

Vilniaus universiteto Kauno fakulteto
Socialinių mokslų ir taikomosios informatikos institutas

Lauros Čekavičiūtės
II kurso, Finansų technologijų
studijų šakos studentės

Magistro darbas

**Finansų įstaigose naudojamų dirbtinio intelekto sistemų kibernetinio saugumo rizikos
scenarijų generavimo metodas**

Vadovas: Doc., Dr. mult. Paulius Astromskis

Kaunas, 2024 m.

Turinys

PAVEIKSLŲ SĄRAŠAS	5
LENTELIŲ SĄRAŠAS	6
SANTRAUKA	7
IVADAS	8
1. DIRBTINIO INTELEKTO SISTEMŲ SAMPRATA IR TAIKYMAS FINANSŲ ĮSTAIGOSE.....	13
1.1. Dirbtinio intelekto sistemų samprata.....	13
1.2. Finansų rinkos struktūra	18
1.3. Dirbtinio intelekto sistemų vystymosi tendencijos finansų srityje.....	21
1.4. Dirbtinio intelekto taikymo finansinių paslaugų saugumui apžvalga	25
2. REIKALAVIMAI RIZIKOS SCENARIJŲ SUDEDAMOSIOMS DALIMS	28
2.1. Dirbtinio intelekto reguliacinė aplinka.....	28
2.2. Dirbtinio intelekto rizikos valdymo standartai	33
2.3. Reikalavimai rizikos scenarijams.....	36
3. DIRBTINIO INTELEKTO SISTEMŲ RIZIKOS SCENARIJŲ GENERAVIMO METODOLOGIJA	38
3.1. Rizikos scenarijaus struktūra ir elementų taksonomija	38
3.2. Generatyvinių DI įrankių palyginimas	43
3.2.1. ChatGPT-4	43
3.2.2. Google Gemini 1.5 Flash	44
3.2.3. IBM Watson AI.....	45
3.2.4. Microsoft Copilot	46
3.2.5. Generatyvinių DI įrankių palyginimas	47
3.3. Užklauskos konteksto formavimas rizikos scenarijams generuoti.....	50
3.4. Užklauskos standartas dirbtinio intelekto rizikos scenarijams generuoti.....	52
4. DI SISTEMOS RIZIKOS SCENARIJŲ GENERATORIUS	59
4.1. Generatoriaus kūrimas.....	59
4.2. Generatoriaus apmokymo procesas.....	63
4.3. Generatoriaus vartotojo sąsaja	67
4.4. Generatoriaus saugumo ir privatumo apsauga	70
5. EKSPERIMENTAS SU SUKURTU GENERATORIUMI	72
5.1. Hipotetinės įmonės.....	72
5.1.1. Kredito ir mokėjimų rinkos hipotetinė įmonė	73
5.1.2. Draudimo rinkos hipotetinė įmonė.....	74
5.1.3. Finansinių priemonių rinkos hipotetinė įmonė.....	74

5.1.4. Kriptoturto rinkos hipotetinė įmonė	75
5.2. Eksperimentas su generatoriumi ir hipotetinėmis įmonėmis.....	76
5.2.1. Eksperimentas su kredito ir mokėjimų rinkos hipotetinė įmone	77
5.2.2. Eksperimentas su draudimo rinkos hipotetinė įmone.....	79
5.2.3. Eksperimentas su finansinių priemonių rinkos hipotetinė įmone.....	83
5.2.4. Eksperimentas su kriptoturto rinkos hipotetinė įmone.....	87
5.3. Eksperimentas su kibernetinio saugumo ekspertais ir hipotetinėmis įmonėmis	91
5.3.1. Eksperimentas su kredito ir mokėjimų rinkos hipotetinė įmone	91
5.3.2. Eksperimentas su draudimo rinkos hipotetinė įmone.....	94
5.4. Apibendrinimas	96
IŠVADOS.....	98
LITERATŪROS ŠALTINIAI.....	100
1 priedas. Priešiškų grėsmių įvykiai	108
2 priedas. Ne priešiškų grėsmių įvykiai.....	113
3 priedas. Turto taksonomija 1	114
4 priedas. Grėsmių taksonomija	115
5 priedas. Turto taksonomija 2	129
6 priedas. Konteksto klausimynas	140
7 priedas. Sukurto generatoriaus vertinimo kriterijai.....	143
8 priedas. Generatoriaus apmokymo instrukcijos.....	144
9 priedas. 1 eksperimentas su kredito ir mokėjimų rinkos hipotetinė įmone	147
10 priedas. 2 eksperimentas su draudimo rinkos hipotetinė įmone.....	149
11 priedas. 3 eksperimentas su finansinių priemonių rinkos hipotetinė įmone.....	151
12 priedas. 4 eksperimentas su kriptoturto rinkos hipotetinė įmone	153
13 priedas. NATO kibernetinio saugumo inžinieriaus parašyti trys labiausiai tikėtini kibernetinio saugumo scenarijai hipotetinei įmonei	158
14 priedas. Nacionalinio kibernetinio saugumo centro Tarptautinio kibernetinio grėsmių analizės skyriaus vedėjo parašyti trys labiausiai tikėtini kibernetinio saugumo scenarijai hipotetinei įmonei.....	159

SANTRUMPŲ SĄRAŠAS

AI (angl. *Artificial intelligence*) – Dirbtinis intelektas (sutrumpintai DI).

EU (angl. *European Union*) – Europos sąjunga (sutrumpintai ES).

NIST (angl. *National Institute of Standards and Technology*) – Nacionalinis standartų ir technologijų institutas

DORA (angl. *Digital Operational Resilience Act*) – Skaitmeninės veiklos atsparumo finansų sektoriuje reglamentas)

NLP (angl. *Natural language processing*) – Natūralios kalbos apdorojimas

GDPR (angl. *General Data Protection Regulation*) – Bendrasis duomenų apsaugos reglamentas

DDoS (angl. *Distributed Denial of Service*) – Paskirstytoji aptarnavimo perkrovos ataka

MFA (angl. *Multi-factor Authentication*) – Daugiafaktorė autentifikacija

PAM (angl. *Privileged Access Management*) – Privilegijuotos prieigos valdymas

EPĮ – Elektroninių pinigų įstaigos

MI – Mokėjimo įstaigos

TDA (angl. *Topological Data Analysis*) – Topologinė duomenų analizė

PSD2 (angl. *Payment Services Directive*) – Mokėjimo paslaugų direktyva

IRT – Informacinių ir ryšių technologijos

PAVEIKSLŲ SĄRAŠAS

1 pav. Dirbtinio intelekto sistema	15
2 pav. Dirbtinio intelekto gyvavimo ciklas	16
3 pav. Finansų rinkos struktūra	19
4 pav. Dirbtinio intelekto taikymas finansinių paslaugų saugumui.	25
5 pav. Dirbtinio intelekto akto kritiškumo piramidė	29
6 pav. Rizikos saugumo valdymo procesas.....	35
7 pav. Rizikos scenarijaus sudedamosios dalys	38
8 pav. Motyvacijos elementai.....	41
9 pav. Užklausos forma grėsmių šaltinių nustatymui (angl. Actors)	54
10 pav. Užklausos forma grėsmių įvykių nustatymui (angl. Threat Events)	54
11 pav. Užklausos forma poveikio analizei (angl. Impact)	55
12 pav. Užklausos forma motyvų analizei (angl. Intentions/Motives)	55
13 pav. Užklausos forma turto analizei (angl. Assets)	56
14 pav. Užklausos forma laiko aspektams (angl. Time)	56
15 pav. Bendra užklausos forma	57
16 pav. Rizikos scenarijų generatoriaus kūrimo proceso eiga	60
17 pav. ChatGPT apmokymo schema	61
18 pav. Pagrindiniai generatoriaus komponentai ir jų tarpusavio sąveika	62
19 pav. Generatoriaus pagrindinis langas.....	65
20 pav. Generatoriaus vartotojo sąsaja.....	68
21 pav. Generatoriaus sukurtų užklausų istorija	69
22 pav. Pirmo eksperimento pateikta užklausa generatoriui	77
23 pav. Antro eksperimento pateikta užklausa generatoriui	80
24 pav. Antro eksperimento pateikta užklausa generatoriui	84

LENTELIŲ SĄRAŠAS

1 lentelė. Dirbtinio intelekto sąvokų analizė.....	14
2 lentelė. Finansų įstaigos ir jų pagrindinės funkcijos, kurioms naudojamas DI.....	24
3 lentelė. Generatyvinių DI įrankių palyginamas	48
4 lentelė. Palyginimas tarp generatoriaus ir kibernetinio saugumo specialisto sukurtų scenarijų kredito ir mokėjimų rinkos hipotetinei įmonei.....	92
5 lentelė. Palyginimas tarp generatoriaus ir kibernetinio saugumo specialisto sukurtų scenarijų draudimo rinkos hipotetinei įmonei.....	95

SANTRAUKA

Čekavičiūtė, Laura (2024). A method for generating cybersecurity risk scenarios for artificial intelligence systems in financial institutions. MBA Graduation Paper. Kaunas: Vilnius University, Kaunas Faculty, Institute of Social Sciences and Applied Informatics. 92 p.

SUMMARY

The aim of this master's thesis is to develop a cyber security risk scenario generation method for artificial intelligence (AI) systems used in financial institutions. The research seeks to identify and evaluate potential cyber security threats that arise in AI systems within the financial sector.

The first part of the thesis examines the concept of artificial intelligence, its core elements, and levels of autonomy. It explores the development of AI systems in the financial sector while highlighting emerging security challenges. Additionally, the structure of financial institutions and the regulatory environment for AI systems are analysed, along with the requirements for risk management to ensure the reliability and security of these systems.

The second part focuses on the methodology for generating cyber security risk scenarios. It discusses the fundamental components of risk scenarios, the taxonomy of scenario elements, and compares existing tools available in the market. Furthermore, it explores cyber security risk management standards, providing a foundation for systematic and structured risk modelling.

In the practical part, a cyber security risk scenario generator is implemented using the GPT-4 (OpenAI) system. This generator is designed to address the specific needs of AI systems within financial institutions. The thesis analyses the development process, training stages, user interface features, and technical characteristics of the generator. The functionality of the generator is tested through experiments involving hypothetical financial institution cases to create and evaluate realistic cyber security risk scenarios. Practical examples illustrate potential threats and highlight opportunities to strengthen the security of AI systems. This tool assists organizations in effectively managing cyber security risks, enhancing the protection of AI systems in the financial sector, and contributing to the safer application of AI technologies.

IVADAS

Temos aktualumas

Kibernetinio saugumo sritis nuolat tobulėja, tačiau kartu didėja ir kibernetinių grėsmių sudėtingumas, ypač finansų sektoriuje. Šis sektorius yra vienas pagrindinių kibernetinių atakų taikinių, dėl jame saugomų jautrių finansinių duomenų ir vykdomų didelės apimties operacijų. Siekdamas užtikrinti efektyvumą ir saugumą, finansų įstaigos vis dažniau naudoja dirbtinio intelekto (DI) technologijas – nuo operacijų optimizavimo iki klientų aptarnavimo gerinimo bei sukčiavimo prevencijos. Tačiau šių technologijų plėtra kelia ir naujus iššūkius: DI įrankių pažeidžiamumai gali tapti reikšminga grėsme sistemų saugumui, todėl būtina skirti ypatingą dėmesį jų saugumui užtikrinti (Deloitte, 2024a).

Naujausi tyrimai rodo, kad tradiciniai kibernetinio saugumo metodai nepakankamai efektyviai prisitaiko prie nuolat kintančio grėsmių kraštovaizdžio, todėl būtina taikyti pažangius ir dinamiškus rizikos valdymo įrankius (OECD, 2024a). Šiuo atžvilgiu DI pagrįstas kibernetinio saugumo rizikos scenarijų generatorius gali tapti itin reikšmingu įrankiu, kuris padėtų modeliuoti įvairius, realistiškus grėsmių scenarijus, imituojant galimas atakas ir įvertinant jų poveikį finansų įstaigų sistemoms. Tai suteiktų finansų įstaigoms galimybę ne tik identifikuoti naujus rizikos veiksnius, bet ir pasirengti prevencinėms priemonėms, didinant organizacijos atsparumą (KPMG, 2024).

Kibernetinio saugumo rizikos scenarijų generatorius taip pat gali būti aktualus, dėl galimybės mokytis iš realių duomenų ir tokiu būdu imituoti realistiškus atakų modelius. DI technologijos gali leisti generuoti didelį skaičių skirtingų atakų scenarijų, kuriuos sunku numatyti naudojant tradicinius metodus, todėl šis generatorius gali būti itin efektyvus norint identifikuoti kibernetines grėsmes, kurios galėtų nepasireikšti naudojant standartinius saugumo testus. Be to, moksliniai tyrimai rodo, kad DI sprendimai, kuriuose naudojami mašininio mokymosi ir gilaus mokymosi algoritmai, yra žymiai efektyvesni kibernetinių grėsmių aptikime, lyginant su tradiciniais būdais (Zacas Amosas, 2023).

Apibendrinant, galima teigti, kad augantis DI sistemų ir algoritmų naudojimas finansų įstaigų veikloje pabrėžia ne tik technologinių sprendimų naudą, bet ir jų saugumo iššūkius. DI pagrindu sukurtas kibernetinio saugumo rizikos scenarijų generatorius gali pasiūlyti galimybę generuoti realistiškus atakų modelius, kurie padėtų identifikuoti galimus kibernetinius pavojus, dar iki jiems pasireiškiant. Tokie sprendimai padėtų ne tik stiprinti organizacijos gebėjimą valdyti rizikas, bet ir užtikrintų finansų sektoriaus atsparumą kibernetinėms grėsmėms, prisitaikant prie nuolat kintančios kibernetinės aplinkos (Benjamin Christopher Ilg, 2024).

Problemos ištirimo lygis

Dirbtinio intelekto (DI) sistemų kibernetinio saugumo rizikos scenarijų generavimas yra svarbi, tačiau vis dar nepakankamai ištirta sritis, ypač finansų įstaigų srityje. Nors DI kibernetinio saugumo problematika yra plačiai nagrinėjama, dabartiniai tyrimai daugiausia orientuojasi į bendrus saugumo sprendimus ir grėsmių analizę, tačiau scenarijų generavimas kaip sistemingas ir automatizuotas procesas finansų sektoriuje iki šiol nėra pakankamai ištirtas.

