Copilot (toliau – Security Copilot). Microsoft dokumentacijoje nurodoma, jog tai yra generatyviniu dirbtiniu intelektu pagrįstas saugumo sprendimas, skirtas didinti saugumo specialistų darbo efektyvumą ir išplėsti jų galimybes, kad jie veiktų kompiuterio greičiu bei mastu (Microsoft Build, n.d.). Ši sistema gali būti taikoma įvairiuose kibernetinio saugumo scenarijuose, pavyzdžiui, reaguojant į incidentus, ieškant grėsmių, renkant žvalgybos duomenis, valdant rizikas ir kitose srityse; ir yra prieinama visame pasaulyje (Microsoft Build, n.d.). Kayode, Adebola ir Akerele (2025) taip pat pažymi, kad šis įrankis sujungia „OpenAI“ GPT-4 architektūrą su Security Copilot duomenimis ir veikia per pokalbio sąsają, pagrįstą natūralios kalbos modeliais. Praktikoje tai atrodo taip: naudotojui paprašius ištirti konkrečius kenkėjiškos programinės įrangos indikatorius, sistema gali surasti susijusią informaciją, ir pokalbio lange su naudotoju pasiūlyti jam galimą reagavimo strategiją (Kayode ir kt., 2025). Viešai prieinama šios sistemos techninė informacija leidžia gan išsamiai nagrinėti tokius sistemos aspektus kaip atskirų agentų veikimo logiką, naudotojų autentifikavimo modelius, prieigos kontrolės mechanizmus, asmens duomenų apsaugos klausimus ir kt.

Antrasis šiame tyrime analizuojamas atvejis yra Darktrace ActiveAI Security Platform (toliau – Darktrace). Oficialioje Darktrace svetainėje ši platforma pristatoma kaip plačiai pasaulyje taikomas sprendimas, leidžiantis vienoje sistemoje užtikrinti proaktyvų požiūrį į kibernetinį atsparumą, suteikiant organizacijai išankstinį saugumo būklės matomumą, grėsmių aptikimą realiuoju laiku ir autonominę reakciją į kibernetines grėsmes (Darktrace, n.d.). Pagrindinės Darktrace funkcijos apima grėsmių aptikimą, savarankiškai besimokantį dirbtinį intelektą, grėsmių aptikimo automatizavimą ir elgsenos analizę (ITButler, n.d.). Remiantis viešai pateikiama sistemos aprašymo informacija, galima vertinti tokius aspektus kaip reagavimo į incidentus procesą, integracijų sąveiką, savarankiško sistemos mokymosi procesą, nuolatinę tinklo stebėseną ir kitus su sistemos veikimu susijusius elementus(Darktrace, n.d.).

Trečiasis tiriamas atvejis yra Google Security Operations (toliau – SecOps). SecOps yra visame pasaulyje prieinama debesijos pagrindu veikianti saugumo operacijų sistema, suteikianti saugumo specialistams galimybę veiksmingiau aptikti, tirti kibernetinio saugumo grėsmes ir į jas reaguoti. Exabeam (n.d.) duomenimis, SecOps apima tokias pagrindines funkcijas kaip duomenų rinkimas, grėsmių aptikimas ir analizė, atvejų valdymas, reagavimo automatizavimas, ataskaitų rengimas ir taisyklių, skirtų grėsmėms aptikti, nustatymas. viešai skelbiama informaciją apie šį sprendimą leidžia analizuoti taikomą dirbtinio intelekto modelį, integruotą į saugumo operacijas, duomenų analizės procesus, prieigos valdymo procesus ir kt.

Taigi, autorės manymu, aukščiau aprašyti atvejai sudaro pakankamą pagrindą šias dirbtinio intelekto sistemas analizuoti pagal metodologijos dalyje nustatytus kriterijus. Manoma, jog toks

pasirinkimas leis pakankamai išsamiai įvertinti dirbtinio intelekto sistemų kūrimo ypatumus kibernetinio saugumo užtikrinimo kontekste.

3.2.1. Dirbtinio intelekto sistemų kūrimo ypatumų lyginamoji analizė kibernetinio saugumo kontekste

Šioje tyrimo dalyje bus lyginami pasirinkti atvejai, taikant vienodus analizės kriterijus. Manoma, kad toks palyginimas sudarys sąlygas įvertinti, kokie sistemų kūrimo ypatumai išryškėja naudojant dirbtinio intelekto sistemas, užtikrinančios kibernetinį saugumą.

Lyginimo kriterijai pasirinkti atsižvelgiant į ankstesniuose šio darbo skyriuose aptartus teorinius aspektus ir atliktos rinkos analizės rezultatus. Šio darbo teorinėje dalyje buvo nustatyta, kad kuriant dirbtinio intelekto sistemas svarbu vertinti ne tik jų pažangumą ir taikymo galimybes, bet ir naudojamus metodus, duomenų tvarkymo principus, žmogaus vaidmenį bei technines ir organizacines saugumo priemones. Rinkos analizė taip pat leido suprasti, kad kibernetinio saugumo srityje taikomi dirbtinio intelekto sprendimai gali būti įvairūs. Jie gali skirtis savo paskirtimi, technologine veikimo logika, taikomomis saugumo priemonėmis ir kt. Kartu apibendrinus šiuos rezultatus, buvo prieita išvados, kad kibernetinį saugumą užtikrinančios dirbtinio intelekto sistemos gali pasireikšti tokiais aspektais, kaip 1) sistemos atliekama funkcija, 2) taikomais sprendimais ir 3) jos autonomiškumo lygis.

Dėl šių priežasčių tolesnėje analizėje pasirinkti atvejai bus lyginami pagal šiuos kriterijus:

1. **Sistemų paskirtis ir funkcinė logika**. Šiuo kriterijumi siekiama įvertinti, kokiam tikslui sistema buvo sukurta ir kaip konkrečiai ji pasireiškia siekiant užtikrinti kibernetinį saugumą.
2. **Naudojamų dirbtinio intelekto sprendimų, duomenų tvarkymo ir integracijos ypatumai**. Šiuo kriterijumi bus analizuojama, kokie dirbtinio intelekto metodai konkrečioje sistemoje yra taikomi, kokiais duomenimis sistema remiasi, iš kur gauna tuos duomenis, ir kaip ji integruojama į platesnę saugumo infrastruktūrą ar kitus organizacijos procesus.
3. **Žmogaus priežiūros, autonomiškumo ir saugumo priemonių santykis.** Šiuo kriterijumi bus vertinama, kiek sistema gali veikti savarankiškai, kiek žmogus ją prižiūri ir (ar) ar atlieka kitus veiksmus, ir kokios techninės ar organizacinės priemonės taikomos siekiant užtikrinti saugų sistemos naudojimą.

Manytina, kad šie kriterijai sudaro pagrindą nuosekliai ir kryptingai palyginti pasirinktus praktinius atvejus, taip pat gali padėti nubrėžti vaizdą, kaip praktikoje pasireiškia dirbtinio intelekto sistemų kūrimo logika kibernetinio saugumo kontekste.

Sistemų paskirties ir funkcinės logikos palyginimas.

Pirmasis lyginamosios analizės kriterijus yra pasirinktų dirbtinio intelekto sistemų paskirtis ir funkcinė logika. Pažymėtina, kad visi nagrinėjami sprendimai buvo sukurti siekiant užtikrinti kibernetinį saugumą, tačiau siaurąja prasme jų vaidmuo šiame kontekste nėra vienodas. Pirmasis analizės kriterijus leidžia geriau suprasti, kokiai problemai spręsti buvo kuriama konkreti dirbtinio intelekto sistema ir kokią funkciją ji atlieka organizacijos saugumo procesuose.

Kaip nurodo Microsoft (2026), Security Copilot pirmiausia orientuotas į dirbtinio intelekto agentų pagalbą saugumo ir IT specialistams, taip pat darbo procesams optimizuoti ir produktyvumo didinimui. Šis sprendimas padeda greičiau tirti incidentus, atlikti pasikartojančias užduotis, analizuoti gautą informaciją, rinkti su grėsmėmis susijusius duomenis ir išsamiau sudaryti bendrą organizacijos saugumo situacijos vaizdą (Microsoft, 2026). Antriniuose šaltiniuose taip pat pabrėžiama jo reikšmė grėsmių aptikimui, incidentų tyrimui ir saugumo operacijų centralizavimui (The Knowledge Academy, 2026). Ši sistema pasižymi tokiais dirbtinio intelekto metodais, kaip a) natūralios kalbos apdorojimas (generatyvinis dirbtinis intelektas), t.y. su šia sistema galima bendrauti natūralia kalba, todėl specialistams lengviau suprasti situaciją ir nuspręsti, kokių veiksmų reikėtų imtis; ir b) mašininio mokymo algoritmai, kurie analizuoja vartotojų elgseną ir pateikia jų elgesio prognozes (Microsoft, 2026). Taigi, Security Copilot galima vertinti kaip generatyvinio dirbtinio intelekto pagrindu veikiantį saugumo asistentą, kuris sustiprina žmogaus atliekamą analizę, taip pat naudoja mašininio mokymo algoritmus, siekiant analizuoti ir daryti prognozes.

Darktrace (n.d.) teigia, kad jos sistema pasižymi kitokia veikimo logika nei didžioji dalis kibernetinio saugumo pramonėje esamų sprendimų. Pasak juos, ši platforma sujungia kiekvieną dirbtinio intelekto sąveiką į vieną vaizdą ir skirta nuolat stebėti organizacijos aplinką, aptikti galimas grėsmes bei reaguoti į jas itin greitai ir efektyviai. Praktikoje tai reiškia, jog šiuo atveju dirbtinis intelektas nėra vien tik pagalbinė priemonė specialistui, o tampa aktyvesne saugumo sistemos dalimi, galinčia tiesiogiai prisidėti prie grėsmių suvaldymo (Darktrace, n.d.). Tokį vertinimą pagrindžia ir Darktrace ActiveAI Security Platform pristatymas, kuriame platforma siejama su savimokos technologijomis, grėsmių aptikimu, kontekstiniu situacijos analizės bei supratimu ir autonominio atsako galimybėmis (Darktrace, n.d.). Atsižvelgiant į tai, Darktrace galima būtų apibūdinti kaip aktyvų savimokos ir pažangų reagavimo dirbtinio intelekto sprendimą.