Dauguma mokslinių tyrimų sutelkia dėmesį į kibernetinio saugumo iššūkių identifikavimą ir rizikos valdymo sistemų kūrimą. ENISA, (2020) nagrinėja DI kibernetinio saugumo grėsmes ir pateikia bendras gaires jų valdymui, tačiau šios gairės neapima specifinių metodų, skirtų rizikos scenarijų generavimui. Be to, NIST (2021a) pateikia bendrą dirbtinio intelekto rizikos valdymo struktūrą, kurioje daugiausia dėmesio skiriama rizikos vertinimo ir valdymo procesams, tačiau šioje struktūroje trūksta praktinių gairių, kaip kurti ir taikyti konkrečius dirbtinio intelekto scenarijus. De Silva & Alahakoon, (2022) taip pat akcentuoja DI gyvavimo ciklo saugumo svarbą ir kibernetinių rizikų vertinimą kiekviename etape, tačiau jų tyrime nėra nagrinėjamas scenarijų generavimas, kuris padėtų finansų įstaigoms iš anksto pasirengti kibernetinio saugumo problemų sprendimui.

Kita vertus, organizaciniai ir etiniai kibernetinio saugumo aspektai yra plačiai diskutuojami mokslinėje literatūroje. Raimondo, (2023) analizuoja rizikos valdymo strategijas, orientuotas į DI sistemų integraciją organizacijose, tačiau šie tyrimai daugiausiai koncentruojasi į strateginį valdymą, o ne į konkrečių kibernetinių grėsmių scenarijų modeliavimą ir prognozavimą. ISO/IEC 38507 (2022) taip pat pateikia DI valdymo standartus, apimančius saugumo užtikrinimo priemones, tačiau trūksta praktinių gairių, skirtų realių grėsmių imitacijai ir scenarijų kūrimui finansų sektoriuje.

Dabartinėje praktikoje kibernetinių rizikų scenarijams kurti dažniausiai taikomi grėsmių modeliavimo metodai (angl. „Threat Modelling“) bei rizikos vertinimo analizės, tokios kaip OCTAVE ir FAIR metodologijos (Cisco Press, 2017; Yusuf Enoch et al., 2021). Šie metodai remiasi grėsmių identifikavimu, vertinimu ir valdymu pagal konkrečius techninius ir organizacinius parametrus. Tačiau tradiciniai metodai turi esminių trūkumų: jie yra statiški, orientuoti į istorinius duomenis ir dažniausiai reikalauja daugiau žmogiškųjų resursų. Be to, tokios metodologijos neatsižvelgia į DI specifiką, kaip, pavyzdžiui, modelių pažeidžiamumą, duomenų šališkumą išnaudojimą ar autonominių sprendimų keliamą riziką (Dr. Emmanuel Tsukerman, 2020; Yusuf Enoch et al., 2021).

Nepakankamas tyrimų lygis taip pat atsiskleidžia ir automatizuoto kibernetinių rizikų scenarijų generavimo kontekste, kur DI sistemos būtų naudojamos ne tik grėsmių identifikavimui, bet ir scenarijų modeliavimui (Aleksas MakFarlandas, 2024). Finansų įstaigų DI sistemos reikalauja specifinių priemonių, gebančių dinamiškai ir greitai modeliuoti kibernetines grėsmes, kad būtų galima prognozuoti galimus atakų scenarijus realiuoju laiku (Microsoft, 2024c). Šiuo metu tokie įrankiai nėra išsamiai aprašyti nei moksliniuose tyrimuose, nei praktikoje. Todėl šiame tyrime siūlomas kibernetinio saugumo rizikos scenarijų generatorius finansų įstaigoms sprendžia egzistuojantį metodologinį trūkumą. Kur generatorius remtųsi dirbtinio intelekto technologijomis, leisdamas automatizuotai generuoti įvairius kibernetinių rizikų scenarijus, atsižvelgiant į finansų įstaigų veiklos specifiką bei DI sistemų pažeidžiamumus. Toks įrankis padėtų dinamiškai prisitaikyti

prie nuolat kintančių grėsmių ir pasiūlyti realiais duomenimis pagrįstus scenarijus, taip užtikrinant efektyvesnę finansų įstaigų pasirengimą kibernetinėms atakoms. Be to, šis įrankis galėtų padėti sumažinti žmoniškųjų resursų poreikį, tuo pačiu didinant saugumo priemonių greitį ir efektyvumą.

Taigi, mokslinės publikacijos ir tyrimai pateikia išsamias kibernetinių rizikų valdymo gaires, bet esamų metodų ribotumas, jų statiškumas bei nepritaikymas DI sistemų kontekstui atskleidžia poreikį naujam sprendimui. Siūlomas finansų įstaigoms skirtas kibernetinių rizikų scenarijų generatorius galėtų būti reikšmingas indėlis į šią sritį, kuris padėtų integruoti automatizuotus DI sprendimus, kurie leistų kurti dinamiškus, praktikoje pritaikomus kibernetinio saugumo rizikos scenarijus.

Darbo problema

Kaip automatizuotai generuoti finansų įstaigose naudojamų dirbtinio intelekto sistemų kibernetinio saugumo rizikos scenarijus, reikalingus šios rizikos valdymui?

Darbo objektas – Finansų įstaigose naudojamų dirbtinio intelekto sistemų kibernetinio saugumo rizikos valdymas.

Darbo tikslas – Pasiūlyti finansų įstaigose naudojamų dirbtinio intelekto sistemų rizikos scenarijų generavimo metodą, kuris leistų automatizuoti dalį kibernetinio saugumo rizikos valdymo proceso.

Darbo uždaviniai:

1. Atskleisti dirbtinio intelekto sistemų sampratą ir taikymą finansų įstaigose.
2. Nustatyti reikalavimus rizikos scenarijų sudedamosioms dalims.
3. Sudaryti dirbtinio intelekto sistemų rizikos scenarijų generavimo metodologiją.
4. Sukurti dirbtinio intelekto sistemų rizikos scenarijų generatorių.
5. Atlikti eksperimentinį dirbtinio intelekto sistemos rizikos scenarijų generavimą.

Darbo struktūra

Pirmojoje darbo dalyje „Dirbtinio intelekto sistemų samprata ir taikymas finansų įstaigose“ nagrinėjamos dirbtinio intelekto sąvokos, elementai ir autonomiškumo lygiai. Analizuojamas dirbtinio intelekto gyvavimo ciklas, finansų rinkos struktūra bei dirbtinio intelekto taikymas finansų sektoriuje. Pateikiamos dirbtinio intelekto vystymosi tendencijos finansų srityje ir apžvelgiamos jo pritaikymo galimybės finansinių paslaugų saugumui užtikrinti.

Antrojoje dalyje „Reikalavimai rizikos scenarijų sudedamosioms dalims“ nagrinėjama dirbtinio intelekto reguliacinė aplinka bei rizikos valdymo standartai. Taip pat aptariami reikalavimai rizikos scenarijų formavimui.

Trečiojoje dalyje „Dirbtinio intelekto sistemų rizikos scenarijų generavimo metodologija“ aprašomos rizikos scenarijų sudedamosios dalys ir jų elementų taksonomijos. Atliekamas dirbtinio intelekto įrankių – ChatGPT-4, Google Gemini 1.5 Flash, IBM Watson AI, Microsoft Copilot – palyginimas ir įrankių analizė. Taip pat suformuojama užklausoje struktūra rizikos scenarijams generuoti.

Ketvirtojoje dalyje „DI sistemos rizikos scenarijų generatorius“ sukuriamas rizikos scenarijų generatorius, naudojant GPT-4 įrankį. Detaliai aprašomas generatoriaus kūrimo, apmokymo procesas, vartotojo sąsajos ypatumai bei užtikrinama generuojamų scenarijų saugumo ir privatumo apsauga.

Penktojoje dalyje „Eksperimentas su sukurtu generatoriumi“ atliekamas eksperimentas su hipotetinėmis finansų įmonėmis: kredito ir mokėjimų, draudimo, finansinių priemonių, kriptoturto rinkose. Taip pat atliekami eksperimentai bendradarbiaujant su kibernetinio saugumo ekspertais, siekiant įvertinti generuojamų rizikos scenarijų efektyvumą ir pritaikomumą praktikoje.

Darbe naudoti literatūros šaltiniai

Teorinėje darbo dalyje daugiausiai buvo naudotasi užsienio autorių moksliniais darbais, susijusiais su dirbtiniu intelektu ir kibernetiniu saugumu, bei įvairiomis kibernetinio saugumo publikacijomis, paskelbtomis tokių organizacijų kaip ENISA, Europos Parlamento ir Europos Sąjungos Tarybos, NIST, MITRE. Šaltiniuose buvo nagrinėjamos pagrindinės direktyvos ir aktai, susiję su dirbtinio intelekto rizikomis kibernetinio saugumo kontekste. Finansų temos kontekstui pagrįsti buvo pagrinde remtasi DORA reglamentu, OECD gairėmis bei Lietuvos banko paskelbtomis publikacijomis, kurie detaliai analizuoja finansų įstaigų kibernetinio saugumo reguliavimą. Praktiniams vertinimams ir pastebėjimams finansų įstaigų dirbtinio intelekto sistemų kibernetinio saugumo rizikų generatoriaus naudą bei problematiką pagrįsti naudotasi straipsniais ir kitais Lietuvos bei užsienio autorių šaltiniais internete.

Darbo ir tyrimo metodai

Analizuojant teorinę dalį – dirbtinio intelekto sistemų taikymą finansų įstaigose ir kibernetinio saugumo rizikos scenarijų generavimą, buvo naudojami šie tyrimo metodai: lyginamoji mokslinė literatūros analizė, sintezė, abstrahavimas ir sisteminimas. Šie metodai padėjo įvertinti esamas teorines nuostatas, identifikuoti dirbtinio intelekto taikymo tendencijas finansų sektoriuje ir apibrėžti pagrindinius kibernetinio saugumo rizikos elementus.

Metodologinėje dalyje buvo pasitelkti analizės, sintezės, sisteminimo ir klasifikavimo metodai. Šių metodų taikymas leido apibrėžti rizikos scenarijaus sudedamąsias dalis, atlikti rizikos elementų taksonomijos sudarymą bei užklausoje standarto kūrimą. Be to, buvo įvertinti skirtingi dirbtinio intelekto įrankiai, tokie kaip ChatGPT-4, Google Gemini 1.5 Flash, IBM Watson AI ir

Microsoft Copilot, siekiant nustatyti tinkamiausią įrankį finansų įstaigų rizikos scenarijų generavimo procesui.

Darbo teorinė reikšmė

- Išanalizuota dirbtinio intelekto samprata: apibrėžta dirbtinio intelekto sąvoka, išanalizuoti pagrindiniai DI elementai ir jų autonomiškumo lygiai ir identifikuoti dirbtinio intelekto gyvavimo ciklo etapai.
- Ištirta finansų įstaigų struktūra ir DI sistemų vystymosi tendencijos finansų sektoriuje, siekiant suprasti dirbtinio intelekto įtaką šio sektoriaus plėtrai ir saugumui.
- Įvertinti galimi dirbtinio intelekto taikymo privalumai ir iššūkiai finansinėse paslaugose, ypatingą dėmesį skiriant kibernetinio saugumo aspektams.
- Ištirti DI kibernetinės saugos rizikos valdymui taikomi rizikos valdymo standartai ir rekomendacijos, siekiant sudaryti prielaidas efektyviam rizikos valdymui.
- Nustatyti reikalavimai kibernetinės saugos rizikos scenarijams, siekiant užtikrinti, kad scenarijai tinkamai atspindėtų galimus kibernetinio saugumo incidentus.
- Atlikta esamų reikalavimų ir standartų analizė, susijusi su DI sistemų rizikos scenarijų kūrimu ir valdymu.
- Apsibrėžta rizikos scenarijų elementų taksonomija, kuri padės sistemingai klasifikuoti kibernetinio saugumo grėsmes, remiantis teoriniais tyrimais ir standartais.
- Atliktas dirbtinio intelekto įrankių palyginimas, įskaitant ChatGPT-4, Google Gemini 1.5 Flash, IBM Watson AI ir Microsoft Copilot, siekiant nustatyti efektyviausią priemonę kibernetinio saugumo rizikos scenarijų generavimui finansų sektoriuje.
- Sukurta užklauso formavimo metodika, padedanti automatizuotai generuoti kibernetinio saugumo rizikos scenarijus, pritaikytus finansų įstaigoms.

Darbo praktinė reikšmė

- Sukurtas generatorius pasinaudojant ChatGPT, kuris pritaikytas efektyviam kibernetinio saugumo scenarijų generavimui, orientuotam į finansų sektoriaus specifiką.
- Atliktas eksperimentas su sukurtu rizikos scenarijų generatoriumi, pritaikant jį hipotetinėms finansų įstaigoms (kredito, draudimo, kriptoturto rinkoms). Eksperimento metu buvo vertinamas generatoriaus efektyvumas, sugeneruotų scenarijų kokybė ir pritaikomumas finansų įstaigų rinkoms.

Darbo struktūra ir apimtis

Darbą sudaro įvadas, 3 dalys, išvados. Pagrindinė darbo medžiaga aprašyta 92 puslapių, įskaitant 5 lenteles, 24 paveikslus. Taip pat pateikiama 14 priedų. Panaudotos literatūros sąrašą sudaro 145 šaltiniai.

1. DIRBTINIO INTELEKTO SISTEMŲ SAMPRATA IR TAIKYMAS FINANSŲ ĮSTAIGOSE

Šioje darbo dalyje analizuojama dirbtinio intelekto samprata: sąvoka, elementai ir autonomiškumo lygiai. Aprašomas dirbtinio intelekto gyvavimo ciklas ir jo pagrindiniai etapai. Atliekama finansų rinkų struktūros analizė, aptariamas dirbtinio intelekto taikymas finansų įstaigose.

1.1. Dirbtinio intelekto sistemų samprata

Dirbtinio intelekto (DI) sistemų samprata yra plati ir įvairiapusė, atspindinti šios technologijos raidą nuo pirmųjų bandymų imituoti žmogaus intelektą iki pažangių autonominių sistemų, gebančių savarankiškai spręsti sudėtingas problemas. Šiame poskyryje toliau bus nagrinėjama, kaip įvairūs autoriai ir institucijos apibrėžia DI sąvoką, kokie yra DI pagrindiniai elementai ir kokią reikšmę turi autonomiškumo lygiai bei gyvavimo ciklas efektyviam šių sistemų pritaikymui.

„Dirbtinio intelekto“ sąvoka siejama su 1956 m., kai Dartmuto koledže Naujajame Hampšyre buvo surengtas Dartmuto vasaros dirbtinio intelekto tyrimų projektas (DSRPAI), kuriame pirmą kartą buvo pavartotas šis terminas. DSRPAI projekto tikslas buvo suvienyti įvairių sričių mokslininkus, kad būtų sukurta nauja mokslinių tyrimų sritis, skirta kurti mašinas, galinčias imituoti žmogaus intelektą (Haenlein ir Kaplan, 2019). Toliau bus palyginama dirbtinio intelekto sąvoka, aprašyta skirtingų mokslininkų.

Tiek mokslininkai, tiek praktikai neprieina vienos nuomonės dėl dirbtinio intelekto sąvokos reikšmės, todėl dirbtinis intelektas, kaip technologija turi keletą apibrėžimų, tokių kaip: skaičiavimo metodas, mašina, rinkinys, programinė įranga ar mašininė sistema (1 lentelė).

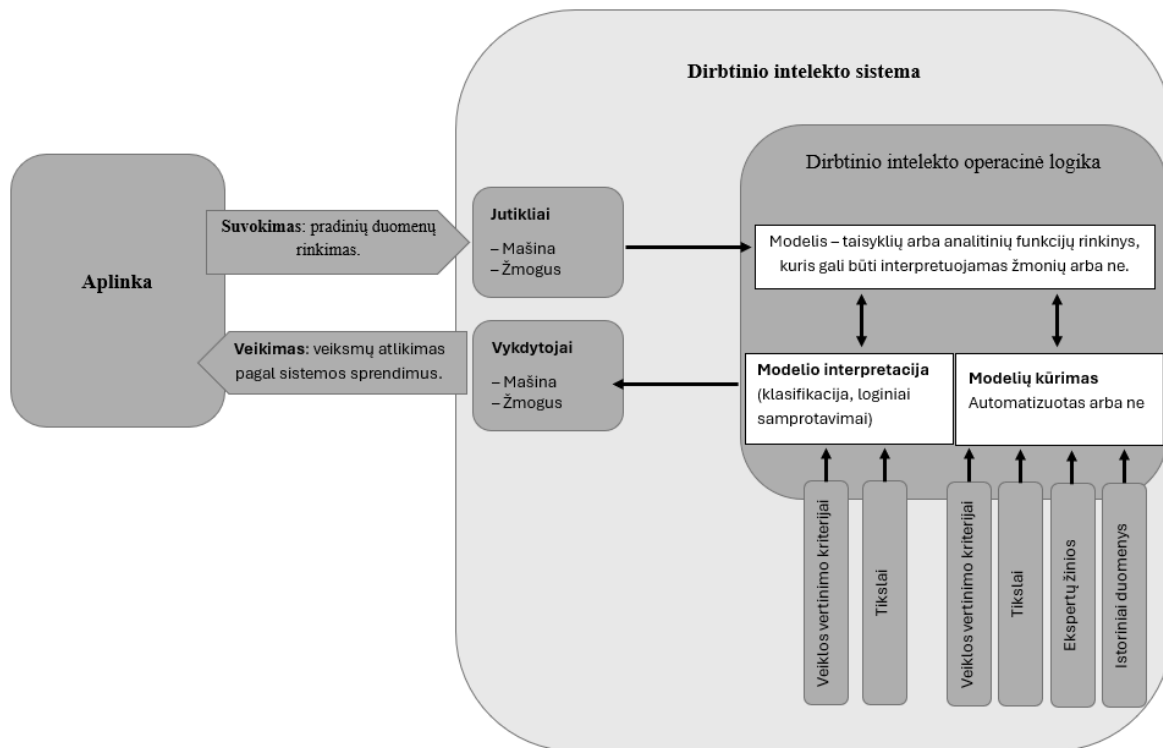
1 lentelė. Dirbtinio intelekto sąvokų analizė

Šaltinis	Apibrėžimas
Salehi ir Burgueño (2018)	Skaičiavimo metodas, kuriuo bandoma imituoti žmogaus pažinimo galimybes, manipuliuojant simboliais ir simboliškai struktūruotomis žinių bazėmis.
Rai et al. (2019)	Mašinos gebėjimas atlikti kognityvines funkcijas, kurias mes siejame su žmogaus protu.
Ai Hleg (2019)	Sistema, kuri demonstruoja protingą elgesį analizuodama savo aplinką ir imdamasi veiksmų – su tam tikru savarankiškumo laipsniu – konkreitiems tikslams pasiekti.
European Commission (2020)	Technologijų, jungiančių duomenis, algoritmus ir skaičiavimo galią, rinkinys.
Leitner-Hanetseder et al. (2021)	Dirbtinis intelektas yra savarankiškai besimokanti sistema.
OECD (2024)	Mašininė sistema, skirta iš gautos įvesties padaryti aiškių ar numanomų objektų išvadas.
Europos parlamentas ir Europos sąjungos taryba (2024)	Programinė įranga, kuri naudoja vieną ar daugiau metodų, galinčių generuoti turinį, prognozes, rekomendacijas ar sprendimus, turinčius įtakos fizinei ar virtualiai aplinkai.

Šaltinis: sudaryta autorės.

1 lentelėje matyti, kad Dirbtinio intelekto (DI) sąvokos varijuoja tarp skirtingų autorių, atspindėdamos įvairius požiūrius apie dirbtinio intelekto sampratą. Salehi ir Burgueño, (2018) išryškina DI technologiją, kaip skaičiavimo metodą, pabrėžiant žmogaus imitavimą, kuomet bandoma manipuluoti įvairiais simboliais ir žinių bazėmis. Kai, Rai ir kt. (2019) akcentuoja, kad DI, tai mašina, kuri siejama su žmogaus protu, galinti mokytis, samprotauti, spręsti problemas ir priimti sprendimus. O Ai Hleg, (2019) ir Leitner-Hanetseder ir kt. (2021), teigia, kad dirbtinis intelektas yra savarankiškai besimokanti sistema. Kita vertus, European Commission (2020) ir OECD (2024) DI apibūdina kaip technologijų, algoritmų ir skaičiavimo galios rinkinį ar mašininę sistemą, kuri geba generuoti išvadas iš gautos įvesties, akcentuodami DI technologinį pobūdį. Tuo tarpu, ES Dirbtinio intelekto akte yra įtvirtinta oficiali DI sistemos sąvoka, kur pagrindinė DI sistemų savybė yra jų gebėjimas daryti išvadas, t. y. procesas, kuriam vykstant gaunami išvediniai, pavyzdžiui, predikcijos, turinys, rekomendacijos ar sprendimai, kurie gali turėti įtakos fizinei ir virtualiai aplinkai (Europos parlamentas ir Europos sąjungos taryba, 2024).

Ši oficiali DI sistemos sąvoka, įteisinta ES Dirbtinio intelekto akte, ne tik apibūdina DI funkcionalumą, bet ir atskleidžia, jog DI sistemos yra sudėtingos struktūros, sudarytos iš elementų, kurie kartu sąveikauja ir vykdo užduotis. DI sistemos elementai apima tris pagrindines dalis: jutiklius, operacinę logiką ir vykdytojus (1 pav.).



Sudaryta autorės pagal: OECD, (2019). Scoping the OECD AI principles, p. 6.

1 pav. Dirbtinio intelekto sistema

Kaip matyti pagal 1 pav. dirbtinio intelekto sistemos elementas – jutikliai (angl. sensors) yra atsakingi už pradinio duomenų rinkimą iš aplinkos, kuri gali būti tiek fizinė, tiek virtuali. Pavyzdžiui, DI sistema gali naudoti kameras, mikrofonus ar kitus jutiklius, kad gautų informaciją iš aplinkos. Surinkti duomenys toliau yra apdorojami ir perduodami operacinei logikai. Operacinė logika (angl. operational logic) yra pagrindinis sistemos blokas, kuriame vyksta duomenų apdorojimas, modelių kūrimas ir interpretavimas. Modelio kūrime yra apdorojami duomenys rankiniu būdu arba naudojant automatizuotus įrankius, tokius kaip mašininio mokymosi algoritmai. Apdorojus duomenis modelio interpretacijos procese, sistema analizuoja sukurtus modelius ir generuoja sprendimus ar rekomendacijas, kurias vėliau įgyvendina vykdytojai. (OECD, 2019)

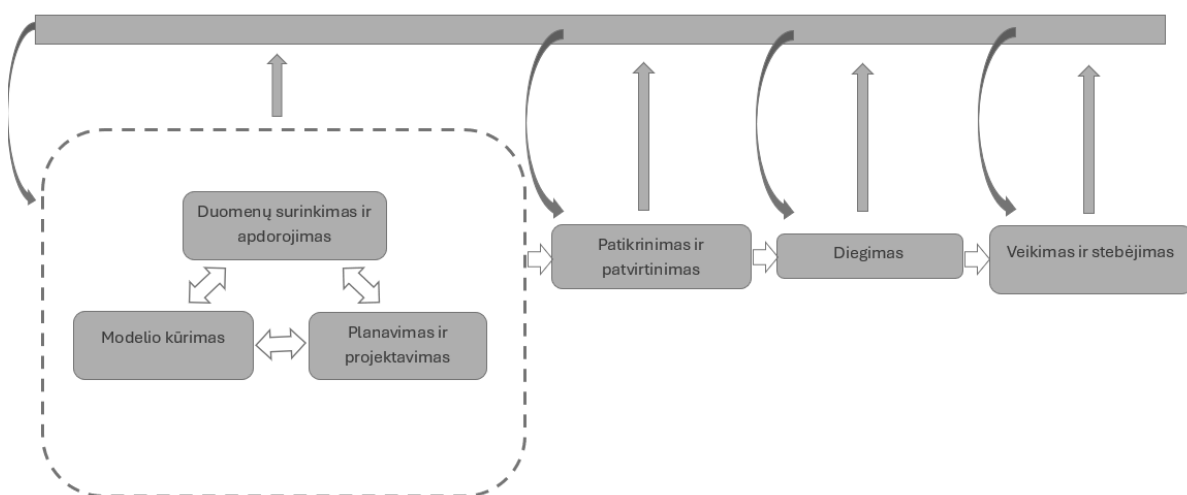
Vykdytojai (angl. actuators) yra trečiasis pagrindinis DI sistemos elementas, atsakingas už veiksmų atlikimą remiantis operacinės logikos pateiktais sprendimais. Vykdytojai gali būti tiek mechaniniai, tiek informaciniai, kurie naudoja gautas modelio interpretacijas, kad suformuluotų galutinius rezultatus. Pagal suformuluotus rezultatus yra atliekamas nustatytas veiksmas, kuris daro tam tikrą įtaką aplinkai, kurioje sistema veikia. Šio veiksmo atlikimas lemia tam tikrą dirbtinio intelekto sistemos autonomijos lygį, tai yra, koku mastu sistema gali veikti be žmogaus įsikišimo (OECD, 2022):

- **Nėra veiksmo autonomijos** (dar vadinama "žmogaus parama"). Kai sistema negali veikti pagal savo rekomendacijas ar išvestis. Žmogus savo nuožiūra naudoja arba ignoruoja dirbtinio intelekto sistemos rekomendacijas ar išvestis.

- **Maža veiksmo autonomija** (dar vadinama „žmogus cikle“). Sistema įvertina įvestis ir veikia pagal savo rekomendacijas ar išvestis, jei žmogus sutinka.
- **Vidutinė veiksmo autonomija** (dar vadinama „žmogus ant ciklo“). Sistema įvertina įvestis ir veikia pagal savo rekomendacijas ar išvestis, nebent žmogus vetuoja.
- **Didelė veiksmo autonomija** (dar vadinama „žmogus už ciklo ribų“). Sistema įvertina įvestis ir veikia pagal savo rekomendacijas ar išvestis be žmogaus įsikišimo.

Šie autonomijos lygiai yra svarbūs, nes jie lemia DI sistemos veiklos laipsnį ir atsakomybės pasiskirstymą tarp žmogaus ir dirbtinio intelekto sistemos. Kuo didesnis autonomijos lygis, tuo daugiau atsakomybės tenka sistemai, o tai kelia klausimų dėl skaidrumo ir atskaitomybės už jos veiksmus (OECD, 2022).

Be to, kaip papildomą struktūrą pagrindinėms dirbtinio intelekto sistemos techninėms charakteristikoms suprasti yra naudojamas dirbtinio intelekto gyvavimo ciklas, kuris yra svarbus siekiant užtikrinti DI sistemų efektyvumą, saugumą ir atitikimą etikos bei teisiniams reikalavimams. DI gyvavimo ciklą paprastai sudaro keturi pagrindiniai etapai: projektavimas, duomenų rinkimas ir modelio kūrimas; tikrinimas ir tvirtinimas; diegimas, bei veikimas ir stebėsena (De Silva & Alahakoon, 2022) (2 pav.).



Sudaryta autorės, pagal OECD, (2019). Scoping the OECD AI principles. ir Pupillo et al., (2021). Artificial Intelligence and Cybersecurity.

2 pav. Dirbtinio intelekto gyvavimo ciklas

Pagal 2 paveikslą matyti, kad projektavimo, duomenų rinkimo ir modeliavimo aspektai gali skirtis eiliškumo tvarka, priklausomai nuo kuriamos dirbtinio intelekto sistemos. Šiame gyvavimo ciklo etape planavimo ir projektavimo stadija yra fundamentinė, joje yra formuojami pagrindiniai sistemos tikslai ir reikalavimai, kurie lemia tolimesnių etapų eigą (OECD, 2019). Anot, ENISA, (2020), projektavimo ir duomenų surinkimo etapuose būtina atsižvelgti į dirbtinio intelekto kuriamos sistemos tikslus, naudojamų duomenų šaltinius, bei jų kokybę. Tai apima išsamią duomenų analizę, jų atrinkimą ir pirminį apdorojimą, siekiant užtikrinti, kad duomenys būtų tinkami sistemos mokymui.