SecOps savo paskirtimi galima apibūdinti kaip tarpinį variantą tarp pirmų dviejų minėtų sistemų. SecOps aprašyme nurodoma, jog ši platforma padeda saugumo komandoms aptikti, tirti ir valdyti daugiau grėsmių įdedant mažiau pastangų ir greičiau į jas reaguoti, o integracija su Gemini naudojama informacijos paieškai, santraukų rengimui, galimų veiksmų siūlymui ir tyrimo procesų palengvinimui, panašiai kaip SecurityCopilot atveju (Google Cloud, 2026). Papildomai, sistema gali atlikti pirminį grėsmių įspėjimų vertinimą ir pateikti jų trumpą pristatymą bei paaiškinimą. Tokiu būdu dirbtinis intelektas reikšmingai prisideda prie analizės ir operacijų efektyvinimo, tačiau galutinis sprendimas priklauso žmogaus

atsakomybei (Google Cloud, 2026). Remiantis šita informacija, SecOps gali būti vertinama kaip dirbtiniu intelektu sustiprinta saugumo operacijų platforma.

Apibendrinant galima teigti, kad šių sistemų paskirtis panaši tik bendrąja prasme, nes visos jos susijusios su kibernetinio saugumo užtikrinimu ir gali yra prieinamos visame pasaulyje. Vis dėl to žvelgiant siauriau, jų funkcinė logika skiriasi: Security Copilot veikia kaip saugumo specialisto asistentas, Darktrace labiau orientuota į aktyvų grėsmių stebėjimą ir autonominį reagavimą, o SecOps stiprina platesnį saugumo operacijų procesą. Autorės manymu, šis skirtumas yra reikšmingas, nes nuo sistemos paskirties priklauso jos kūrimo prioritetai, numatomos funkcijos, autonomiškumo lygis ir žmogaus vaidmuo priimant sprendimus.

Apibendrintas palyginimas pateikiamas toliau lentelėje (žr. 3 lentelę).

3 lentelė. Analizuojamų sistemų paskirties ir funkcinės logikos palyginimas.

Sistema	Pagrindinė paskirtis	Funkcinė logika
Security Copilot	Padėti saugumo ir IT specialistams tirti incidentus ir greičiau priimti sprendimus; darbo procesams optimizuoti ir produktyvumo didinimui.	Generatyvinio dirbtinio intelekto pagrindu veikiantis saugumo asistentas; mašininio mokymo algoritmas, siekiant analizuoti ir daryti prognozes.
Darktrace	Aptikti grėsmes realiuoju laiku ir parengti atsako strategija.	Savimoka, elgsenos analizė ir savarankiškas (autonominis) reagavimas.
SecOps	Stiprinti saugumo operacijų procesą ir tyrimą.	Sisteminis modelis su integruotu generatyviniu dirbtiniu intelektu.

Šaltinis: sudaryta autorės remiantis Microsoft (2026), Darktrace (n.d.) ir Google Cloud (2026).

Naudojamų dirbtinio intelekto sprendimų, duomenų tvarkymo ir integracijos ypatumai.

Antrasis lyginamosios analizės kriterijus yra naudojami dirbtinio intelekto sprendimai, duomenų tvarkymas ir integracija su kitomis vidinėmis ir (ar) išorinėmis sistemomis. Lyginant pasirinktus sistemų atvejus pagal aptariamą kriterijų, galima matyti, kad jų veikimo pagrindas gana aiškiai skiriasi. Kaip minėta, manoma, jog šis kriterijus gali leisti išsiaiškinti ne tik kokiu technologiniu principu veikia kiekviena sistema, bet ir iš kokių šaltinių ji gauna informaciją, kaip ją apdoroja ir toliau tvarko bei kiek yra pati sistema yra susieta su atitinkama organizacijos technologine aplinka.

Kaip jau buvo minėta, Security Copilot remiasi generatyviniu dirbtiniu intelektu. Ši sistema naudoja jau pačios organizacijos turimą informaciją, saugumo papildinius, papildomai išorinius patikimus šaltinius ir istorinius grėsmių duomenis tam, kad galėtų pateikti atsakymus į sistemai užduodamus klausimus bei

padėti saugumo specialistams atlikti kasdienes jų užduotis. Taip pat sistemos aprašyme pažymima, kad šioje platformoje yra įdiegti agentai, kurie taip pat gali padėti susieti turimus duomenis su dirbtinio intelekto galimybėmis ir integruoti šias funkcijas į platesnius darbo procesus (Microsoft, 2026).

Duomenų tvarkymo požiūriu Security Copilot veikia remdamasis jau organizacijos pačios nustatytomis prieigos teisėmis ir esamomis integracijomis. Praktikoje tai reiškia, kad pati sistema gali pasiekti tik tuos duomenis, kuriems turi organizacijos suteiktą prieigą ir pateikti tik tokią informaciją jos naudotojui, kurią jam leidžiama pasiekti pagal prieigos teises. Aptariama sistema gali tvarkyti tokias įvestis kaip naudotojo užklausos, į jas sistemos sugeneruoti atsakymai, įkelti dokumentai ir tam tikri sistemos įrašai. Remiantis tuo, galima teigti, kad šio sprendimo veiksmingumas labai priklauso nuo to, kaip organizacijoje yra sutvarkytas informacijos prieinamumas ir kaip Security Copilot bus susietas su kitais Microsoft saugumo sprendimais (Microsoft, 2026).

Darktrace labiau grindžiama nuolatine organizacijos aplinkos stebėsena ir jos veiklos modelio supratimu, kad būtų galima aptikti nukrypimus nuo įprastos organizacijos saugumo praktikos. Darktrace (n.d.) nurodo, kad sistema realiuoju laiku vertina organizacijos duomenis ir gali būti integruojama su įvairiais kitais sprendimais, pavyzdžiui, debesijos, tinklo ar saugumo įrankiais. Vadinasi, jos veikimas priklauso nuo visos organizacijos technologinės aplinkos stebėjimo, o ne nuo vieno konkretaus įrankio ar vien tik tekstu pagrįsto modelio (Darktrace, n.d.). Šią veikimo logiką papildo Darktrace aprašomas „Cyber AI Loop“ principas, pagal kurį yra tarpusavyje susiejamos prevencijos, aptikimo ir reagavimo funkcijos, o skirtingų sistemos dalių duomenys gali būti naudojami grėsmei stebėti per kelis atakos etapus (Darktrace, 2023). Taigi, Darktrace atveju ypač svarbi tampa duomenų integracija tarp skirtingų organizacijos infrastruktūros dalių ir gebėjimas vertinti organizacijos elgsenos pokyčius realiuoju laiku, kad būtų į juos tinkamai reaguota.

SecOps išsiskiria tuo, kad dirbtinis intelektas čia integruojamas į platesnę saugumo operacijų valdymo sistemą. Google Cloud (2026) nurodo, kad Gemini papildomai gali padėti ieškoti informacijos, rengti santraukas, siūlyti taisykles ir naudoti saugumo operacijų duomenis, ir tokiu būdu padėti specialistams dirbti su tokia informacija greičiau. Praktiniu požiūriu tai reiškia, kad šie du įrankiai yra tarpusavyje susieti ir veikia kartu, t.y. vienas ieško reikiamos informacijos ir padeda ją apdoroti suprantamu būdu, o kitas skirtas visai saugumo informacijai saugoti, valdyti ir analizuoti. Dokumentacijoje taip pat matyti, kad daug dėmesio skiriama prieigos valdymui, t. y. tam, kas ir kokią informaciją gali matyti bei naudoti. Tai rodo, kad šiuo atveju dirbtinis intelektas neveikia kaip visiškai atskiras sprendimas, bet papildo jau veikiančią saugumo operacijų sistemą (Google Cloud, 2026).

Apibendrinant galima teigti, kad Security Copilot labiau remiasi generatyviniu modeliu, papildiniais ir sąsajomis su kitais Microsoft sprendimais. Darktrace labiau siejama su nuolatine aplinkos stebėsena, elgsenos analize ir duomenų vertinimu realiuoju laiku. Tuo tarpu SecOps veikia kaip platesnės saugumo operacijų sistemos dalis, kurioje dirbtinis intelektas naudojamas papildomai analizei, informacijos

apibendrinimui ir tam tikrų procesų automatizavimui. Manoma, jog šie skirtumai atspindi, kad dirbtinio intelekto sistemų kūrimas kibernetinio saugumo srityje priklauso ne tik nuo pasirinktos technologijos, bet ir nuo to, kaip ji susiejama su organizacijos duomenimis, darbo procesais ir naudojama infrastruktūra.

Apibendrintas palyginimas pateikiamas toliau lentelėje (žr. 4 lentelę).

4 lentelė. Naudojamų dirbtinio intelekto sprendimų, duomenų tvarkymo ir integracijos ypatumai.

Sistema	**Dirbtinio intelekto sprendimo tipas**	**Duomenų logika**	**Integracijos ypatumai**
Security Copilot	Generatyvinis dirbtinis intelektas.	Užklausos, atsakymai, įkelti dokumentai, prieiga pagal suteiktas teises.	Gali būti susieta su kitais Microsoft saugumo sprendimais.
Darktrace	Elgsenos analizė ir savarankiškas mokymasis.	Nuolatinis organizacijos aplinkos stebėjimas realiuoju laiku.	Jungiama su įvairiomis saugumo ir tinklo sistemomis.
SecOps	Generatyvinis dirbtinis intelektas.	Didelės apimties saugumo duomenys, prieiga pagal suteiktas teises.	Veikia kaip platesnės saugumo sistemos dalis.

Šaltinis: sudaryta autorės remiantis Microsoft (2026), Darktrace (n.d.) ir Google Cloud (2026).

Žmogaus priežiūros, autonomiškumo ir saugumo priemonių palyginimas.

Trečiasis lyginamosios analizės kriterijus pagal kurį buvo lyginami šio tyrimo atvejai apima žmogaus priežiūros, sistemos savarankiškumo ir taikomų saugumo priemonių santykį. Šis kriterijus leidžia įvertinti, kiek veiksmų gali atlikti pati sistema savarankiškai, kiek ji yra prižiūrima žmogaus ir kiek kontrolės išlieka žmogaus rankose. Autorės manymu, tai yra vienas svarbiausių kriterijų, nes nuo jo priklauso ne tik sistemos veikimo pobūdis, bet ir pasitikėjimo ja praktikoje lygis.

Security Copilot labiausiai atitinka tokį modelį, kuriame dirbtinis intelektas veikia kaip žmogaus pagalbininkas, o ne kaip savarankiškas sprendimų priėmėjas. Kaip jau buvo minėta, Microsoft (2026) nurodo, kad ši sistema gali automatizuoti pasikartojančias užduotis, tačiau ji veikia pagal pačios organizacijos tiek sistemai, tiek naudotojui suteiktas prieigos teises ir yra integruojama į įprastus organizacijos darbo procesus. Iš esmės tai reiškia, kad žmogus visada išlieka pagrindinis sprendimų priėmėjas atsižvelgdamas į sistemos pateikiamus duomenis. Be to, Microsoft (2026) pabrėžia mažiausių teisių principo taikymą ir nurodo, kad klientų duomenys niekada nėra naudojami atitinkamo modelio mokymui (Microsoft, 2026). Atsižvelgiant į tai, galima teigti, kad Security Copilot sukurtas taip, jog

palengvintų specialistų darbą, pateikdamas jam išsamų vaizdų apie organizacijos saugumą, tačiau visa galutinių sprendimų atsakomybė tenka žmogui.

Darktrace atveju sistemos savarankiškumo lygis yra kur kas didesnis. Darktrace (n.d.) nurodo, kad ši platforma gali savarankiškai greitai reaguoti į grėsmes ir padėti jas stabdyti realiuoju laiku. Taip pat akcentuojama jos autonomiška nuolatinė organizacijos aplinkos stebėsena, grėsmių aptikimas ir reagavimas į jas. Vis dėl to nepaisant jos autonomiškumo, Darktrace (n.d.) deklaruoja, jog sistema yra saugi šiuo atžvilgiu ir užtikrina pakankamą skaidrumą dėl to viešai pateikiama atskira informacija apie sistemos saugumą, atitikties ir privatumo užtikrinimą *(angl. Trust centre)* (Darktrace, n.d.). Šis atvejis iš esmės rodo, kad šiame modelyje dirbtiniam intelektui suteikiamas kur kas aktyvesnis vaidmuo, o pati sistema orientuota į greitą ir savarankiškesnį veikimą. Šią išvadą papildomai pagrindžia ir paslaugos aprašymas Digital Marketplace aplinkoje, kuriame pabrėžiama šios sistemos savarankiško mokymosi technologija, anomalios elgsenos aptikimas, nuolatinis tyrimas ir autonominio atsako realiuoju laiku galimybės (UK Digital Marketplace, n.d.).

SecOps pagal šį kriterijų ir vėl užima tarpinę poziciją. Google Cloud (2026) nurodo, kad sistema gali padėti tirti įspėjimus, rengti grėsmių santraukas ir suprantamu bei paprastu būdu siūlyti galimus riziką mažinančius veiksmus, tačiau šie rezultatai pirmiausia skirti saugumo specialistų darbui palengvinti, o ne automatiniam sprendimų priėmimui. Vadinasi, dirbtinis intelektas čia labai aktyviai prisideda prie analizės, bet galutinis sprendimas vis tiek lieka žmogaus atsakomybė. Kalbant apie skaidrumą, Google Cloud (2026) pakankamai aiškiai aprašo technines ir organizacines saugumo priemones, susijusias su prieigos valdymu, duomenų prieinamumu, saugojimo terminais ir kt. (Google Cloud, 2026). Tai leidžia teigti, kad šiame modelyje siekiama suderinti dirbtinio intelekto teikiamą pagalbą, žmogaus priežiūrą ir duomenų apsaugos principus.

Apibendrinant galima teigti, kad Security Copilot pasižymi mažiausiu savarankiškumo lygiu ir stipriausia žmogaus kontrole, Darktrace išsiskiria kur kas didesniu autonomiškumu ir aktyvesniu vaidmeniu reaguojant į aptinkamas grėsmes, tuo tarpu SecOps tarp jų užima tarpinę poziciją, nes dirbtinis intelektas čia reikšmingai padeda analizuoti informaciją, tačiau sprendimų priėmimas visada išlieka žmogaus rankose. Manytina, kad šis skirtumas yra reikšmingas, nes jis parodo, kiek atsakomybės sistemų kūrėjai perduoda pačiam dirbtiniam intelektui ir kiek jos palieka žmogaus kontrolei, ko pasėkoje organizacijai lengviau pasirinkti sprendimą pagal jos siekiamą tikslą.

Apibendrintas palyginimas pateikiamas toliau lentelėje (žr. 5 lentelę).

5 lentelė. Žmogaus priežiūros, autonomiškumo ir saugumo priemonių palyginimas.

Sistema	Žmogaus vaidmuo	Autonomiškumo lygis	Pagrindinės saugumo priemonės
Security Copilot	Žmogus visada yra sistemos kontrolės grandyje.	Žemas–vidutinis	Mažiausių prieigos teisių principas, stiprus autentifikavimas, vaidmenimis grindžiama prieiga, klientų duomenys nenaudojami sistemos modelio mokymui.
Darktrace	Žmogus labiau prižiūri sistemą, nei tiesiogiai valdo kiekvieną veiksmą.	Aukštas	Nuolatinė stebėsena, Trust Centre su išsamia dokumentacija, operacinis atsparumas, skaidrumas dėl techninių ir organizacinių saugumo priemonių.
SecOps	Žmogus visada priima galutinius sprendimus.	Vidutinis	Prieigos teisės prie sistemos, apibrėžtas duomenų tvarkymas, aiškiai nustatyti duomenų saugojimo terminai, įdiegtas informacijos naudojimo kontrolės mechanizmas.

Šaltinis: sudaryta autorės remiantis Microsoft (2026), Darktrace (n.d.) ir Google Cloud (2026).

3.2.2. Lyginamosios analizės apibendrinimas

Atlikta lyginamoji dirbtinio intelekto sistemų, naudojamų kibernetinio saugumo užtikrinimui, atvejų analizė parodė, kad visos nagrinėtos sistemos, nepaisant savo skirtumų, turi ir bendrų bruožų. Pirmiausia, jas sieja vienas bendras tikslas stiprinti kibernetinį saugumą, padėti aptikti grėsmes, jas tirti ir, prireikus, į jas tinkamai reaguoti. Visos aukščiau analizuotos sistemos taip pat remiasi duomenų analize, yra integruojamos į platesnę organizacijos saugumo aplinką ir (ar) su kitais išoriniais sprendimais, ir bent iš dalies automatizuoja tam tikras specialistų atliekamas užduotis. Bendras bruožas pasireiškia ir tuo, kad kiekvienu atveju žmogus išlieka svarbia saugumo proceso dalimi, nors jo vaidmuo skirtingose sistemose gali būti skirtingas. Visuose aptartuose sprendimuose taip pat matyti dėmesys prieigos kontrolei, duomenų apsaugai ir saugiam sistemos veikimui.

Vis dėl to, kartu analizė atskleidė ir reikšmingų skirtumų tarp šių sistemų. Jie labiausiai išryškėjo vertinant sistemų paskirtį, jos veikimo logiką, duomenų naudojimo pobūdį, autonomiškumo lygį ir taikomas saugumo priemones. Iš analizuotų atvejų buvo nustatyta, kad Security Copilot labiau orientuotas į specialisto darbo palaikymą, informacijos apibendrinimą ir greitesnį sprendimų priėmimą; Darktrace daugiau dėmesio skiria nuolatinei stebėsenai, galimų grėsmių atpažinimui ir aktyvesniam reagavimui

realiuoju laiku, o SecOps yra tarpinis variantas, kur sistema ne tik padeda analizuoti informaciją, bet ir yra glaudžiai įtraukta į platesnį saugumo operacijų valdymo procesą.

Vienas ryškiausių nustatytų skirtumų susijęs su tuo, kiek savarankiškumo suteikiama pačiai sistemai priimti atitinkamus sprendimus. Pavyzdžiui, Security Copilot atveju žmogus aiškiai išlieka pagrindinis kontrolės subjektas; Darktrace modelyje dirbtinis intelektas veikia savarankiškumo prasme daug aktyviau, tad ir šios platformos autonomiškumo lygis yra didesnis palyginus; o SecOps šiuo požiūriu vėlgi yra tarpinėje pozicijoje, nes sistema reikšmingai prisideda prie analizės ir tyrimo proceso, tačiau vis tiek galutinis sprendimas paliekamas žmogui priimti. Visi šie skirtumai geriausiai atspindi, kiek atsakomybės perduodama dirbtinio intelekto sistemai ir kiek jos lieka žmogaus priežiūroje bei atsakomybėje.