Duomenų rinkimo strategijos kūrimas apima duomenų šaltinių identifikavimą ir jų integravimą į bendrą sistemą. Tuo tarpu, OECD, (2019) pabrėžia, kad duomenų privatumas ir saugumas turi būti, taip pat, aprašytas pirminiuose projektavimo etapuose, siekiant išvengti potencialių teisinių ir saugumo problemų ateityje. Tačiau, teigia ir, kad pernelyg didelis dėmesys duomenų kokybei projektavimo etape gali užgožti kitus svarbius aspektus, tokius kaip etiniai ir privatumo klausimai.

Modelio kūrimo etapas prasideda nuo surinktų duomenų analizės ir modelio algoritmų kūrimo. Šiame etape vykdomas įvairių modelių testavimas ir optimizavimas, siekiant rasti geriausią sprendimą kuriamai sistemai. Anot, ISO/IEC 23894 (2023), svarbu užtikrinti, kad modelio mokymas vyktų su įvairiais duomenų rinkiniais, kad būtų išvengta modelio šališkumo. Modelio interpretavimas, taip pat yra esminė dalis, nes reikia užtikrinti, kad modelio išvados būtų aiškios ir pagrįstos, o tai reikalauja nuodugnios modelio veikimo analizės ir rezultatų vizualizacijos. Pasak, ENISA (2020), taip pat, būtinas nuolatinis tikrinimas ir atnaujinimas modelio parametrų, kur būtų įtraukti įvairių sričių ekspertai, kurie užtikrintų modelio tinkamumą praktiniam naudojimui.

Tikrinimo ir tvirtinimo etapas yra esminis siekiant užtikrinti, kad sukurtas modelis atitiktų nustatytus reikalavimus ir būtų pasiruošęs diegimui. Šiame etape atliekami įvairūs testai, tokie kaip modelio tikslumo, stabilumo ir patikimumo vertinimas, naudojant realius duomenis (Shahriar et al., 2023). Anot, ENISA (2020), tikrinimo procesas apima modelio veikimo testavimą įvairiose sąlygose, siekiant nustatyti galimus trūkumus ir užtikrinti modelio patikimumą. ISO/IEC 23894 (2023) akcentuoja, kad netinkamai praeitas tikrinimo ir tvirtinimo etapas gali sukelti netikėtų sistemos veikimo sutrikimų arba kokybės pablogėjimą, kas pabrėžia būtinybę skirti pakankamai dėmesio ir išteklių šiam etapui, siekiant užtikrinti, kad modelis būtų ne tik techniškai teisingas, bet ir saugus, bei patikimas.

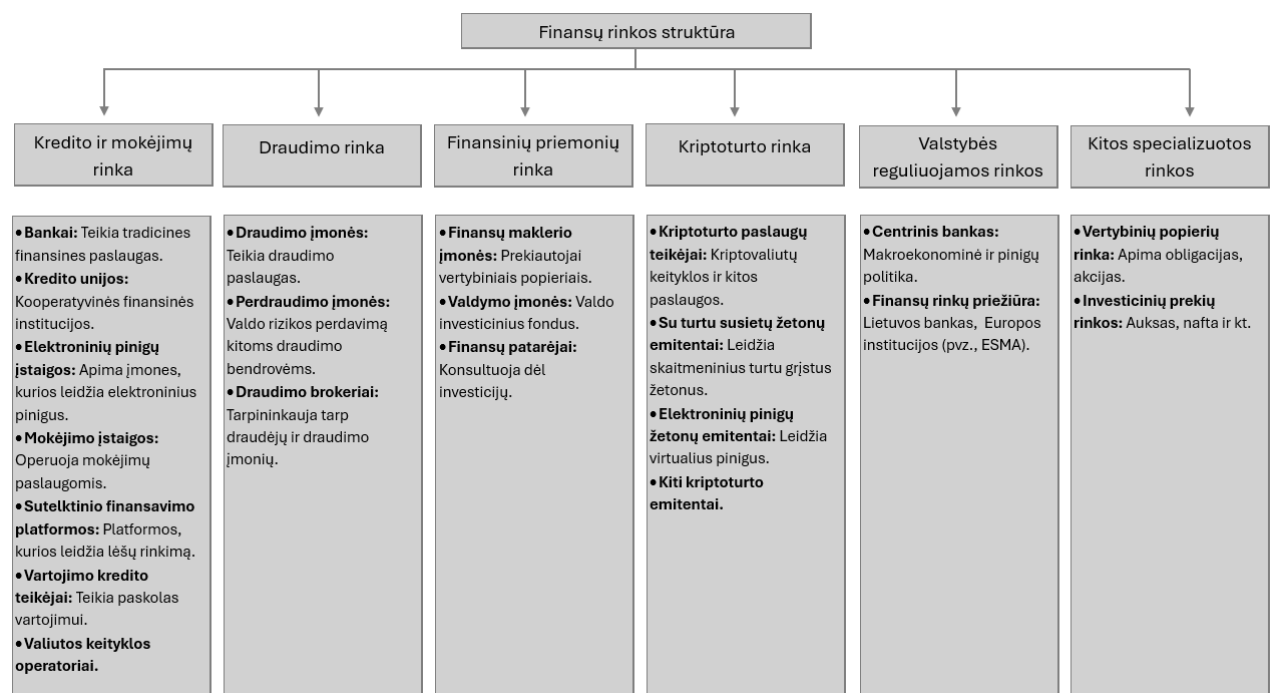
Gyvavimo ciklo diegimo etape yra paruošiamas modelis integravimui į produkcinę aplinką. Tai gali apimti modelio perkėlimą į debesų kompiuterijos infrastruktūrą arba įdiegti į konkrečius įrenginius. Diegimo metu būtina užtikrinti, kad modelis turėtų reikalingus kompiuterinius ir atminties resursus, kad galėtų tinkamai veikti. Po diegimo svarbus nuolatinis stebėjimas modelio veikimo ir atlikimas reguliarių atnaujinimų, bei optimizacijų, siekiant užtikrinti sistemos efektyvumą ir patikimumą (Shahriar et al., 2023). ISO/IEC 23894 (2023) nustato, kad diegimo etape svarbu įtraukti nuolatinis atnaujinimus ir modelio priežiūrą, siekiant palaikyti modelio efektyvumą ir atitikti kintančius organizacijų reikalavimus. Pagal, NIST (2023) diegimas tiesioginėje gamyboje apima bandomąjį bandymą, suderinamumo su senosiomis sistemomis tikrinimą, reguliavimo atitikties užtikrinimą, organizacinių pokyčių valdymą ir vartotojų patirties vertinimą.

Veikimo ir stebėsenos etapas apima nuolatinį modelio veikimo stebėjimą ir priežiūrą. Tai apima duomenų įvesties ir modelio rezultatų stebėjimą, siekiant nustatyti galimus pokyčius ar nukrypimus nuo normos, kurie gali rodyti, kad reikia atlikti modelio atnaujinimus arba persitreniravimą. Veikimo stebėseną, taip pat apima modelio veikimo vertinimą pagal išsikeltus tikslus ir rezultatų analizę (De Silva & Alahakoon, 2022). Pagal, ENISA, (2020) stebėseną, taip pat, yra nuolatinis procesas, kuris turi būti integruotas į kasdienės organizacijos operacijas, siekiant užtikrinti, kad modelis veiktų optimaliai ir atitiktų nustatytus tikslus.

Taigi, apibendrinant, dirbtinio intelekto sampratą apima DI sąvoka, elementai, autonomiškumo lygiai ir gyvavimo ciklas. Kur DI sąvokos apima įvairius aspektus, kurios atspindi technologijos plėtimąsi - nuo skaičiavimo metodų ir matematinių apibrėžčių iki savarankiškai besimokančių sistemų. Tačiau Europos parlamento ir tarybos išleistas reglamentas pateikia oficialią dirbtinio intelekto sąvoką įteisintą teisės aktais, kur DI tai, programinė įranga, kuri naudoja vieną ar daugiau metodų, galinčių generuoti turinį, prognozes, rekomendacijas ar sprendimus, turinčius įtakos fizinei ar virtualiai aplinkai. DI sistemų veikimas yra grindžiamas trimis pagrindiniais elementais – jutikliais, operacine logika ir vykdytojais, kurių tarpusavio sąveika užtikrina efektyvų duomenų apdorojimą ir rezultatų generavimą. Autonomiškumo lygiai rodo sistemų gebėjimą veikti nuo žmogaus patvirtinimo iki visiškai savarankiškų sprendimų priėmimo, o gyvavimo ciklas, susidedantis iš projektavimo, modelio kūrimo, tikrinimo, diegimo ir stebėsenos etapų, garantuoja sistemų patikimumą ir pritaikymą praktikoje. Taigi, DI sistemų samprata yra kompleksiška, apimanti tiek teorinius pagrindus, tiek praktinio pritaikymo mechanizmus, kurie sudaro pagrindą technologijų plėtrai ir jų naudai įvairiose srityse. Sekančiame poskyryje bus atliekama finansų rinkos struktūros analizė.

1.2. Finansų rinkos struktūra

Finansų įstaigos yra vienas iš pagrindinių ekonomikos sistemos elementų, užtikrinantis kapitalo valdymą, investavimą, skolinimą ir rizikos valdymą. Finansų įstaigos ne tik leidžia valdyti kasdienes finansines operacijas, bet ir padeda formuoti ilgalaikius ekonominius sprendimus. Įvairūs šių įstaigų veiklos modeliai ir reguliavimo skirtumai sukuria sudėtingą sistemą, kurioje kiekviena rūšis atlieka savo specifinį vaidmenį. Nors tradiciniai bankai ir draudimo įmonės daugiausia užtikrina stabilumą, bet tokios įstaigos kaip: mokėjimo įstaigos ir investicijų valdymo organizacijos skatina inovacijas ir efektyvumą. Tinkamas jų klasifikavimas pagal funkcijas, reguliavimo reikalavimus ir technologinį pagrindą leidžia geriau suprasti jų poveikį ekonomikai bei tarpusavio sąryšį (Loesch, 2018a). Toliau šiame poskyryje bus atliekama pagrindinių finansinių įstaigų rūšių analizė (3 pav.).



Šaltinis: sudaryta autorės

3 pav. Finansų rinkos struktūra

Kredito įstaigos, kaip bankai ir kredito unijos, vykdo finansinio tarpininkavimo funkciją, kur daugiausia dėmesio skiria kapitalo paskirstymui ir kreditavimui. Bankai yra viena iš pagrindinių finansų sistemos dalyvių, kuri teikia plačiausią paslaugų spektrą – nuo paskolų iki investicinių sprendimų. Universalūs bankai siūlo įvairias paslaugas, įskaitant taupymo sprendimus, mokėjimus, paskolas ir investicines paslaugas, o specializuoti bankai, tokie kaip hipotekos bankai, orientuojasi į konkrečias nišas. Universalūs bankai taip pat pasižymi stabilumu ir kapitalo diversifikacija, tačiau dėl sudėtingos struktūros jie dažnai yra lėtesni diegiant technologines inovacijas. Specializuoti bankai yra labiau efektyvesni savo srityse, tačiau dėl siauresnio veiklos spektro jie jautresni rinkos svyravimams. Toks bankų pasiskirstymas atskleidžia funkcinio klasifikavimo aspektą, kur universalūs bankai labiau orientuojasi į plataus spektro stabilumą, o specializuoti į efektyvumą specifinėse srityse (Europos Parlamentas, 2023; Loesch, 2018b). Tuo tarpu, kredito unijos veikia pagal kooperatinį modelį, orientuotą į specifines klientų grupes, tačiau jų mažesnė kapitalo bazė riboja galimybes diversifikuoti riziką. Skirtingai nei bankai, kredito unijos savo kapitalą kaupia per narių įnašus, o jų veikla yra labiau socialiai orientuota. Tai leidžia kredito unijoms geriau prisitaikyti prie vietos rinkos poreikių ir užtikrinti lanksčias sąlygas nariams, pavyzdžiui, teikiant mažesnes paskolas arba siūlant specializuotas taupymo programas (Lietuvos bankas, 2021).

Mokėjimo paslaugų teikėjai, koncentruojasi į skaitmeninių technologijų taikymą, siekdamas pagerinti operacijų efektyvumą ir sumažinti transakcijų kaštus. Elektroninių pinigų įstaigos (EPI) ir mokėjimo įstaigos (MI) užtikrina greitus, skaitmeninius mokėjimus. Kur jų veikla labiausiai orientuojasi į elektroninius mokėjimus, pinigų pervedimus ir bekontakčius atsiskaitymus, kurie tampa vis populiariesni skaitmeninės ekonomikos plėtros kontekste. Šių įstaigų pagrindinis privalumas yra

jų lankstumas ir gebėjimas greitai prisitaikyti prie technologinių naujovių, tokių kaip mobilieji mokėjimai, skaitmeninės valiutos ar „blockchain“ technologijos. Tačiau EPĮ ir MI veikla yra ribota lyginant su bankais – jos negali teikti paskolų ar indėlių paslaugų, kas riboja jų verslo modelio tvarumą ir mastą. Šis apribojimas išryškina technologinio klasifikavimo svarbą, kur technologijos yra jų pagrindas, tačiau tai kartu sukuria pažeidžiamumą kibernetiniams pavojams, ypač jei reguliavimo priemonės nėra pakankamos (Europos Komisija, 2020; Lietuvos respublikos finansų ministerija, 2023). Kadangi EPĮ ir MI veikla remiasi technologinėmis platformomis, kibernetinės atakos ar sistemos sutrikimai gali sukelti didelius nuostolius tiek pačioms įstaigoms, tiek jų klientams. Šis aspektas akcentuoja griežtos reguliacinės aplinkos svarbą. Europos Sąjungos direktyvos ir DORA reglamentas numato aukštus reikalavimus kibernetinio atsparumo ir operacinės rizikos valdymo srityse, siekiant užtikrinti šių įstaigų veiklos stabilumą ir klientų apsaugą (The European Parliament and the councils of the European union, 2022). EPĮ ir MI taip pat išsiskiria savo indėliu į finansinę įtrauktį. Šios įstaigos suteikia galimybę naudotis finansinėmis paslaugomis klientams, kurie neturi tradicinių banko sąskaitų, ypač besivystančiose rinkose. Tačiau jų konkurencinė padėtis dažnai ribojama bankų dominavimo ir sudėtingų reguliacinių reikalavimų, kurie gali lėtinti plėtrą ir inovacijų diegimą. Nepaisant to, EPĮ ir MI nuolat stengiasi didinti paslaugų spektrą, bendradarbiaudamos su kitomis finansų sektoriaus įmonėmis, taip stiprindamos savo poziciją rinkoje ir prisidėdamos prie skaitmeninės ekonomikos augimo (Europos Komisija, 2020; Lietuvos respublikos finansų ministerija, 2023).