Apibendrinant gautus rezultatus, galima teigti, kad dirbtinio intelekto sistemų kūrimas kibernetinio saugumo kontekste negali būti grindžiamas vienu universaliu modeliu, kuris tiktų visiems kibernetinio saugumo tikslams pasiekti. Nepaisant to, kad bendras tikslas yra kibernetinio saugumo užtikrinimas, vis dėl to konkretūs kūrimo sprendimai priklauso nuo atitinkamos sistemos funkcijos, naudojamų duomenų pobūdžio, integracijos su kitais organizacijos sprendimais ir jai suteikiamo savarankiškumo lygio. Iš atliktos analizės galima daryti išvadą, kad dirbtinio intelekto sistemų kūrimo ypatumai kibernetinio saugumo srityje būtent ir pasireiškia per jų skirtingą kūrimo logiką, kad kiekviena organizacija turėtų galimybę rinktis iš tokių sprendimų įvairovės, kuris labiausiai atitiktų jos saugumo politiką, keliamus tikslus, infrastruktūrą ir priimtinas technines bei organizacines saugumo priemones.

Siekiant aiškiau susisteminti atliktos lyginamosios analizės rezultatus, toliau lentelėje (žr. 6 lentelę) pateikiami pagrindiniai bendri ir skirtingi analizuotų sistemų kūrimo požymiai kibernetinio saugumo kontekste.

6 lentelė. Lyginamosios analizės apibendrinimas.

Analizės aspektas	**Bendri požymiai**	**Skirtingi požymiai**
Paskirtis	Visos sistemos skirtos padėti aptikti, tirti ar valdyti kibernetines grėsmes.	Security Copilot labiau orientuota į specialisto darbą, Darktrace į aktyvesnį reagavimą, o SecOps į bendro saugumo proceso stiprinimą.
Dirbtinio intelekto taikymo logika	Visose sistemose dirbtinis intelektas naudojamas tam, kad saugumo analizė vyktų greičiau ir efektyviau.	Security Copilot pagrinde veikia kaip pagalbininkas, Darktrace kaip aktyvi stebėjimo ir reagavimo platforma, o SecOps kaip platesnės saugumo sistemos dalis.
Duomenų naudojimas	Visos sistemos remiasi saugumo duomenimis ir jų analize.	Security Copilot remiasi jau turima organizacijos informacija pagal prieigos teises, Darktrace nuolatiniu organizacijos aplinkos ir veiksmų stebėjimu, o SecOps didesnės apimties saugumo duomenų kaupimu ir valdymu.
Integracija	Visos sistemos yra susietos su platesne organizacijos saugumo aplinka.	Security Copilot susieta su savo saugumo sistemomis, Darktrace su vidinėmis ar išorinėmis tinklo ir saugumo sistemomis, o SecOps su platesne saugumo valdymo platforma.

Analizės aspektas	Bendri požymiai	Skirtingi požymiai
Žmogaus vaidmuo	Visais atvejais žmogus išlieka svarbi proceso dalis.	Security Copilot atveju žmogus išlieka kontrolės centre, Darktrace atveju žmogus daugiau priežiūri nei tiesiogiai daro įtaką, o SecOps atveju žmogus priima tik galutinį sprendimą.
Autonomiškumas	Visose sistemose yra bent dalinis automatizavimas.	Security Copilot savarankiškumo lygis mažiausias, Darktrace didžiausias, o SecOps vidutinis.
Saugumo priemonės	Visose sistemose svarbi prieigos teisių kontrolė, duomenų apsauga ir saugus veikimas.	Security Copilot yra griežtas prieigų teisių valdymas, Darktrace turi nuolatinę stebėseną ir sistemos atsparumą, o SecOps aiškų prieigos tiesių ir duomenų valdymas.
Bendras kūrimo modelis	Visos sistemos kuriamos siekiant stiprinti kibernetinį saugumą pasitelkiant dirbtinį intelektą.	Security Copilot atitinka pagalbininko modelį, Darktrace aktyvaus reagavimo modelį, SecOps platesnės saugumo sistemos modelį.

Šaltinis: sudaryta autorės remiantis Microsoft (2026), Darktrace (n.d.) ir Google Cloud (2026).

3.3. Rekomendacinės gairės kuriant dirbtinio intelekto sistemas kibernetinio saugumo kontekste

Lyginamosios atvejų analizės rezultatai leidžia daryti išvadą, kad kuriant dirbtinio intelekto sistemas kibernetinio saugumo srityje neužtenka vien tik remtis tokios sistemos pažangumo kriterijumi. Šiuo atveju svarbu ne tik tai, koks dirbtinio intelekto modelis bus naudojamas, bet ir tai, kokiam tikslui sistema bus naudojama, kaip bus taikoma praktikoje, kokius duomenis naudos, kiek savarankiškai galės atlikti tam tikras užduotis bei kaip bus užtikrinamas tokios sistemos saugumas kruopščiai atrenkant tinkamas saugumo priemones. Išanalizuoti atvejai taip pat atskleidė, kad dirbtinio intelekto sistemos kibernetinio saugumo srityje gali būti kuriamos labai skirtingai. Vienos jų gali labiau padėti specialistams analizuoti informaciją ir greičiau priimti sprendimus, kitos gali būti skirtos nuolat stebėti organizacijos veiksmus bei aplinką ir atitinkamai reaguoti į grėsmes, o dar kitos gali veikti kaip platesnės saugumo operacijų sistemos dalis. Tai yra svarbūs aspektai, nes organizacijos turėtų galimybę rinktis iš galimų sprendimų įvairovės pagal tai, kas jiems labiausiai tinkama ir atitinka jų saugumo politiką. Atsižvelgiant į tai, galima teigti, kad dirbtinio intelekto sistemos užtikrinančios kibernetinį saugumą negali būti kuriamos pagal vieną bendrą modelį. Kiekvienu atveju turi būti įvertinta, kokią konkrečią funkciją dirbtinio intelekto sistema atliks, kokio lygio kontrolės jai prireiks bei ar jos priemonės ir galimybės yra pakankamos konkrečiai organizacijai. Pažymėtina, kad požiūris atitinka ir tarptautinėse gairėse pabrėžiamą mintį, kad dirbtinio intelekto sistemos turi būti kuriamos atsižvelgiant į rizikas, naudojimo kontekstą ir patikimumo reikalavimus (Tabassi, 2023; NCSC ir CISA, 2023; ENISA, 2023).

Remiantis atliktu tyrimu, galima išvesti kelias pagrindines gaires, į kurias reikėtų atsižvelgti kuriant dirbtinio intelekto sistemas kibernetinio saugumo srityje.

Pirma, dar prieš kuriant sistemą reikia aiškiai apsibrėžti jos paskirtį. Atliktas tyrimas parodė, kad skirtingos dirbtinio intelekto sistemos kibernetinio saugumo srityje gali spręsti skirtingas problemas. Pavyzdžiui, viena sistema gali būti skirta padėti analitikams greičiau suprasti incidentą, kita gali stebėti tinklo elgseną ir aptikti anomalijas, o trečia bus skirta palengvinti bendrą saugumo operacijų valdymą. Iš to daroma išvada, jog konkrečiai organizacijai pirmasis klausimas neturėtų būti, kokį dirbtinio intelekto modelį pasirinkti, o reikėtų aiškiai atsakyti, kokią problemą sistema turi išspręsti, o tuomet jau svarstyti apie konkretų modelį. Jeigu sistemos paskirtis nėra tiksliai apibrėžta, kyla rizika, kad ji bus kuriama pernelyg abstrakčiai ir vėliau nebus aišku, kaip ją pritaikyti praktikoje. Taigi, sistemos tikslas, naudojimo aplinka ir galimos rizikos turėtų būti įvertinti dar pradiniame sistemos kūrimo etape. Tai padėtų išvengti situacijos, kai technologija pasirenkama pirmiau, o jos praktinis pritaikymas apgalvojamas tik vėliau.

Antra, sistemos architektūra turėtų būti pasirenkama pagal jos paskirtį. Atlikta analizė leido suprasti, kad dirbtinio intelekto sistemos kibernetinio saugumo srityje gali veikti skirtingais principais. Vienais atvejais naudojamas generatyvinis dirbtinis intelektas, kitais savarankiško mokymosi technologijos ir elgsenos analizė, dar kitais dirbtinis intelektas gali būti integruojamas į platesnę saugumo operacijų platformą. Iš esmės tai reiškia, kad nėra vieno visiems atvejams tinkamo sprendimo. Jeigu sistema kuriama kaip specialisto pagalbininkas, svarbios gali būti informacijos paieškos, apibendrinimo, paaiškinimo ir rekomendacijų teikimo funkcijos. Tuo tarpu, jeigu siekiama nuolatinės stebėsenos ir greito reagavimo į grėsmes, gali būti reikalinga visai kitokia architektūra, paremta realaus laiko duomenų analize ir automatizuotu atsaku. Šiuo požiūriu svarbu nepamiršti, kad technologija neturėtų būti pasirenkama vien dėl to, kad ji yra nauja ar populiari. Ji turėtų būti skirta ir tinkama konkrečiai funkcijai atlikti. Kitaip tariant, sistemos techninis sprendimas turi sekti iš jos paskirties, o ne atvirkščiai.

Trečia, ypatingas dėmesys turi būti skiriamas duomenims. Dirbtinio intelekto sistema negali veikti be duomenų, tačiau vien didelis jų kiekis savaime dar nereiškia, kad sistema veiks patikimai. Tyrimas atskleidė, kad analizuotos sistemos gali skirtis tuo, kokius duomenis naudoja ir kaip juos susieja su savo veikimo logika. Vienos sistemos remiasi pačios organizacijos saugumo duomenimis, kitos nuolat analizuoja tos organizacijos veiksmus ir aplinkos elgseną, o dar kitos naudoja platesnį saugumo operacijų kontekstą. Kuriant tokią sistemą būtina aiškiai žinoti, kokie duomenys jai bus reikalingi ir kokiems tikslams pasiekti, iš kur jie bus gaunami, įvertinti jų kokybė ir kas gali prie jų prieiti. Taip pat svarbu atsižvelgti į tai, ar sistema nenaudos perteklinių duomenų ir ar jų tvarkymas nesukels papildomų saugumo rizikų asmens duomenų saugumo prasme. Būtent duomenų tvarkymas praktikoje gali tapti viena jautriausių vietų. Jeigu duomenys bus netikslūs, nepakankamai prižiūrimi arba bus per plačiai prieinami, sistema gali pateikti klaidingus rezultatus arba pati tapti papildomu rizikos šaltiniu. Atitinkamai, duomenų kokybė, jų kilmės aiškumas, prieigos kontrolė ir apsauga nuo manipuliacijų turėtų būti laikomi būtina patikimos sistemos sąlyga. Panaši mintis pabrėžiama ir mokslinėje literatūroje, kur dirbtinio intelekto sistemų patikimumas siejamas su duomenų kokybe ir jų apsauga (Junklewitz ir kt., 2023; Polito ir Pupillo, 2024).

Ketvirta, reikia aiškiai nustatyti, kiek savarankiškai sistema gali veikti. Vienas svarbiausių klausimų kuriant dirbtinio intelekto sistemą yra jos autonomiškumo lygis. Analizės rezultatai leido pamatyti, kad vienose sistemose žmogus išlieka pagrindinis sprendimų priėmėjas, o dirbtinis intelektas tik padeda analizuoti informaciją. Kitose sistemose dirbtinis intelektas veikia kur kas savarankiškiau, pavyzdžiui, pats aptinka saugumo grėsmes ar inicijuoja reagavimo į juos veiksmus. Iš to daroma išvada, jog neužtenka tik pasakyti, kad sistema bus automatizuota arba pusiau automatizuota. Reikia aiškiai apibrėžti, kokius veiksmus ji gali atlikti pati, kada ji turi tik pateikti rekomendaciją, o kada jau yra būtinas žmogaus galutinis sprendimas. Tai ypač svarbu tais atvejais, kai sistema gali turėti tiesioginį poveikį organizacijos saugumo procesams. Kibernetinio saugumo srityje pernelyg didelis pasitikėjimas automatizuotais sprendimais gali būti labai rizikingas. Jeigu specialistai pradeda manyti, kad sistema viską atliks pati už juos, gali sumažėti jų budrumas ir kritinis sistemos rezultatų vertinimas. Būtent dėl to žmogaus priežiūra turėtų būti reali, o ne tik formaliai nurodyta dokumentuose. Ši problema aptariama ir moksliniuose šaltiniuose, kur pabrėžiama, kad žmogaus įsitraukimas išlieka svarbus net ir naudojant pažangias dirbtinio intelekto sistemas (Jabbarova, 2023; Polito ir Pupillo, 2024).

Penkta, turi būti saugomas ne tik dirbtinio intelekto modelis, bet ir visa sistema. Atliktas tyrimas leidžia daryti išvadą, kad dirbtinio intelekto saugumas neturėtų būti suprantamas per siaurai, o priešingai, turi būti skiriamas ypatingas dėmesys. Neužtenka užtikrinti, kad pats modelis veiktų tiksliai ar būtų apsaugotas nuo klaidų. Kibernetinio saugumo srityje dirbtinio intelekto sistema paprastai veikia kartu su kitais organizacijos sprendimais, todėl svarbi tampa visa jos veikimo aplinka. Tai reiškia, kad reikia saugoti duomenų srautus, naudotojų prieigas, integracijas, papildinius, veiksmų žurnalus, išorinius komponentus ir reagavimo mechanizmus. Pažeidžiamumas gali atsirasti ne tik pačiame modelyje, bet ir bet kurioje kitoje sistemos dalyje, pavyzdžiui, netinkamai sukonfigūruotoje prieigoje ar nesaugioje integracijoje. Taigi, kuriant dirbtinio intelekto sistemą svarbu vertinti ne tik jos rezultatą, bet ir visą techninę bei organizacinę aplinką. Šią mintį pabrėžia ir Europos Komisijos Jungtinis tyrimų centras, nurodydamas, kad kibernetinio saugumo reikalavimai turėtų būti taikomi visai dirbtinio intelekto sistemai, o ne tik jos modeliui (Junklewitz ir kt., 2023).

Šešta, sistema turi būti nuolat prižiūrima ir atnaujinama. Dirbtinio intelekto sistema kibernetinio saugumo srityje neturėtų būti suprantama kaip vieną kartą sukurtas ir įdiegtas produktas. Kibernetinės grėsmės nuolat keičiasi ir stiprėja, todėl ir pati sistema turi būti reguliariai prižiūrima, tikrinama bei atnaujinama. Tai reiškia, kad jau kūrimo etape reikėtų numatyti veiksmus, kaip sistema bus valdoma ir atnaujinama po jos įdiegimo. Šiuo atveju reikėtų apgalvoti, kas stebės jos veikimą, kaip bus vertinami klaidingi rezultatai, kaip bus koreguojamos taisyklės, tikrinamos integracijos ir prisitaikoma prie naujų kibernetinių grėsmių. Be tokios priežiūros net ir pažangi sistema ilgainiui gali tapti mažiau veiksminga arba pradėti kelti papildomų rizikų. Atsižvelgiant į tai, dirbtinio intelekto sprendimas turėtų būti suprantamas kaip nuolat prižiūrima ir atnaujinama sistema, o ne kaip galutinis produktas, kurio po

įdiegimo nebereikia aktyviai prižiūrėti. Toks požiūris atitinka ir saugaus dirbtinio intelekto gyvavimo ciklo idėją, kuri pabrėžiama NCSC gairėse ir Jungtinės Karalystės praktikos kodekse (NCSC ir CISA, 2023; UK Government, 2025).

Apibendrinant kas buvo pasakyta, galima teigti, kad atliktas atvejų tyrimas leidžia išskirti kelis svarbiausius klausimus, į kuriuos reikėtų atsakyti kuriant dirbtinio intelekto sistemas kibernetinio saugumo srityje, t.y. kokiai funkcijai sistema bus kuriama, kokia architektūra jai tinkamiausia, kokiais duomenimis ji remsis, kiek savarankiškai galės veikti, kaip bus užtikrinamas jos saugumas ir kaip ji bus prižiūrima bei atnaujinama po įdiegimo. Iš to daroma išvada, jog efektyvi dirbtinio intelekto sistema neturėtų būti kuriama kaip universali technologija, kurią vėliau bandoma pritaikyti prie atitinkamos organizacijos poreikių. Priešingai, veiksminga dirbtinio intelekto sistema turėtų būti kuriama kaip aiškiai apibrėžtas, konkrečiai problemai spręsti skirtas ir tinkamai valdomas sprendimas. Tai leistų geriau išnaudoti dirbtinio intelekto galimybes ir kartu sumažinti su jo taikymu susijusias saugumo, valdymo bei atsakomybės iššūkius.

IŠVADOS

1. Dirbtinio intelekto sistemų sąvoka pasaulyje nėra suprantama vienodai ir gali būti aiškinama skirtingai, priklausomai nuo teisinio, technologinio ar bendrojo konteksto. Kuriant tokias sistemas svarbu ne tik apibrėžti pačią sistemą, bet ir įvertinti, kokiais metodais ji bus grindžiama, nes nuo jų priklauso sistemos veikimo principas.
2. Kibernetinio saugumo srityje dirbtinis intelektas gali būti naudojamas grėsmių aptikimui, rizikos analizei, elgsenos vertinimui, incidentų valdymui, prevencijai, autentifikavimo ir prieigos kontrolės stiprinimui. Vis dėl to tos pačios technologijos gali būti naudojamos ir kenkėjiškiems tikslams, todėl šių sistemų kūrimas turi būti siejamas ne tik su technologine pažanga, bet ir su saugiu, atsakingu bei valdomu jų taikymu.
3. Dirbtinio intelekto sistemų kūrimas kibernetinio saugumo kontekste yra susijęs su technologinėmis, organizacinėmis, teisinėmis ir etinėmis rizikomis. Šios rizikos pasireiškia dėl galimų sistemos pažeidžiamumų, pernelyg didelio pasitikėjimo automatizuotais sprendimais, žmogaus ir sistemos sąveikos problemų, duomenų kokybės ir apsaugos klausimų, privatumo, atsakomybės bei reguliavimo neapibrėžtumo. Atsižvelgiant į tai, jų veikimas turi būti ne tik techniškai efektyvus, bet ir saugus, skaidrus, prižiūrimas bei praktiškai valdomas.
4. Pasauliniu mastu reikalavimai, taikomi dirbtinio intelekto sistemų kūrimui kibernetinio saugumo srityje, nėra vienodi. Europos Sąjungoje labiau vyrauja privalomomis teisės normomis grindžiamas reguliavimas, o JAV, Jungtinėje Karalystėje, Singapūre, Australijoje ir Indijoje didesnė reikšmė teikiama gairėms, principams ir gerajai praktikai. Nepaisant šių skirtumų, išskirtini bendri reikalavimai, kurie yra svarbūs kuriant tokias sistemas: rizika grindžiamas požiūris, saugumas nuo kūrimo pradžios, visos sistemos apsauga, duomenų konfidencialumas ir vientisumas, prieigos kontrolė, dokumentavimas, auditas, reguliarūs atnaujinimai, tiekimo grandinės saugumas ir žmogaus priežiūra.
5. Pasaulinės rinkos tendencijų ir pasirinktų praktinių atvejų analizė parodė, kad dirbtinio intelekto sprendimų taikymas kibernetinio saugumo srityje stiprėja, tačiau jų kūrimas negali būti grindžiamas vien tik inovatyvumu ar universalaus sprendimo paieška. Tokios sistemos turi būti kuriamos atsižvelgiant į konkrečią jų paskirtį, organizacijos rizikas, naudojamus duomenis, integraciją su kitais saugumo sprendimais, autonomiškumo lygį ir žmogaus priežiūros apimtį. Remiantis atlikto tyrimo rezultatais, darbe pateiktos rekomendacinės gairės, pagal kurias dirbtinio intelekto sistema kibernetinio saugumo srityje turėtų būti aiškios paskirties, tinkamos architektūros, duomenimis pagrįstas, saugiai integruotas, prižiūrimas ir praktikoje valdomas sprendimas.