Draudimo įmonės ir kredito įstaigos atlieka skirtingas, tačiau viena kitą papildančias funkcijas. Draudimo įmonės, orientuotos į rizikos valdymą, užtikrina finansinį stabilumą asmenims ir verslams, o perdraudimo bendrovės mažina sisteminės rizikos (pvz., stichinės nelaimės) poveikį visai rinkai. Draudimo įmonės taip pat teikia paslaugas, apsaugančias nuo finansinių nuostolių, susijusių su turto, sveikatos ar gyvybės rizika. Kur šių įmonių veiklos pagrindas yra rizikos vertinimas ir valdymas, siekiant užtikrinti, kad surinktos įmokos būtų pakankamos padengti draudimo išmokas bei veiklos sąnaudas. Funkcinio klasifikavimo požiūriu, draudimo įmonės papildo bankus ir investicijų valdymo įmones, suteikdamos ilgalaikius sprendimus, bet gali būti mažiau lanksčios, reaguojant į greitus rinkos pokyčius. Kredito įstaigos, priešingai, teikia tiesioginį kapitalą investicijoms ir vartojimui, skatindamos ekonominį augimą. Vis dėlto šių dviejų sektorių veikla turi ir trūkumų: draudimo įmonėms trūksta lankstumo likvidumo poreikių tenkinimui, o kredito įstaigos yra labiau pažeidžiamos ekonominių krizių metu dėl negražinamų paskolų rizikos (Lietuvos bankas, 2021; Loesch, 2018b).

Finansinių priemonių rinkos dalyviai, tokie kaip makleriai ir turto valdymo įmonės, suteikia tradicines investavimo galimybes ir užtikrina rinkos likvidumą. Šie dalyviai paprastai veikia reguliuojamoje aplinkoje, kuri užtikrina rinkos skaidrumą ir investuotojų apsaugą, o direktyva kaip MiFID 2 prisideda prie sąžiningos konkurencijos. Priešingai, kriptoturto paslaugų teikėjai, veikiantys decentralizuotose platformose, siūlo inovatyvius sprendimus, kurie sumažina tarpininkų vaidmenį ir operacijų kaštus. Jie apima tokias platformas kaip kriptovaliutų biržos, tokenų (žetonų) emitentai ir decentralizuotos finansų (DeFi) sistemos, kurios suteikia vartotojams prieigą prie inovatyvių finansinių sprendimų. Tačiau jų veikla susiduria su esminiais iššūkiais, susijusiais su reguliavimo neapibrėžtumu ir didelėmis kibernetinio saugumo rizikomis. Todėl ES reglamentas „Markets in Crypto-Assets“ (MiCA) suteikia aiškų reguliavimo pagrindą šiai sparčiai augančiai rinkai, apsaugoti

vartotojus nuo nepastovumo ir užtikrinti rinkos stabilumą. Kriptoturto paslaugų tiekėjų veikla taip pat išsiskiria technologijų integracija – blockchain technologija leidžia užtikrinti sandorių skaidrumą, decentralizuotumą ir patikimumą. Tačiau šios technologijos yra jautrios specifiniams rizikos veiksniams, tokiems kaip kibernetinės atakos ar neteisėtas kriptoturto naudojimas, įskaitant pinigų plovimą ir terorizmo finansavimą. Be to, nepastovumas kriptovaliutų rinkoje sukuria neprognozuojamą aplinką tiek investuotojams, tiek paslaugų teikėjams (Europos Komisija, 2020; Lietuvos respublikos finansų ministerija, 2023).

Centriniai bankai ir finansų priežiūros įstaigos, yra atsakingos už sisteminį finansų rinkų stabilumą, makroekonominį balanso palaikymą ir sisteminės rizikos valdymą. Pavyzdžiui, Lietuvos bankas, reguliuojantis atsakingo skolinimo politiką, gali užkirsti kelią kredito burbulų susidarymui. Tuo tarpu privatūs rinkos dalyviai, įskaitant komercinius bankus, draudimo bendroves, mokėjimo paslaugų teikėjus, ar kriptovaliutų platformas, veikia daug dinamiškiau, orientuodamiesi į inovacijų diegimą ir greitą prisitaikymą prie rinkos poreikių. Komerciniai bankai veikia kaip kapitalo paskirstytojai, siūlydami kreditus ir investicinius produktus, o EPI / MĮ, naudodamos skaitmenines technologijas, skatina mokėjimo ir skolinimo paslaugų modernizavimą. Vis dėlto, valstybės įstaigų lėtumas reaguojant į rinkos pokyčius dažnai riboja inovacijų tempą, o privatūs dalyviai, siekdami pelno, kartais imasi pernelyg rizikingų veiksmų, kas kelia grėsmę finansų rinkos stabilumui (Loesch, 2018b, 2018a).

Apibendrinant, galima teigti, kad finansų įstaigos atlieka kompleksines funkcijas, kurios apima kapitalo perskirstymą, rizikos valdymą ir rinkos likvidumo užtikrinimą. Tarpusavio sąveika tarp finansų įstaigų yra būtina siekiant užtikrinti sisteminį rinkos stabilumą. Tradicinės kredito ir draudimo įmonės išlieka svarbiausi finansinės infrastruktūros elementais, tačiau inovatyvios FinTech ir kriptoturto platformos keičia nusistovėjusias rinkos taisykles. Kur jų tinkamas klasifikavimas, veiksmingas reguliavimas ir subalansuota rinkos priežiūra leidžia efektyviau įgyvendinti reguliacines nuostatas bei prisitaikyti prie sparčių technologinių pokyčių, kurios yra būtinos sąlygos ilgalaikės finansų sistemos tvarumui užtikrinti. Sekančiame poskyryje bus aptariamos dirbtinio intelekto vystymosi tendencijos finansų srityje.

1.3. Dirbtinio intelekto sistemų vystymosi tendencijos finansų srityje

Dirbtinio intelekto technologijų plėtra yra viena iš reikšmingiausių šiuolaikinės technologinės revoliucijos krypčių, kurios poveikis jaučiamas įvairiose ekonomikos srityse, įskaitant ir finansų sektorių (Vinothkumar & Karunamurthy, 2023). Dirbtinio intelekto technologijos sudaro galimybes efektyvinti veiklą, sumažinti sąnaudas ir pasiūlyti inovatyvias paslaugas vartotojams. Šių technologijų naudojimas leidžia gerinti klientų aptarnavimą, valdyti riziką, aptikti sukčiavimo atvejus ir automatizuoti rutininę veiklą. Dėl DI pritaikymo galimybių finansų sektorius tampa dinamiškesnis ir lankstesnis reaguojant į rinkos pokyčius bei didėjančią reguliavimo sudėtingumą. Todėl, dirbtinio intelekto sistemos tampa svarbia ir neatsiejama šiuolaikinės finansų pramonės dalimi (OECD, 2021a).

Generatyvinio dirbtinio intelekto technologijos yra viena iš populiarių naujovių finansų sektoriuje, pasižyminčiu dideliu potencialu procesų automatizavimo srityje, kaip būdu užtikrinti efektyvumą ir pagerinti produktyvumą finansinių paslaugų operacijose ir rizikos valdyme (OECD, 2023). Pasak McKinsey & Company, (2023a), generatyvinis DI gali pagerinti procesų efektyvumą ir prisidėti prie naujų produktų, bei paslaugų kūrimo finansų srityje, leidžiant analizuoti didelius duomenų kiekius ir generuoti naujus sprendimus. Taip pat, nurodo, kad vienas pagrindinių generatyvinio DI privalumų būtų gebėjimas automatizuoti laiko reikalaujančias užduotis, tokias kaip finansinių ataskaitų rengimas ar kredito vertinimas (McKinsey & Company, 2023b). Anot, BCG (2023), generatyvinis dirbtinis intelektas leistų kurti sudėtingas simuliacijas ir scenarijus, kurios padėtų finansų specialistams modeliuoti ir numatyti galimus rinkos pokyčius.

Bankai yra viena iš pagrindinių įstaigų, plačiai taikančių DI sprendimus. Pagrindinės funkcijos apima kreditų rizikos vertinimą, sukčiavimo prevenciją ir klientų aptarnavimo automatizavimą (OECD, 2024c). Pavyzdžiui, mašininio mokymosi algoritmai, tokie kaip „XGBoost“ ar „Random Forest“, leidžia analizuoti didelius klientų duomenų rinkinius, siekiant tiksliai įvertinti jų mokumo galimybes. Tai gali būti ypač naudinga mažinant kredito nuostolius ir užtikrinant tikslingą paskolų suteikimą (Rachel Meltzer, 2023; Varun Raskar, 2024). Sukčiavimo aptikimas taip pat yra kritiškai svarbi funkcija, kurią užtikrina realaus laiko duomenų analizės sprendimai, tokie kaip „Visa Advanced Authorization“. Tokios technologijos ne tik padeda apsaugoti finansinius sandorius, bet ir stiprina vartotojų pasitikėjimą banko teikiamomis paslaugomis (Visa, 2024).

Draudimo sektoriuje DI naudojamas rizikos vertinimui bei žalų administravimui. Draudimo kompanijos, remdamosi didelių duomenų analize, gali kurti individualizuotus draudimo pasiūlymus, atitinkančius konkrečius klientų poreikius (Cao, 2020). Pavyzdžiui, „DataRobot“ platforma gali automatizuoti rizikos vertinimo procesą, kur būtų galimybė analizuoti didelius duomenų kiekius, leidžiant draudikams pateikti personalizuotus pasiūlymus ir taip pat efektyviau prognozuoti galimas rizikas, sumažinant apdorojimo laiką ir sąnaudas (DataRobot, 2024). Žalų administravimo procesų automatizavimas leidžia spartinti žalos vertinimą ir atsiskaitymus, kas didina klientų pasitenkinimą ir mažina administracines išlaidas (ENISA, 2015). Toks DI įrankis, kaip „Shift Technology“ galėtų padėti sukčiavimo atvejams aptikti ir žalų administravimui optimizuoti (SHIFT, 2024).

Investicijų valdymo srityje dirbtinis intelektas gali būti naudojamas portfelio optimizavimui, rinkos analizei ir algoritminės prekybos srityse. Algoritminė prekyba, paremta DI, leidžia investicinėms įmonėms greitai reaguoti į rinkos pokyčius, taip sumažinant riziką ir pagerinant investicijų grąžą, o portfelio optimizavimo įrankiai, tokie kaip „BlackRock Aladdin“, gali padėti investuotojams priimti pagrįstus sprendimus, mažinant riziką ir didinant investicijų grąžą (BlackRock, 2024; Cao, 2020). „Kensho“ platforma, gali būti taip pat naudinga rinkos duomenų analizei ir investicinių galimybių identifikavimui. Šis DI sprendimas leistų analizuoti makroekonominis duomenis ir sudaryti tikslines prognozes, kurios yra vertingos tiek investuotojams, tiek portfelio valdytojams (Kensho, 2024). Be to, „Two Sigma“ dirbtinio intelekto įrankis naudingas gali būti prekybos strategijoms, remiantis dideliais kiekiais istorinių ir realaus laiko duomenų taškų (Two Sigma, 2024). Taigi, šios tendencijos rodo, kad investicinės įmonės DI sprendimus vis dažniau naudoja siekdamos konkurencinio pranašumo.

Kriptoturto rinkoje DI technologijos tampa itin svarbios dėl didelio sandorių skaičiaus ir decentralizuotų sistemų, kur DI taip pat gali būti taikomas rinkos prognozavimui, naudojant algoritmus, analizuojančius kriptovaliutų kainų svyravimus (Cao, 2020). „Chainalysis“ dirbtinio intelekto pagrįstas sprendimas gali padėti kriptoturto biržoms stebėti ir analizuoti sandorius, siekiant užkirsti kelią pinigų plovimui ir kitiems finansiniams nusikaltimams (Chainalysis, 2024). Be to, „Elliptic“ sistema gali suteikti galimybę biržoms identifikuoti rizikingus adresus ir užtikrinti atitiktį reguliavimo reikalavimams (Elliptic, 2024). Kriptoturto rinkos plėtros tendencijos rodo, kad saugumo užtikrinimas ir sandorių greitis yra prioritetinės sritys.

Mokėjimų platformos yra kritiškai svarbios finansų sektoriuje, nes jos tvarko didelius kiekius transakcijų kasdien ir turi užtikrinti ne tik saugumą, bet ir efektyvumą. Dirbtinio intelekto technologijos leidžia realiu laiku analizuoti sandorius, aptikti anomalijas ir užkirsti kelią sukčiavimo atvejams (ENISA, 2014; OECD, 2024c). Tokios platformos kaip „Visa Advanced Authorization“ gali būti naudojamos realaus laiko transakcijų analizei ir padėti atpažinti anomalijas sandorių duomenyse, užkertant kelią galimiems sukčiavimo atvejams. Tai ne tik apsaugo finansines operacijas, bet ir padeda įstaigoms laikytis griežtų reguliavimo reikalavimų (Visa, 2024). „Mastercard Decision Intelligence“ platforma, kuri naudodama dirbtinį intelektą, taip pat gali analizuoti klientų operacijų modelius ir padėti nustatyti anomalijas, apsaugodama vartotojus nuo galimų grėsmių (Mastercard, 2024). O „Forter“ dirbtinio intelekto sprendimas gali padėti e. prekybos mokėjimams analizuoti ir apsaugoti nuo sukčiavimo. „Forter“ sistema naudoja mašininio mokymosi algoritmus, kad analizuotų vartotojų elgseną, transakcijų duomenis ir sukurtų sprendimus, kurie realiu laiku užtikrina tiek sandorių saugumą, tiek klientų patirtį (Forter, 2024). Tendencijos rodo, kad DI sprendimai mokėjimų sektoriuje vis labiau orientuojasi į klientų patirties gerinimą, mažinant neteisingus operacijų atmetimus, taip pat sukčiavimo prevencijos optimizavimą naudojant pažangius mašininio mokymosi algoritmus.