PASIŪLYMAI

1. Rengiant, peržiūrint bei koreguojant tarptautinius ir nacionalinius teisės aktus, gaires ar kitus dokumentus, reglamentuojančius dirbtinio intelekto sistemų kūrimą ir taikymą, siūloma suvienodinti dirbtinio intelekto sistemos sąvoką ir ją formuluoti remiantis Europos Parlamento ir Tarybos reglamento (ES) 2024/1689 3 straipsnio 1 dalyje įtvirtinta sąvoka. Siūloma vartoti šią formuluotę: *„dirbtinio intelekto sistema – mašina pagrįsta sistema, suprojektuota veikti skirtingais autonomiškumo lygiais, kuri po įdiegimo gali pasižymėti prisitaikomumu ir kuri, siekdama aiškių arba numanomų tikslų, pagal gautą įvestį nustato, kaip generuoti rezultatus, įskaitant prognozes, turinį, rekomendacijas ar sprendimus, galinčius daryti poveikį fizinei ar virtualiai aplinkai“*.
2. Nacionaliniam kibernetinio saugumo centrui prie Krašto apsaugos ministerijos, bendradarbiaujant su kitomis už kibernetinio saugumo, duomenų apsaugos ir skaitmeninės politikos formavimą atsakingomis institucijomis, siūloma parengti praktines rekomendacines gaires dėl dirbtinio intelekto sistemų kūrimo kibernetinio saugumo užtikrinimo kontekste. Šiose gairėse siūloma nustatyti, kad prieš kuriant ar diegiant tokią sistemą turi būti įvertinama: 1) kokiai konkrečiai kibernetinio saugumo problemai spręsti sistema kuriama; 2) kokia sistemos architektūra ir dirbtinio intelekto metodai pasirenkami; 3) kokie duomenys bus naudojami, kokia jų kilmė, kokybė ir prieigos ribos; 4) kokius veiksmus sistema galės atlikti savarankiškai, o kada bus būtinas žmogaus patvirtinimas; 5) kaip bus užtikrinama visos sistemos, o ne tik modelio, apsauga; 6) kaip sistema bus prižiūrima, testuojama ir atnaujinama po jos įdiegimo.

LITERATŪROS SĄRAŠAS

Teisės aktai ir kiti normatyvinio pobūdžio šaltiniai:

1. Australian Government Department of Industry, Science and Resources. (2024). *Voluntary AI Safety Standard*. Prieiga per internetą: https://www.industry.gov.au/sites/default/files/2024-09/voluntary-ai-safety-standard.pdf;
2. Australian Signals Directorate. (n. d.). *Essential Eight explained*. Žiūrėta 2026 m. kovo 15 d. Prieiga per internetą: https://www.cyber.gov.au/business-government/asds-cyber-security-frameworks/essential-eight/essential-eight-explained;
3. CNIL. (n. d.). *Ensuring the security of AI systems development*. Žiūrėta 2026 m. kovo 15 d. Prieiga per internetą: https://www.cnil.fr/en/ensuring-security-ai-systems-development;
4. Cyber Security Agency of Singapore. (n. d.). *Launch of Guidelines and Companion Guide on Securing Artificial Intelligence Systems*. Žiūrėta 2026 m. kovo 15 d. Prieiga per internetą: https://www.csa.gov.sg/news-events/press-releases/launch-of-guidelines-and-companion-guide-on-securing-artificial-intelligence-systems/;
5. ENISA. (2023). *Cybersecurity of AI and standardisation*. Prieiga per internetą: https://www.enisa.europa.eu/publications/cybersecurity-of-ai-and-standardisation;
6. Europos Komisija. (2025, vasario 6). *Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act)*. Prieiga per internetą: https://ec.europa.eu/newsroom/dae/redirection/document/112455;
7. Europos Parlamento ir Tarybos direktyva (ES) 2022/2555 dėl priemonių aukštam bendram kibernetinio saugumo lygiui visoje Sąjungoje užtikrinti, kuria iš dalies keičiamas Reglamentas (ES) Nr. 910/2014 ir Direktyva (ES) 2018/1972 ir panaikinama Direktyva (ES) 2016/1148 (TIS 2 direktyva). Prieiga per internetą: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2555;
8. Europos Parlamento ir Tarybos reglamentas (ES) 2024/1689, kuriuo nustatomos suderintos dirbtinio intelekto taisyklės ir iš dalies keičiami reglamentai (EB) Nr. 300/2008, (ES) Nr. 167/2013, (ES) Nr. 168/2013, (ES) 2018/858, (ES) 2018/1139 ir (ES) 2019/2144 ir direktyvos 2014/90/ES, (ES) 2016/797 ir (ES) 2020/1828 (Dirbtinio intelekto aktas), 2024 m. birželio 13 d. Prieiga per internetą: https://eur-lex.europa.eu/legal-content/LT/TXT/HTML/?uri=OJ:L_202401689#tit_1;
9. Lietuvos Respublikos kibernetinio saugumo įstatymas. Prieiga per internetą: https://e-tar.lt/portal/lt/legalAct/5468a25089ef11e4a98a9f2247652cf4/asr?csrt=340014198151981820;
10. National Cyber Security Centre (NCSC) ir Cybersecurity and Infrastructure Security Agency (CISA). (2023). *Guidelines for secure AI system development*. Prieiga per internetą: https://www.ncsc.gov.uk/sites/default/files/documents/Guidelines-for-secure-AI-system-development.pdf;

11. National Institute of Standards and Technology (NIST). (2024). *Secure software development practices for generative AI and dual-use foundation models: An SSDF community profile*. Prieiga per internetą: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-218A.pdf;
12. OECD. (2024). *Explanatory memorandum on the updated OECD definition of an AI system*. Prieiga per internetą: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_3c815e51/623da898-en.pdf;
13. Press Information Bureau, Government of India. (2026). *India AI Governance Guidelines*. Prieiga per internetą: https://static.pib.gov.in/WriteReadData/specificdocs/documents/2026/feb/doc2026215790801.pdf;
14. Tabassi, E. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. National Institute of Standards and Technology. Prieiga per internetą: https://doi.org/10.6028/NIST.AI.100-1;
15. UK Government. (2025). *Code of Practice for the Cyber Security of AI*. Prieiga per internetą: https://www.gov.uk/government/publications/ai-cyber-security-code-of-practice/code-of-practice-for-the-cyber-security-of-ai.

Moksliniai šaltiniai:

16. Adewusi, A. O., Okoli, U. I., Olorunsogo, T., Adaga, E., Daraojimba, D. O. ir Obi, O. C. (2024). Artificial intelligence in cybersecurity: Protecting national infrastructure: A USA review. *World Journal of Advanced Research and Reviews, 21*(1), 2263–2275. Prieiga per internetą: https://pdfs.semanticscholar.org/a19c/c68e55b791c3c0156f8c96eb15287476c2fe.pdf;
17. Akhtar, Z. B. ir Rawol, A. T. (2024). Enhancing cybersecurity through AI-powered security mechanisms. *IT Journal Research and Development, 9*(1), 50–67. Prieiga per internetą: https://journal.uir.ac.id/index.php/ITJRD/article/view/16852/7172;
18. Binns, R., Veale, M., Van Kleek, M. ir Shadbolt, N. (2018). 'It's reducing a human being to a percentage': Perceptions of justice in algorithmic decisions. *Proceedings of Machine Learning Research, 81*, 1–15. Prieiga per internetą: https://proceedings.mlr.press/v81/binns18a/binns18a.pdf;
19. Brundage, M., Avin, S., Clark, J., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Rinehart, W. ir kt. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. arXiv. Prieiga per internetą: https://arxiv.org/pdf/1802.07228;
20. Das, R. ir Sandhane, R. (2021). Artificial intelligence in cyber security. *Journal of Physics: Conference Series, 1964*(4), 042072. Prieiga per internetą: https://iopscience.iop.org/article/10.1088/1742-6596/1964/4/042072/pdf;
21. Delipetrev, B., Tsinaraki, C. ir Kostić, U. (2020). *Historical evolution of artificial intelligence: Analysis of the three main paradigm shifts in AI*. Prieiga per internetą: https://publications.jrc.ec.europa.eu/repository/handle/JRC120469;

22. Jabbarova, K. ir Jafarov, S. (2023). *AI and cybersecurity – New threats and opportunities*. Prieiga per internetą: https://www.researchgate.net/profile/Sarkhan-Jafarov/publication/376308829_AI_AND_CYBERSECURITY_-NEW_THREATS_AND_OPPORTUNITIES/links/65723144ea5f7f02054cec06/AI-and-Cybersecurity-New-Threats-and-Opportunities.pdf;
23. Katsoulacos, Y. (2021). Algorithmic collusion and oligopoly theory. *Journal of Antitrust Enforcement, 9*(2), 270–297. Prieiga per internetą: https://academic.oup.com/antitrust/article/9/2/270/5903500;
24. Kayode, B. F., Adebola, N. ir Akerele, S. (2025). The state of AI-driven cybersecurity: Trends, challenges and opportunities. *Journal of Artificial Intelligence, Machine Learning and Data Science, 3*(2), 2731–2739. Prieiga per internetą: https://d1wqtxts1xzle7.cloudfront.net/124021355/2_AI_577-libre.pdf;
25. Khan, M. I., Arif, A. ir Khan, A. R. A. (2024). The most recent advances and uses of AI in cybersecurity. *BULLET: Jurnal Multidisiplin Ilmu, 3*(4), 566–578. Prieiga per internetą: https://www.researchgate.net/profile/Muhammad-Ismaeel-Khan/publication/390740851_The_Most_Recent_Advances_and_Uses_of_AI_in_Cybersecurity/links/67fb8256bd3f1930dd5d5d8f/The-Most-Recent-Advances-and-Uses-of-AI-in-Cybersecurity.pdf;
26. Kritikos, M. (2019). *Artificial intelligence ante portas: Legal and ethical reflections*. Prieiga per internetą: https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/634427/EPRS_BRI(2019)634427_EN.pdf;
27. Lazić, L. (2019). *Benefit from AI in cybersecurity*. Prieiga per internetą: https://www.researchgate.net/profile/Ljubomir-Lazic/publication/336826190_BENEFIT_FROM_AI_IN_CYBERSECURITY/links/5db42935299bf111d4ca24bf/BENEFIT-FROM-AI-IN-CYBERSECURITY.pdf;
28. Michael, K., Abbas, R. ir Roussos, G. (2023). AI in cybersecurity: The paradox. *IEEE Transactions on Technology and Society, 4*(2), 104–109. Prieiga per internetą: https://ieeexplore.ieee.org/abstract/document/10153442;
29. Polito, C. ir Pupillo, L. (2024). Artificial intelligence and cybersecurity. *Intereconomics, 59*(1), 10–13. Prieiga per internetą: https://reference-global.com/download/article/10.2478/ie-2024-0004.pdf;
30. Shanbhag, N., Dawson, M. ir Etori, N. A. (2024). Artificial intelligence's role in cybersecurity and global dynamics. *MWAIS 2024 Proceedings*. Prieiga per internetą: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1029&context=mwais2024;
31. Ucheji, C., Ekeneme, J. ir Ezekwem, C. (2025). Global trends in AI-driven cybersecurity: A systematic and bibliometric analysis. *Asian Journal of Research in Computer Science, 18*(9), 103–115. Prieiga per internetą: https://www.researchgate.net/profile/Chidiebere-

Ucheji/publication/395755991_Global_Trends_in_AI-Driven_Cybersecurity_A_Systematic_and_Bibliometric_Analysis/links/68d2be3f220a341aa14ea1c5/Global-Trends-in-AI-Driven-Cybersecurity-A-Systematic-and-Bibliometric-Analysis.pdf;

32. Yampolskiy, R. V. ir Spellchecker, M. S. (2016). *Artificial intelligence safety and cybersecurity: A timeline of AI failures*. arXiv. Prieiga per internetą: https://arxiv.org/pdf/1610.07997.

Kiti internetiniai šaltiniai:

33. Amazon Web Services. (n. d.). *AI-augmented threat actor accesses FortiGate devices at scale*. Žiūrėta 2026 m. kovo 28 d. Prieiga per internetą: https://aws.amazon.com/blogs/security/ai-augmented-threat-actor-accesses-fortigate-devices-at-scale/;
34. Darktrace. (2024). *Darktrace ActiveAI Security Platform Solution Brief. Data Sheet*. Prieiga per internetą: https://cdn.prod.website-files.com/626ff4d25aca2edf4325ff97/66cdda61d3ade4ac4f201dd3_ActiveAI%20Security%20Platform%20Solution%20Brief.pdf;
35. Darktrace. (n. d.). *ActiveAI Security Platform*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://www.darktrace.com/platform;
36. Darktrace. (n. d.). *Autonomous Response*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://www.darktrace.com/darktrace-autonomous-response;
37. Darktrace. (n. d.). *Integrations*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://www.darktrace.com/integrations;
38. Darktrace. (n. d.). *Trust Centre*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://www.darktrace.com/trust;
39. Encyclopaedia Britannica. (n. d.). *History of artificial intelligence*. Prieiga per internetą: https://www.britannica.com/science/history-of-artificial-intelligence;
40. Europos Komisija. (n. d.). *Regulatory framework on artificial intelligence*. Prieiga per internetą: https://digital-strategy.ec.europa.eu/lt/policies/regulatory-framework-ai;
41. Europos Parlamentas. (2023). *Kas yra dirbtinis intelektas ir kaip jis naudojamas?* Prieiga per internetą: https://www.europarl.europa.eu/topics/lt/article/20200827STO85804/kas-yra-dirbtinis-intelektas-ir-kaip-jis-naudojamas;
42. Europos Parlamentas. (n. d.). *Dirbtinis intelektas: grėsmės ir galimybės*. Žiūrėta 2026 m. vasario 23 d. Prieiga per internetą: https://www.europarl.europa.eu/topics/lt/article/20200918STO87404/dirbtinis-intelektas-gresmes-ir-galimybes;
43. Google Cloud. (n. d.). *Data RBAC overview*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://docs.cloud.google.com/chronicle/docs/administration/datarbac-overview;

44. Google Cloud. (n. d.). *Data retention*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://docs.cloud.google.com/chronicle/docs/about/data-retention;
45. Google Cloud. (n. d.). *Gemini in Google SecOps*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://docs.cloud.google.com/chronicle/docs/secops/gemini-secops;
46. Google Cloud. (n. d.). *Google Security Operations*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://cloud.google.com/security/products/security-operations;
47. Google Cloud. (n. d.). *Role-Based Access Control (RBAC)*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://docs.cloud.google.com/chronicle/docs/administration/rbac;
48. Google Cloud. (n. d.). *Triage and Investigation Agent*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://docs.cloud.google.com/chronicle/docs/secops/triage-investigation-agent;
49. Google Cloud. (n. d.). *Using the Gemini Case summary widget*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://docs.cloud.google.com/chronicle/docs/soar/investigate/working-with-cases/using-the-gemini-case-summary-widget;
50. GOV.UK Digital Marketplace. (n. d.). *Darktrace Active AI Security Platform*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://www.applytosupply.digitalmarketplace.service.gov.uk/g-cloud/services/827362662792006;
51. Grand View Research. (n. d.). *Artificial intelligence in cybersecurity market report*. Žiūrėta 2026 m. kovo 17 d. Prieiga per internetą: https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-cybersecurity-market-report;
52. Hernandez, G. (2023). Analyst's guide to the ActiveAI Security Platform. *Darktrace*. Prieiga per internetą: https://www.darktrace.com/blog/exploring-the-cyber-ai-loop-as-an-analyst-prevent-asm-detect;
53. IBM. (2025). *Cost of a Data Breach Report*. Prieiga per internetą: https://www.ibm.com/downloads/documents/us-en/131cf87b20b31c91;
54. IBM. (n. d.). *The history of artificial intelligence*. Prieiga per internetą: https://www.ibm.com/think/topics/history-of-artificial-intelligence;
55. ISO. (n. d.). *Artificial intelligence: What it is, how it works and why it matters*. Žiūrėta 2026 m. vasario 23 d. Prieiga per internetą: https://www.iso.org/artificial-intelligence;
56. ISO. (n. d.). *What is AI?* Žiūrėta 2026 m. vasario 23 d. Prieiga per internetą: https://www.iso.org/artificial-intelligence/what-is-ai;
57. MarketsandMarkets. (n. d.). *Generative AI cybersecurity market*. Žiūrėta 2026 m. kovo 17 d. Prieiga per internetą: https://www.marketsandmarkets.com/Market-Reports/generative-ai-cybersecurity-market-164202814.html;

58. Microsoft. (n. d.). *Get started with onboarding to Microsoft Security Copilot*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://learn.microsoft.com/en-us/copilot/security/get-started-security-copilot;
59. Microsoft. (n. d.). *Microsoft Security Copilot agents overview*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://learn.microsoft.com/en-us/copilot/security/agents-overview;
60. Microsoft. (n. d.). *Privacy and data security in Microsoft Security Copilot*. Prieiga per internetą: https://learn.microsoft.com/en-us/copilot/security/privacy-data-security;
61. Microsoft. (n. d.). *Security Copilot Agent Development Overview*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://learn.microsoft.com/en-us/copilot/security/developer/custom-agent-overview;
62. Microsoft. (n. d.). *Understand authentication in Microsoft Security Copilot*. Žiūrėta 2026 m. balandžio 3 d. Prieiga per internetą: https://learn.microsoft.com/en-us/copilot/security/authentication;
63. National Institute of Standards and Technology. (n. d.). *Artificial intelligence*. Prieiga per internetą: https://www.nist.gov/artificial-intelligence;
64. OECD. (n. d.). *OECD.AI Policy Observatory*. Žiūrėta 2026 m. kovo 15 d. Prieiga per internetą: https://oecd.ai/en/;
65. Swiss Cyber Institute. (n. d.). *The history of artificial intelligence: A timeline from Turing to today*. Žiūrėta 2026 m. vasario 23 d. Prieiga per internetą: https://swisscyberinstitute.com/blog/history-artificial-intelligence/;
66. Turner, L. (2026). *What Is Microsoft Security Copilot and How Can It Be Utilized?* The Knowledge Academy. Prieiga per internetą: https://www.theknowledgeacademy.com/blog/microsoft-security-copilot/;
67. White & Case LLP. (n. d.). *AI Watch: Global regulatory tracker – United States*. Žiūrėta 2026 m. kovo 28 d. Prieiga per internetą: https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states;
68. World Economic Forum. (2025). *Global Cybersecurity Outlook 2025*. Prieiga per internetą: https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2025.pdf;
69. World Economic Forum. (2026). *Global Cybersecurity Outlook 2026*. Prieiga per internetą: https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2026.pdf.

ANOTACIJA LIETUVIŲ IR ANGLŲ KALBOMIS

Magistro baigiamajame darbe nagrinėjami dirbtinio intelekto sistemų kūrimo ypatumai kibernetinio saugumo užtikrinimo kontekste. Siekiama atskleisti, kaip tokios sistemos kuriamos ir taikomos pasauliniu mastu, atsižvelgiant į rinkos tendencijas, reguliavimo skirtumus ir praktinius sprendimus. Darbe analizuojama dirbtinio intelekto samprata, raida ir pagrindiniai metodai, taip pat jo taikymo galimybės, rizikos ir kibernetinio saugumo reikalavimai. Siekiant įvertinti praktinius šių sistemų kūrimo ypatumus, lyginami trys atvejai: Microsoft Security Copilot, Darktrace ActiveAI Security Platform ir Google Security Operations su Gemini.