Reguliavimo institucijos taip pat aktyviai diegia DI sistemas, siekdamos stebėti finansų rinkų skaidrumą ir laikytis sudėtingų reguliavimo reikalavimų. Tai leidžia efektyviau prižiūrėti finansų rinkas, užtikrinti skaidrumą ir greičiau reaguoti į galimus piktnaudžiavimo atvejus (OECD, 2021a). Pavyzdžiui, „ComplyAdvantage“ padeda identifikuoti neteisėtą veiklą finansų sektoriuje ir užtikrina atitiktį vietiniams bei tarptautiniams standartams. Tai tampa ypač aktualu tarptautinėse rinkose, kur reguliavimo reikalavimai nuolat griežtėja (ComplyAdvantage, 2024). Taip pat „Ayasdi“ gali būti naudingas duomenų klasterizacijai ir anomalijų aptikimui. „Ayasdi“ naudoja topologinę duomenų analizę (TDA), kad identifikuotų sudėtingus ryšius tarp sandorių ir nustatytų įtartinus veiksmus, kurie galėtų rodyti neteisėtą veiklą (Ayasdi, 2021). Taigi, reguliavimo institucijose DI įrankiai gali padėti greitai ir efektyviai apdoroti didžiulius duomenų kiekius, kurie būtų neįveikiami tradiciniais metodais. DI sprendimai taip pat prisideda prie prevencijos, užkertant kelią neteisėtoms finansinėms veikloms ankstyvose jų stadijose.

2 lentelė. Finansų įstaigos ir jų pagrindinės funkcijos, kurioms naudojamas DI

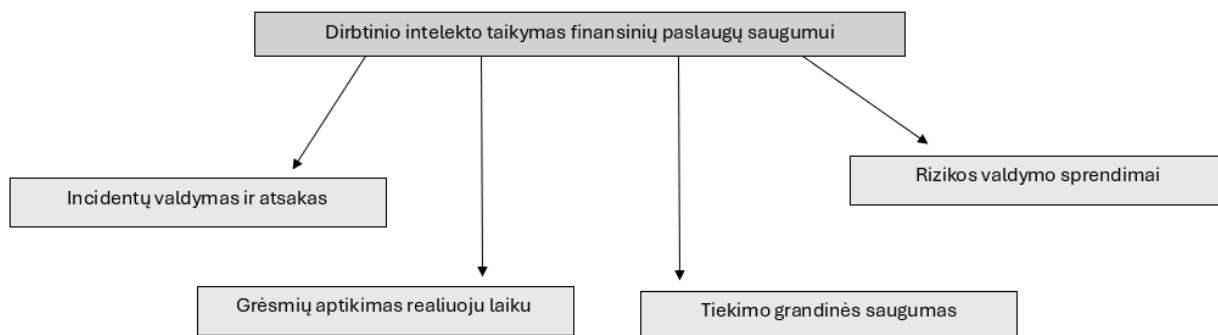
Finansų įstaigos rūšis	Pagrindinės funkcijos	DI įrankiai/pavyzdžiai
Bankai	Kreditų rizikos valdymas, sukčiavimo aptikimas, klientų aptarnavimo automatizavimas	„XGBoost“, „Random Forest“
Investicinės įmonės	Portfelio optimizavimas, algoritminė prekyba, rinkos analizavimas	„BlackRock Aladdin“, „Kensho“, „Two Sigma“
Draudimo kompanijos	Rizikos vertinimas, žalų administravimas	„Shift Technology“, „DataRobot“
Mokėjimų platformos	Transakcijų monitoringas, sukčiavimo aptikimas	„Visa Advanced Authorization“, „Mastercard Decision Intelligence“, „Forter“
Kriptoturto rinka	Sandorių saugumas, pinigų plovimo prevencija, rinkos prognozavimas	„Chainalysis“, „Elliptic“
Reguliavimo institucijos	Atitikties užtikrinimas, anomalijų aptikimas	„ComplyAdvantage“, „Ayasdi“

Šaltinis: sudaryta autorės

Taigi, dirbtinio intelekto sistemų vystymosi tendencijos finansų srityje atspindi spartų technologinį progresą ir augantį šių sprendimų pritaikymą finansų sektoriuje. Kur skirtingos finansų įstaigos orientuojasi į specifines funkcijas (2 lentelė): bankai daugiausia dėmesio skiria sukčiavimo prevencijai ir rizikos valdymui, investicinės įmonės optimizuoja portfelius ir prekybos strategijas, draudimo kompanijos naudoja DI žalų administravimo automatizavimui, o mokėjimų platformos pabrėžia saugumo užtikrinimą. Reguliavimo institucijos, taikydamos DI, padeda užtikrinti rinkos skaidrumą ir laikytis tarptautinių reikalavimų, o kriptoturto rinkoje pagrindinis dėmesys skiriamas rinkos prognozavimams ir saugumui vykdant transakcijas. *Apibendrinant, galima teigti, kad dirbtinio intelekto sistemų naudojimas finansų sektoriuje transformuoja ne tik esamas veiklos sritis, bet ir skatina naujų sprendimų kūrimą. Bankai, investicinės įmonės, draudimo kompanijos ir kitos finansų įstaigos taiko DI įvairioms funkcijoms - nuo rizikos vertinimo iki klientų aptarnavimo. DI technologijos padeda įstaigoms geriau suprasti klientų poreikius, efektyviau valdyti riziką ir užtikrinti didesnę veiklos skaidrumą. Šios tendencijos finansų įstaigose rodo, kad dirbtinis intelektas ir toliau plės savo pritaikymo sritis finansų sektoriuje, didindamas tiek efektyvumą, tiek paslaugų kokybę. Sekančiame poskyryje bus apžvelgiamas dirbtinio intelekto taikymas finansinių paslaugų saugume.*

1.4. Dirbtinio intelekto taikymo finansinių paslaugų saugumui apžvalga

Finansinių paslaugų sektorius yra viena iš labiausiai reguliuojamų ir jautriausių sričių, kurioje kyla daugybė kibernetinių rizikų dėl didelių finansinių srautų ir klientų duomenų apdorojimo. Dirbtinis intelektas šioje srityje įgauna vis didesnę reikšmę, nes jis gali padėti identifikuoti ir valdyti kibernetines grėsmes (Europos komisija, 2020a). Europos Sąjungos NIS2 direktyva (Network and Information Systems) ir DORA (Digital Operational Resilience Act) reglamentas, kartu su kibernetinio saugumo strategijomis, pabrėžia poreikį stiprinti finansinių paslaugų sektoriaus kibernetinį atsparumą naudojant pažangias technologijas (Europos Komisija, 2020b; Europos Parlamentas ir Europos Sąjungos Taryba, 2022; The European Parliament and The Council of the European Union, 2022). Toliau, šiame poskyryje bus aptariama, kaip DI gali būti taikomas finansinių paslaugų saugumui, analizuojant pagrindinius kibernetinių rizikų valdymo aspektus bei prevencijos priemonės (4 pav.).



Sudaryta autorės, pagal: Europos komisija, Europos Sąjungos skaitmeninio dešimtmečio kibernetinio saugumo strategija (2020b), ES skaitmeninių finansų strategijos (2020a); Europos parlamentas ir Europos sąjungos taryba, Europos parlamento ir tarybos direktyva (ES) 2022/2555 (2022)

4 pav. Dirbtinio intelekto taikymas finansinių paslaugų saugumui.

Dirbtinio intelekto pagrįstos sistemos padeda bankams ir kitoms finansų įstaigoms greitai aptikti kibernetines grėsmes, įskaitant duomenų vagystes, kenkėjiškas programas ir vidines grėsmes, kurios gali būti mažiau pastebimos. Tokios sistemos automatizuoja daugelį procesų ir naudoja didelius duomenų kiekius, kad mokytųsi iš ankstesnių grėsmių ir atitinkamai pritaikytų apsaugos priemones. NIS2 direktyva reikalauja, kad finansų įstaigos aktyviai diegtų rizikos valdymo priemones, pagrįstas šiuolaikinėmis technologijomis (Europos parlamentas ir Europos sąjungos taryba, 2022). Vis dėlto DI naudojimas finansų sektoriuje taip pat kelia naujų iššūkių. DI sistemų veikimas priklauso nuo didelių duomenų kiekių, o tai kelia riziką dėl privatumo pažeidimų ir duomenų saugumo. Įmonės privalo užtikrinti, kad jų DI pagrįstos sistemos atitiktų griežtus Europos Sąjungos reglamentus, susijusius su asmens duomenų tvarkymu ir saugojimu (Ana Hadnes Bruder et al., 2024).

DORA reglamentas (the European Parliament and the council of the European Union, 2022) apima visą finansų sektoriaus technologinę ekosistemą, įskaitant trečiųjų šalių paslaugų teikėjus ir debesijos sprendimus, kuriuos įstaigos naudoja savo veiklai. Šis reglamentas nustato griežtus reikalavimus finansinių įstaigų skaitmeniniam atsparumui ir operaciniam saugumui, pabrėždamas, jog kibernetinės grėsmės valdymas privalo būti vienas iš pagrindinių strateginių prioritetų. Pagal DORA reglamentą, finansinės įstaigos turi nuolat stebėti, vertinti ir valdyti technologinę riziką, o DI gali padėti optimizuoti šiuos procesus, automatizuodamas tiek grėsmių aptikimą, tiek incidentų valdymą realiuoju laiku. Pavyzdžiui, „Darktrace“ sistema naudoja dirbtinio intelekto algoritmus tinklo srauto anomalijoms aptikti ir saugumui užtikrinti. Ši sistema leidžia realiu laiku identifikuoti netipinius veiksmus ir galimas grėsmes. Nepaisant šio privalomumo, „Darktrace“ sistema yra ribota dėl „juodosios dėžės“ modelio veikimo, kur sprendimų priėmimo logika nėra aiški, todėl kyla sunkumų siekiant atitikti DORA skaidrumo reikalavimus, taip pat, „Darktrace“ dažniau orientuojasi į esamų grėsmių atpažinimą, o ne į proaktyvią jų prevenciją (Darktrace, 2024).

Tuo tarpu, „Vectra AI“ naudoja DI algoritmus aptikti anomalijas tinklo srautuose ir identifikuoti vidines grėsmes. Ši sistema ypač naudinga kovojant su grėsmėmis, kurias sukelia kompromituoti vartotojo duomenys ar vidiniai įsilaužimai. Ji taip pat išsiskiria savo galimybe pateikti vizualizacijas, leidžiančias saugumo komandai stebėti tinklo aktyvumą ir aptikti neįprastus veiksmus. Vis dėlto sistema reikalauja nuolatinio duomenų perdavimo į debesijos platformas, keliančio duomenų privatumo ir apsaugos iššūkius, atsižvelgiant į Europos Sąjungos Bendrojo duomenų apsaugos reglamento (GDPR) ir DORA reikalavimus. Be to, „Vectra AI“ veikimas paremtas analizuojant didelius duomenų kiekius, kas gali sukelti dideles veiklos sąnaudas ir apsunkinti mažesnių finansų įstaigų naudojimą (Vectra, 2024). O „Microsoft Azure Sentinel“ yra debesijos pagrindu veikianti saugumo informacijos ir įvykių valdymo bei saugumo automatizavimo ir reagavimo platforma. Kuri finansų įstaigose gali būti naudojama, siekiant centralizuoti grėsmių stebėjimą, aptikimą ir atsaką. Ši sistema naudoja dirbtinio intelekto algoritmus, kad identifikuotų anomalijas, analizuotų tinklo srautus ir optimizuotų kibernetinio saugumo procesus. „Azure Sentinel“ taip pat integruojasi su kitais Microsoft produktais, tokiais kaip „Microsoft Defender“, taip suteikdama platesnį apsaugos spektrą. Tačiau vienas iš pagrindinių šios sistemos trūkumų yra, kad ši sistema priklauso nuo debesijos technologijų. Nors tai suteikia privalumų, susijusių su duomenų prieinamumu ir apdorojimo galia, tačiau taip pat kelia klausimų dėl duomenų privatumo ir apsaugos. Be to, „Azure Sentinel“ modeliai gali generuoti klaidingai teigiamus įspėjimus, dėl ko tokie įspėjimai gali sumažinti saugumo darbuotojų efektyvumą, nes būtų priversti spręsti nereikšmingus incidentus, tuo metu galimai nepastebėdami rimtesnių grėsmių (Microsoft, 2024d; Shlomi Kushchi, 2024).

DORA reglamentas (the European parliament and the council of the European union, 2022) taip pat rekomenduoja finansinėms įstaigoms naudoti pažangias technologijas (9 straipsnis), kad būtų įgyvendinti šie tikslai:

1. **Skaitmeninio atsparumo užtikrinimas** – visos finansinės įstaigos turėtų turėti sistemingą požiūrį į skaitmeninių grėsmių valdymą, įskaitant reguliarius vertinimus, testavimą (24-25 straipsniai). DI gali būti naudojamas siekiant optimizuoti testavimo procesus, automatizuoti grėsmių aptikimą ir testuoti sistemų atsparumą (ENISA, 2020).

2. **IT rizikos valdymas** – finansinės įstaigos turi aktyviai valdyti informacinių technologijų (IT) rizikas (6-7 straipsniai). Kur DI gali būti kaip svarbi priemonė atliekant duomenų analizę, aptikimo ir prevencijos priemones, padedant išvengti IT pažeidimų ir užtikrinant greitą incidentų nustatymą bei sprendimą (IBM, 2024b).
3. **Trečiųjų šalių rizikos valdymas** – DORA akcentuoja tiekimo grandinės saugumą, įpareigojant finansų įstaigas užtikrinti, kad jų IT paslaugų tiekėjai laikytųsi aukščiausių saugumo standartų (28-30 straipsniai). Naudojant DI, galima efektyviau analizuoti ir valdyti šią riziką, vykdyti pažangias sutarčių stebėjimo ir audito procedūras, taip mažinant priklausomybę nuo išorinių paslaugų tiekėjų (Deloitte, 2024b).