Nustatyta, kad dirbtinio intelekto sistemos kibernetinio saugumo srityje negali būti kuriamos pagal vieną universalų modelį. Jų kūrimas priklauso nuo sistemos paskirties, naudojamų duomenų, integracijos su kitais organizacijos sprendimais, autonomiškumo lygio ir žmogaus priežiūros apimties. Taip pat nustatyta, kad tokios sistemos turi būti kuriamos ne tik kaip pažangūs technologiniai sprendimai, bet ir kaip saugiai valdomos, prižiūrimos ir konkrečiam kibernetinio saugumo poreikiui pritaikytos priemonės.

Raktiniai žodžiai: dirbtinis intelektas, dirbtinio intelekto sistemos, kibernetinis saugumas, sistemų kūrimas, žmogaus priežiūra.

This master’s thesis examines the specifics of developing artificial intelligence systems in the context of ensuring cybersecurity. The aim is to reveal how such systems are developed and applied on a global scale, taking into account market trends, regulatory differences, and practical solutions. The thesis analyzes the concept, development, and main methods of artificial intelligence, as well as its application possibilities, risks, and cybersecurity requirements. To assess the practical aspects of developing these systems, three cases are compared: Microsoft Security Copilot, Darktrace ActiveAI Security Platform, and Google Security Operations with Gemini.

It has been determined that artificial intelligence systems in the field of cybersecurity cannot be developed according to a single universal model. Their development depends on the system’s purpose, the data used, integration with other organizational solutions, the level of autonomy, and the extent of human oversight. It has also been established that such systems must be developed not only as advanced technological solutions, but also as tools that are securely managed, monitored, and tailored to specific cybersecurity needs.

Keywords: artificial intelligence, artificial intelligence systems, cybersecurity, system development, human oversight.

SANTRAUKA LIETUVIŲ KALBA

Zorina Julija (2026). *Dirbtinio intelekto sistemų kūrimo ypatumai kibernetinio saugumo užtikrinimo kontekste: pasaulinės rinkos analizė* (magistro baigiamasis darbas). Vilnius: Mykolo Romerio universitetas.

Dirbtinio intelekto sistemų taikymas kibernetinio saugumo srityje tampa vis reikšmingesnis, nes organizacijos susiduria su sudėtingėjančiomis kibernetinėmis grėsmėmis, poreikiu greičiau reaguoti į incidentus, automatizuoti saugumo procesus ir mažinti galimus nuostolius. Vis dėl to dirbtinio intelekto sistemų kūrimas šiame kontekste kelia ne tik naujas galimybes, bet ir rizikas, susijusias su duomenų apsauga, žmogaus priežiūra, autonomiškumu, reguliavimo skirtumais ir saugiu tokių sistemų veikimu. Atsižvelgiant į tai, magistro baigiamojo darbo tyrimo tikslas – atskleisti dirbtinio intelekto sistemų kūrimo ypatumus kibernetinio saugumo užtikrinimo kontekste, remiantis pasaulinės rinkos analize ir praktinių atvejų lyginamuoju vertinimu. Šiam tikslui pasiekti iškelti šie uždaviniai: išanalizuoti dirbtinio intelekto sampratą, raidą bei pagrindinius dirbtinio intelekto metodus; įvertinti dirbtinio intelekto taikymo galimybes, iššūkius, rizikas ir kibernetinio saugumo reikalavimus; ištirti pasaulinės rinkos tendencijas, susijusias su dirbtinio intelekto sprendimų taikymu kibernetinio saugumo srityje; palyginti pasirinktus dirbtinio intelekto sistemų atvejus, siekiant nustatyti jų kūrimo ypatumus kibernetinio saugumo kontekste; pateikti rekomendacines gaires dėl dirbtinio intelekto sistemų kūrimo kibernetinio saugumo užtikrinimo kontekste.

Mokslinį darbą sudaro: įvadas, teorinė dalis, tyrimo metodologija, tiriamoji dalis, išvados, rekomendacijos, literatūros sąrašas ir priedai. Pirmojoje darbo dalyje nagrinėjami dirbtinio intelekto sistemų kūrimo teoriniai aspektai: dirbtinio intelekto samprata, raida ir pagrindiniai metodai, taip pat atskleidžiamos dirbtinio intelekto taikymo galimybės kibernetinio saugumo srityje. Šioje dalyje taip pat analizuojami pagrindiniai iššūkiai, rizikos ir skirtinguose pasaulio regionuose taikomi kibernetinio saugumo reikalavimai kuriant tokias sistemas. Antrojoje darbo dalyje pateikiama tyrimo metodologija. Trečiojoje dalyje atliekamas dirbtinio intelekto sistemų kūrimo ypatumų tyrimas kibernetinio saugumo užtikrinimo kontekste: analizuojamos pasaulinės rinkos raidos kryptys, pristatomi tyrimui pasirinkti atvejai, atliekama jų lyginamoji analizė ir pateikiamos rekomendacinės gairės.

Atlikus tyrimą nustatyta, kad dirbtinio intelekto sistemų sąvoka pasauliniu mastu nėra suprantama vienodai, o šių sistemų kūrimas priklauso nuo pasirinktos sampratos, taikomų metodų ir reguliavimo aplinkos. Taip pat nustatyta, kad kibernetinio saugumo srityje dirbtinis intelektas gali būti naudojamas grėsmių aptikimui, rizikos analizei, elgsenos vertinimui, incidentų valdymui, prevencijai, autentifikavimo ir prieigos kontrolės stiprinimui. Kita vertus, tos pačios technologijos gali būti naudojamos ir kenkėjiškiems tikslams, todėl jų kūrimas turi būti siejamas ne tik su technologiniu efektyvumu, bet ir su saugiu, atsakingu bei valdomu taikymu. Tyrimo metu taip pat nustatyta, kad dirbtinio intelekto sprendimai kibernetinio saugumo srityje neturėtų būti vertinami kaip universalūs ar pasirenkami vien dėl inovatyvumo. Pasaulinės

rinkos tendencijų ir pasirinktų praktinių atvejų analizė parodė, kad tokių sistemų kūrimo ypatumai labiausiai išryškėja per sistemos paskirtį, naudojamų duomenų pobūdį, integraciją su kitais organizacijos sprendimais, autonomiškumo lygį ir žmogaus priežiūros apimtį. Palyginus Microsoft Security Copilot, Darktrace ActiveAI Security Platform ir Google Security Operations su Gemini, nustatyta, kad šios sistemos gali būti kuriamos skirtingais modeliais: kaip specialisto pagalbininkas, kaip aktyvesnė grėsmių stebėsenos ir reagavimo priemonė arba kaip platesnės saugumo operacijų sistemos dalis. Remiantis atlikto tyrimo rezultatais, darbe pateiktos rekomendacinės gairės, pagal kurias dirbtinio intelekto sistema kibernetinio saugumo srityje turėtų būti aiškios paskirties, tinkamos architektūros, duomenimis pagrįstas, saugiai integruotas, prižiūrimas ir praktikoje valdomas sprendimas.

SUMMARY IN ENGLISH

Zorina, Julija (2026). *Specifics of Developing Artificial Intelligence Systems in the Context of Ensuring Cybersecurity: An Analysis of the Global Market (Master's Thesis).* Vilnius: Mykolas Romeris University.

The application of artificial intelligence systems in the field of cybersecurity is becoming increasingly significant as organizations face increasingly complex cyber threats, the need to respond more quickly to incidents, automate security processes, and minimize potential losses. However, the development of artificial intelligence systems in this context presents not only new opportunities but also risks related to data protection, human oversight, autonomy, regulatory differences, and the secure operation of such systems. In light of this, the objective of this master's thesis is to reveal the specific characteristics of developing artificial intelligence systems in the context of ensuring cybersecurity, based on an analysis of the global market and a comparative assessment of practical cases. To achieve this objective, the following tasks have been set: to analyze the concept, development, and main methods of artificial intelligence; to assess the opportunities, challenges, risks, and cybersecurity requirements associated with the application of artificial intelligence; to investigate global market trends related to the application of artificial intelligence solutions in the field of cybersecurity; to compare selected cases of artificial intelligence systems in order to identify the specific features of their development in the context of cybersecurity; provide guidelines for the development of artificial intelligence systems in the context of ensuring cybersecurity.

The research paper consists of: an introduction, a theoretical section, research methodology, a research section, conclusions, recommendations, a bibliography, and appendices. The first part of the work examines the theoretical aspects of developing artificial intelligence systems: the concept, evolution, and main methods of artificial intelligence, as well as the potential applications of artificial intelligence in the field of cybersecurity. This section also analyzes the main challenges, risks, and cybersecurity requirements applied in different regions of the world when developing such systems. The second part of the thesis presents the research methodology. The third section examines the specifics of developing artificial intelligence systems in the context of ensuring cybersecurity: it analyzes global market trends, presents case studies selected for the research, conducts a comparative analysis of them, and provides recommendations. The study found that the concept of artificial intelligence systems is not uniformly understood globally, and the development of these systems depends on the chosen concept, the methods applied, and the regulatory environment. It was also found that in the field of cybersecurity, artificial intelligence can be used for threat detection, risk analysis, behavioral assessment, incident management, prevention, and strengthening authentication and access control. On the other hand, the same technologies can also be used for malicious purposes, so their development must be linked not only to technological effectiveness but also to safe,

responsible, and controlled application. The study also found that artificial intelligence solutions in the field of cybersecurity should not be viewed as universal or chosen solely for the sake of innovation. An analysis of global market trends and selected case studies showed that the specific characteristics of developing such systems are most evident in the system's purpose, the nature of the data used, integration with other organizational solutions, the level of autonomy, and the extent of human oversight. A comparison of Microsoft Security Copilot, Darktrace ActiveAI Security Platform, and Google Security Operations with Gemini revealed that these systems can be designed using different models: as an expert assistant, as a more active threat monitoring and response tool, or as part of a broader security operations system. Based on the results of the study, this paper presents recommended guidelines stating that an artificial intelligence system in the field of cybersecurity should be a solution with a clear purpose, appropriate architecture, data-driven, securely integrated, monitored, and managed in practice.