Be DORA reglamento, kitas svarbus teisės aktas, kuris sustiprina kibernetinio saugumo ir dirbtinio intelekto taikymo aktualumą finansinių paslaugų sektoriuje, yra NIS2 direktyva, kuri siekia užtikrinti aukštą kibernetinio saugumo lygį ir nurodo, kad visos organizacijos, įskaitant finansų įstaigas, privalo įgyvendinti efektyvias kibernetinio saugumo priemones, apimančias rizikos valdymą, incidentų ataskaitų teikimą, tiekimo grandinės saugumą bei operatyvų reagavimą į incidentus. Ši direktyva reikalauja, kad subjektai, patenkantys į jos taikymo sritį, nuolat stebėtų ir analizuotų tinklų bei informacinių sistemų saugumą, taikydami tinkamas technines ir organizacines priemones (Europos parlamentas ir Europos sąjungos taryba, 2022). Tuo tarpu, dirbtinis intelektas čia gali atlikti svarbų vaidmenį, kuris gali padėti efektyviau įgyvendinti reikalaujamą rizikos valdymo politiką. NIS2 direktyva aiškiai nurodo būtinybę reguliariai vertinti kibernetines grėsmes ir incidentus, o DI sistemos leidžia automatizuoti šį procesą. Pavyzdžiui, DI įrankiai padeda sukurti išmaniuosius grėsmių modelius, kurie leidžia įmonėms įvertinti galimų incidentų poveikį ir sudaryti prevencijos strategijas (ENISA, 2023b). Be to, NIS2 direktyva įpareigoja įmones diegti patikimas incidentų valdymo procedūras, kurios užtikrintų greitą atsigavimą po atakų ir nuolatinį kibernetinio saugumo stiprinimą, kur DI gali padėti įgyvendinti šiuos reikalavimus, nes dirbtinio intelekto sistemos gali greitai reaguoti į kibernetines grėsmes ir palengvinti incidentų valdymą. Tai apima realaus laiko stebėjimą, incidentų analizę bei informacijos pateikimą atsakingoms institucijoms, kaip reikalaujama pagal NIS2 (Europos parlamentas ir Europos sąjungos taryba, 2022).

Kitas svarbus DI technologijos aspektas finansinėse paslaugose yra galimybė modeliuoti galimus kibernetinių grėsmių scenarijus. Naudojant DI, galima kurti išankstinius scenarijus, kurie leidžia finansinėms įstaigoms įvertinti galimų kibernetinių atakų poveikį ir imtis prevencinių priemonių. Šios technologijos leidžia ne tik sumažinti grėsmių riziką, bet ir efektyviai panaudoti išteklius pažeidžiamiausių sričių apsaugai. Tai ypač aktualu atsižvelgiant į Europos Sąjungos skaitmeninių finansų strategiją, kuri pabrėžia kibernetinio saugumo stiprinimo svarbą finansinėms paslaugoms ir didesnę pasitikėjimą skaitmenine infrastruktūra (Europos komisija, 2020a).

Apibendrinant, galima teigti, kad dirbtinio intelekto taikymas finansinių paslaugų saugumui yra būtinas siekiant užtikrinti efektyvų rizikos ir incidentų valdymą, grėsmių aptikimą ir tiekimo grandinės saugumą. DORA reglamentas ir NIS2 direktyvos reikalavimai aiškiai nurodo, kad finansų įstaigos turi imtis griežtų saugumo priemonių ir užtikrinti, jog jų infrastruktūros būtų atsparios kibernetiniams išpuoliams. Naudojant DI galima automatizuoti kibernetinių grėsmių stebėseną, prognozuoti atakų scenarijus ir efektyviai valdyti rizikas, kas žymiai prisideda prie bendros finansinių

paslaugų apsaugos. Sekančiame skyriuje bus analizuojami reikalavimai rizikos scenarijų sudedamosioms dalims.

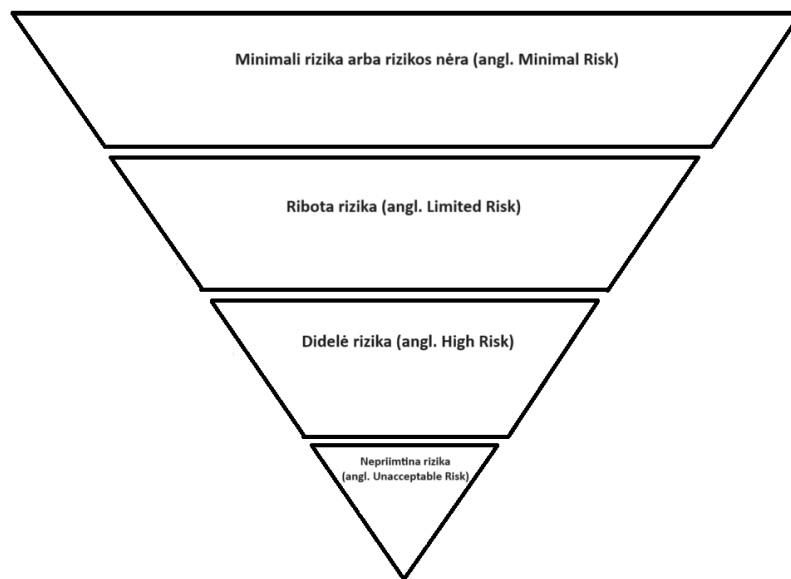
2. REIKALAVIMAI RIZIKOS SCENARIJŲ SUDEDAMOSIOMS DALIMS

Šiame skyriuje bus analizuojama dirbtinio intelekto reguliacinė aplinka, aptariami pagrindiniai rizikos valdymo reikalavimai ir standartai. Taip pat, aprašyti pagrindiniai reikalavimai rizikos scenarijams kurti.

2.1. Dirbtinio intelekto reguliacinė aplinka

Didėjant dirbtinio intelekto paplitimui šiuolaikinėje visuomenėje, atskiros šalys ir tarptautinės bendruomenės deda daug pastangų, kad sukurtų inovacijoms palankią, tačiau saugią reglamentavimo aplinką. Tinkamas reglamentavimas yra labai svarbus siekiant kuo labiau padidinti dirbtinio intelekto technologijų teikiamą naudą ir sumažinti su jomis susijusią riziką. Tačiau tie patys aspektai, kurie sudaro sąlygas siekti socialinės ir ekonominės naudos naudojant dirbtinį intelektą, taip pat gali sukelti naują riziką arba neigiamas pasekmes asmenims arba visuomenei. (Erdélyi & Goldsmith, 2022). Vienas pagrindinių dirbtinio intelekto reguliavimo aplinkos dokumentų Europos Sąjungoje yra Dirbtinio intelekto aktas (DI aktas), kuris kartu su tarptautinių organizacijų, tokių kaip ISO, gairėmis, yra suformavę dirbtinio intelekto reguliacinės aplinkos pagrindą. Toliau šiame poskyryje bus nagrinėjama dirbtinio intelekto reguliacinė aplinka.

Pagrindinė dirbtinio intelekto akto mintis – reguliuoti DI, laikantis rizika grindžiamo požiūrio: kuo didesnė rizika, tuo griežtesnės taisyklės (ES Taryba, 2023). DI akte dirbtinio intelekto rizika nustatoma pagal keturių rizikos lygių kategorijas, kitaip vadinama „Rizikos kritiškumo piramide“ (5 pav.), kuri derinama su moderniu, daugiasluoksniu vykdymo užtikrinimo mechanizmu. Kitaip tariant, kad atitinkamai rizikos kategorijai yra taikomos skirtingos taisyklės. Kuo didesnė rizika, tuo griežtesnės taisyklės yra taikomos. (Kop, 2021)



Sudaryta autorės, pagal Europos komisija, Europos parlamento ir tarybos reglamentas (2021) ir Kop, (2021) EU Artificial Intelligence Act: The European Approach to AI.

5 pav. Dirbtinio intelekto akto kritiškumo piramidė

Nepriimtina rizika laikomi dirbtinio intelekto naudojimo atvejai, kurie yra uždrausti ES. Šie atvejai susiję su kognityviniu elgesio manipuliavimu (manipuliuojami asmenys naudojant pasąmonę veikiančius metodus), emocijų atpažinimu, socialinio vertinimo, bei nuotolinio biometrinio tapatybės nustatymo sistemomis. Manipuliacinės sistemos naudoja pasąmoningus metodus asmens elgesiui iškreipti, sukeliant fizinę ar psichologinę žalą. Tai gali apimti neteisėtą biometrinių informacijos panaudojimą ar duomenų analitiką, kurie pažeidžia asmenų privatumą. Tokios technologijos gali būti naudojamos kenkėjiškoms veikloms, todėl jų reguliavimas ir draudimas yra būtinas siekiant apsaugoti visuomenę nuo pavojingų kibernetinių atakų ir užtikrinti žmogaus teisių laikymąsi. (Veale & Zuiderveen Borgesius, 2021)

Didelės rizikos dirbtinio intelekto sistemomis yra leidžiama naudotis, tačiau joms yra taikomi griežti reikalavimai ir pareigos, kurios yra susijusios su patekimu į ES rinką (Europos komisija, 2021):

- Aukšta duomenų rinkinių, kuriais grindžiama sistema, kokybė, kad būtų kuo labiau sumažinta rizika ir diskriminaciniai rezultatai;
- Veiklos registravimas siekiant užtikrinti rezultatų atsekamumą;
- Išsamūs dokumentai, kuriuose pateikiama visa būtina informacija apie sistemą ir jos paskirtį, kad institucijos galėtų įvertinti, ar ji atitinka reikalavimus;
- Aukštas patvarumo, saugumo ir tikslumo lygis;
- Tinkamos rizikos vertinimo ir mažinimo sistemos;
- Aiški ir tinkama informacija sistemos diegėjui;
- Tinkamos priežiūros priemonės rizikos mažinimui.

Vienas iš pagrindinių reikalavimų didelės rizikos DI sistemoms yra atitikties vertinimas, kurį privalo atlikti tiekėjai prieš pateikiant produktus rinkai. Šis vertinimas grindžiamas esminių

reikalavimų, nustatytų DI akte, laikymusi. Pavyzdžiui, DI sistemų tiekėjai privalo sukurti kokybės valdymo sistemą, įtraukti rizikos valdymo procesus, atnaujinamus per visą sistemos gyvavimo ciklą, ir užtikrinti, kad mokymo duomenų rinkiniai atitiktų kokybės kriterijus, tokius kaip tikslumas, išsamumas ir reprezentatyvumas. Tai ypač svarbu siekiant užkirsti kelią netiesioginei diskriminacijai ir užtikrinti, kad DI sistemos veiktų teisingai ir nešališkai. Be to, DI akte yra numatyti reikalavimai susiję su sistemų saugumu ir patikimumu. Tai apima priemones, skirtas užkirsti kelią kibernetinėms grėsmėms ir užtikrinti sistemų atsparumą netinkamam naudojimui ar kenkėjiškoms atakoms. Didelės rizikos DI sistemų tiekėjai, taip pat privalo užtikrinti techninę dokumentaciją, kuri apima išsamius aprašymus apie sistemos veikimą ir rizikos valdymo priemones, bei užtikrinti tinkamą žurnalo vedimą, leidžiantį atsekti sistemos veikimo istoriją. Visi šie reikalavimai padeda garantuoti, kad didelės rizikos DI sistemos būtų saugios, patikimos ir atitiktų visus teisės aktų reikalavimus. Taigi, dirbtinio intelekto sistemomis Europos rinkoje leidžiama prekiauti užtikrinant atitiktį aukščiau pateiktiems privalomiems reikalavimams. Taip pat, DI sistema priskiriama didelės rizikos sistemoms remiantis numatyta DI sistemos paskirtimi, laikantis galiojančių gaminių saugos teisės aktų. Todėl sistemos priskyrimas didelės rizikos DI sistemai priklauso ne tik nuo sistemos atliekamos funkcijos, bet ir nuo konkrečios tos sistemos paskirties ir jos naudojimo būdų. (Europos komisija, 2021)

Ribota rizika susijusi su dirbtinio intelekto skaidrumu. Dirbtinio intelekto akte yra pateikiami skaidrumo įpareigojimai, kuriais siekiama skatinti pasitikėjimą. Pavyzdžiui, kai naudojamos DI sistemos, kaip: pokalbių robotai, žmonės turi būti informuojami, kad jie sąveikauja su mašina, kad galėtų priimti atitinkamai informaciją. Paslaugų teikėjai, taip pat turi užtikrinti, kad dirbtinio intelekto sukurtą turinį būtų galima atpažinti. Be to, dirbtinio intelekto sukurtas tekstas, skelbiamas siekiant informuoti visuomenę viešojo intereso klausimais, turi būti pažymėtas kaip dirbtinai sukurtas. Ir laisviausiai leidžiamas naudoti minimalios rizikos dirbtinis intelektas, kuris kelia nereikšmingą pavojų sveikatai, saugumui ar pagrindinėms teisėms. Šios sistemos neapima gyvybiškai svarbių funkcijų ar sprendimų ir dažnai naudojamos kasdienėse programose, kur kibernetinės saugos rizikos yra minimalios. Minimali rizika reiškia, kad šios DI sistemos veikia griežtai kontroliuojamose aplinkose, kuriose kibernetinių grėsmių poveikis yra ribotas. Tai gali apimti sistemas, kurios naudojamos asmeniniams asistentams, automatizuotoms klientų aptarnavimo platformoms ar paprastoms rekomendacijų sistemoms internetinėse parduotuvėse. Šios sistemos dažnai naudoja mažiau jautrius duomenis, o jų netinkamas veikimas ar kibernetinis pažeidimas neturi reikšmingo poveikio naudotojų gerovei ar saugumui. (Europos komisija, 2021)

Dirbtinio intelekto reikalavimai saugume ir rizikos valdyme, apima keturias pagrindines temas: techninius saugumo reikalavimus, socialinius-techninius reikalavimus, politinius ir reguliavimo reikalavimus, bei rizikos valdymo strategijas. Techniniai saugumo reikalavimai yra pagrindinė DI sistemų saugumo užtikrinimo dalis. NIST, (2021a) nurodo, kad techniniai rizikos šaltiniai gali būti valdomi naudojant standartinius mašininio mokymosi metodų vertinimo kriterijus, tokius kaip: tikslumas, patikimumas, tvarumas ir atsparumas. Tikslumas (angl. accuracy) nurodo, kiek tiksliai DI sistema atlieka jai pavestas užduotis, o patikimumas (angl. reliability) užtikrina, kad sistema veiktų nuosekliai net ir esant nedideliems pokyčiams duomenyse. Tvarumas (angl. robustness) rodo sistemos gebėjimą išlaikyti efektyvumą įvairiomis sąlygomis, o atsparumas (angl. resilience) apima sistemos gebėjimą atlaikyti kibernetines atakas ir kitus trikdžius. Todėl šis reikalavimas yra neabejotinai svarbus, nes tiesiogiai nulemia DI sistemų patikimumą ir saugumą. Pavyzdžiui, netiksli

arba nepatikima DI sistema gali priimti klaidingus sprendimus, kurie gali turėti rimtų pasekmių. Todėl svarbu nuolat vertinti ir gerinti šiuos rodiklius, siekiant užtikrinti, kad DI sistemos veiktų patikimai ir saugiai.

Socialiniai-techniniai reikalavimai apima ne tik techninius aspektus, bet ir DI sistemų sąveiką su žmonėmis. Kaip nurodo MITRE (2023), svarbu, kad DI sistemos būtų ne tik technologiškai patikimos, bet ir suprantamos vartotojams. Vartotojų pasitikėjimas DI sistemomis didėja, kai jos yra aiškinamos (angl. explainability) ir interpretuojamos (angl. interpretability) suprantamai. Aiškinamumas reiškia, kad vartotojai turi galimybę suprasti, kaip DI sistema priima sprendimus. Tai yra svarbu, nes vartotojai turi žinoti, kodėl sistema priėmė tam tikrą sprendimą, ypač kai tai susiję su svarbiais klausimais, tokiais kaip sveikata ar finansai. Interpretavimas rodo, kad vartotojai gali teisingai interpretuoti sistemos pateiktą informaciją ir priimti pagrįstus sprendimus. Tai padeda sumažinti riziką, kad vartotojai neteisingai supras sistemos veikimą ir priims neteisingus sprendimus.

Politikos ir reguliavimo reikalavimai yra būtini siekiant užtikrinti, kad DI technologijos būtų diegiamos ir naudojamos atsakingai. NIST, (2021b) nurodo, kad DI rizikos valdymas turėtų apimti aiškią ir skaidrią politiką, procesus ir operacines procedūras, taip pat atsakomybės ir vaidmenų apibrėžimą. Šie reikalavimai padeda užtikrinti, kad DI sistemos atitiktų teisinius ir etinius standartus, pvz., žmogaus teisių apsaugą, duomenų privatumą ir atsakomybę. Politiniai ir reguliavimo reikalavimai, taip pat, apima atitiktį tarptautinėms normoms ir standartams. Pavyzdžiui, European commission, (2019) nurodo septynis pagrindinius principus, įskaitant: žmogiškąjį veiksnį ir žmogaus atliekamą priežiūrą, techninį tvarumą ir saugumą, privatumo ir duomenų valdymą, skaidrumą, įvairovę, nediskriminavimą ir teisingumą, bei aplinkos ir visuomenės gerovę ir atskaitomybes. Šie principai padeda užtikrinti, kad DI sistemos būtų kuriamos ir naudojamos atsakingai, atsižvelgiant į visuomenės vertybes ir poreikius.

Be dirbtinio intelekto reikalavimų saugume ir rizikos valdyme, taip pat svarbus yra ir DI sistemų kūrimo ir diegimo aspektas – efektyvus rizikos valdymas. Kur NIST, (2021b) siūlo tris pagrindines funkcijas: konteksto nustatymą ir rizikų išvardijimą (angl. Map function), rizikų analizę ir kiekybinį vertinimą (angl. Measure function), ir rizikų valdymą (angl. Manage function). Šios funkcijos padeda organizacijoms struktūrizuotai identifikuoti, analizuoti ir valdyti rizikas, susijusias su DI sistemomis. Kur konteksto nustatymas apima sistemos aplinkos, tikslų ir galimų rizikų supratimą. Šios funkcijos tikslas yra surasti, atpažinti ir apibūdinti DI sistemos keliamą riziką, kurią vykdant yra surenkama bazinė scenarijaus informacija, kuri padeda priimti sprendimus dėl modelio valdymo, įskaitant sprendimą, kuris gali būti nepagrįstas arba netinkamas. Rizikų analizė ir vertinimas apima rizikų kiekybinį vertinimą, naudojant metrikas ir metodus, siekiant nustatyti rizikų lygį ir jų pasekmes. Rizikos analizės funkcijos tikslas – suvokti rizikos arba poveikio pobūdį, bei jų ypatybes ir palengvinti pačios rizikos valdymą. Ši funkcija yra pagrindas nustatant, kaip turėtų būti valdoma rizika. Ir rizikų valdymas, kuris apima rizikų prioritetų nustatymą ir valdymo strategijų, tokių kaip rizikų vengimas, mažinimas, dalijimasis, perdavimas ar priėmimas, taikymą. Rizikos valdymo funkcijos tikslas yra padėti priimti sprendimus ir atrinkti, bei įgyvendinti rizikos rūšių šalinimo galimybes. Priimant sprendimus turėtų būti atsižvelgiama į aplinkybes ir faktines, bei numanomas pasekmes išorės ir vidaus suinteresuotosioms šalims, taip pat į siūlomos sistemos sąveiką su pasauliu ir į galimas pereinamojo laikotarpio išlaidas (NIST, 2021b).

Standartai, tokie kaip ISO 31000 ir ISO/IEC 27001, yra esminiai įrankiai, siekiant sukurti efektyvią DI reguliacinę aplinką. ISO 31000, (2018) standarte pabrėžiamas sistemingas rizikos valdymo procesas, kuris apima rizikos nustatymą, analizę, vertinimą ir valdymą, kurie yra svarbūs siekiant identifikuoti ir suvaldyti potencialias rizikas, susijusias su DI technologijomis. Šis standartas suteikia bendrą sistemą, kuri gali būti pritaikyta įvairiuose sektoriuose, įskaitant DI, ir padeda organizacijoms suprasti, bei valdyti rizikas, susijusias su naujomis technologijomis. Tuo tarpu, ISO/IEC 27001, (2022) standartas suteikia struktūrą informacijos saugumo valdymui, užtikrinant, kad DI sistemos būtų apsaugotos nuo kibernetinių grėsmių ir duomenų praradimo. Šis standartas nustato informacijos saugumo valdymo sistemos reikalavimus, kurie padeda organizacijoms apsaugoti savo duomenis ir procesus nuo grėsmių ir yra ypač svarbus DI kontekste, kur duomenų saugumas yra kritinis. Be to, ISO/IEC 27001 reikalauja įdiegti tinkamas saugumo kontrolės priemones, kurios apimtų tiek technines, tiek organizacines priemones, siekiant užtikrinti DI sistemų saugumą ir patikimumą. Abu standartai papildė vienas kitą, nes ISO 31000 orientuojasi į bendrą rizikos valdymą, o ISO/IEC 27001 - į konkrečius informacijos saugumo aspektus. Kartu jie suteikia išsamų metodinį pagrindą, leidžiantį organizacijoms ne tik identifikuoti ir valdyti rizikas, bet ir užtikrinti, kad jų DI sistemos būtų saugios ir patikimos. Toks integruotas požiūris yra būtinas siekiant sėkmingai įgyvendinti DI technologijas ir sumažinti galimas rizikas visuomenei.

Nacionalinis standartų ir technologijų institutas (NIST) pateikia išsamias rekomendacijas, kaip užtikrinti dirbtinio intelekto sistemų saugumą, patikimumą ir etiškumą. NIST (2023) akcentuoja atsakingo DI kūrimo svarbą nuo pat sistemos kūrimo pradžios, pabrėžiant pagrindinius principus, tokius kaip: skaidrumas, patikimumas, atsakomybė ir sąžiningumas. Skaidrumas užtikrina, kad DI sprendimai būtų suprantami ir paaiškinami naudotojams, bei suinteresuotiems asmenims, patikimumas rodo, jog DI sistemos veikia pagal numatytus kriterijus, o atsakomybė ir sąžiningumas užtikrina, kad kūrėjai ir naudotojai naudoja technologijas atsakingai, vengdami diskriminacijos ar neetiškų veiksmų. NIST (2024) pateikiamos praktinės rekomendacijos, kaip valdyti ir stebėti generatyvias DI sistemas, pabrėžiant, kad tokios sistemos turi būti reguliariai stebimos ir vertinamos, siekiant užtikrinti jų veikimą pagal numatytus tikslus ir sumažinti galimas rizikas. Generatyvios DI sistemos taip pat turi būti testuojamos realiomis sąlygomis, kad būtų galima identifikuoti ir ištaisyti galimus trūkumus. NIST, (2021) pateikiamos gairės, kaip integruoti atsakingo DI principus į organizacijų veiklą, naudojant rizikos valdymo metodikas, apimančias rizikos nustatymą, analizę, vertinimą ir kontrolę. Šios metodikos padeda užtikrinti, kad DI sistemos būtų saugios ir patikimos. Taigi, NIST pateiktos visos šios gairės suteikia išsamų metodinį pagrindą organizacijoms, siekiant įgyvendinti atsakingą DI kūrimą ir naudojimą, užtikrinant, kad DI sistemos atitiktų aukščiausius etikos, saugumo ir patikimumo standartus. Integruotas požiūris į DI reguliavimą, apimantis skaidrumą, patikimumą, atsakomybę ir sąžiningumą, yra būtinas siekiant sumažinti galimas rizikas ir užtikrinti DI sistemų naudą visuomenei.

Taigi, apibendrinant, dirbtinio intelekto reguliacinė aplinka yra dinamiška ir sudėtinga sritis, reikalaujanti nuolatinio dėmesio ir atnaujinimo. Tarptautiniai standartai, tokie kaip ISO 31000 ir ISO/IEC 27001, bei NIST gairės ir DI aktas, suteikia svarbias gaires ir priemones, kurios padeda užtikrinti DI sistemų saugumą, patikimumą ir etiškumą. ISO 31000 standartas akcentuoja sistemingą rizikos valdymą, o ISO/IEC 27001 nustato reikalavimus informacijos saugumo valdymo sistemoms, siekiant apsaugoti DI sistemas nuo kibernetinių grėsmių ir duomenų praradimo. NIST gairėse,

pabrėžiami atsakingo DI kūrimo ir stebėjimo principai, apimantys skaidrumą, patikimumą, atsakomybę ir sąžiningumą. Šie standartai ir gairės papildo vienas kitą, suteikdami išsamų metodinį pagrindą organizacijoms, siekiant integruoti atsakingo DI principus į savo veiklą. Be to, DI aktas, nustato taisykles ir reglamentus, skirtus užtikrinti DI technologijų saugumą ir etiškumą Europos rinkoje. Todėl, integruotas požiūris į DI reguliavimą, kuris apima tarptautinius standartus, NIST gaires ir DI aktą, yra būtinas siekiant sėkmingai įgyvendinti DI technologijas ir sumažinti galimas rizikas visuomenei, tuo pačiu išlaikant inovacijų potencialą ir skatinant atsakingą technologijų naudojimą. Sekančiame poskyryje bus analizuojami dirbtinio intelekto reikalavimai saugumui ir rizikos valdymui.

2.2. Dirbtinio intelekto rizikos valdymo standartai

Dirbtinio intelekto technologijų plėtra ir pritaikymas įvairiose srityse kelia naujus iššūkius organizacijoms, susijusius su rizikos valdymu. Todėl, rizikos valdymo standartai tampa būtina sąlyga siekiant užtikrinti DI sistemų saugumą, patikimumą ir etiškumą. Šiame poskyryje bus pateikiami vieni iš pagrindinių tarptautinių standartų, kurie reglamentuoja DI rizikos valdymą, ir jų taikymo praktikas.

ISO/IEC 38507 (2022) standarte pateikiamos gairės organizacijoms, kaip valdyti DI naudojimą ir jo keliamas rizikas. Šis standartas pabrėžia valdymo organų atsakomybę užtikrinti, kad DI sistemų naudojimas būtų etiškas, atsakingas ir atitiktų organizacijos strateginius tikslus, kur valdymo organai privalo nuolat stebėti DI technologijų keliamas grėsmes ir galimybes, atsižvelgiant į sparčiai kintančią aplinką. Standarte yra išskiriami keli esminiai aspektai: skaidrumas ir paaiškinamumas, atsakomybės priskyrimas, rizikos valdymas ir vertinimas, bei nuolatinis mokymasis ir prisitaikymas. Skaidrumas ir paaiškinamumas reikalauja, kad organizacijos pateiktų aiškią informaciją apie DI sistemų veikimo principus ir sprendimų priėmimo procesus. Atsakomybės priskyrimas užtikrina, kad organizacijos galėtų identifikuoti ir atsakyti už veiksmus, susijusius su DI sistemų naudojimu, o tai padeda sumažinti riziką ir užtikrinti atitiktį teisiniams reikalavimams. Rizikos valdymas ir vertinimas apima rizikos šaltinių nustatymą, vertinimą ir valdymo priemonių diegimą, siekiant minimizuoti neigiamas pasekmes ir pasinaudoti galimybėmis. Galiausiai, nuolatinis mokymasis ir prisitaikymas yra būtinas norint išlaikyti DI sistemų efektyvumą ir saugumą ilgalaikėje perspektyvoje, nes technologijos nuolat tobulėja ir keičiasi. Šis standartas suteikia išsamias gaires, kaip organizacijos gali valdyti DI rizikas ir užtikrinti, kad DI sistemų naudojimas prisidėtų prie organizacijos sėkmės.

ISO/IEC 27005 (2018) skirtas informacijos saugumo rizikos valdymui ir jo procesams, kurie yra būtini diegiant DI sistemas. Šis standartas pateikia struktūruotą metodologiją, kuri padeda organizacijoms identifikuoti, vertinti ir valdyti rizikas, susijusias su informacijos saugumu. Pagrindinis šio standarto tikslas yra užtikrinti, kad informacijos saugumo rizikos būtų valdomos sistemingai ir efektyviai. Standarte nurodomi pagrindiniai rizikos valdymo proceso etapai: rizikos vertinimas, rizikos valdymas, rizikos stebėjimas ir peržiūra (6 pav.). Rizikos vertinimas apima grėsmių ir pažeidžiamumų identifikavimą, bei jų potencialaus poveikio vertinimą, siekiant nustatyti

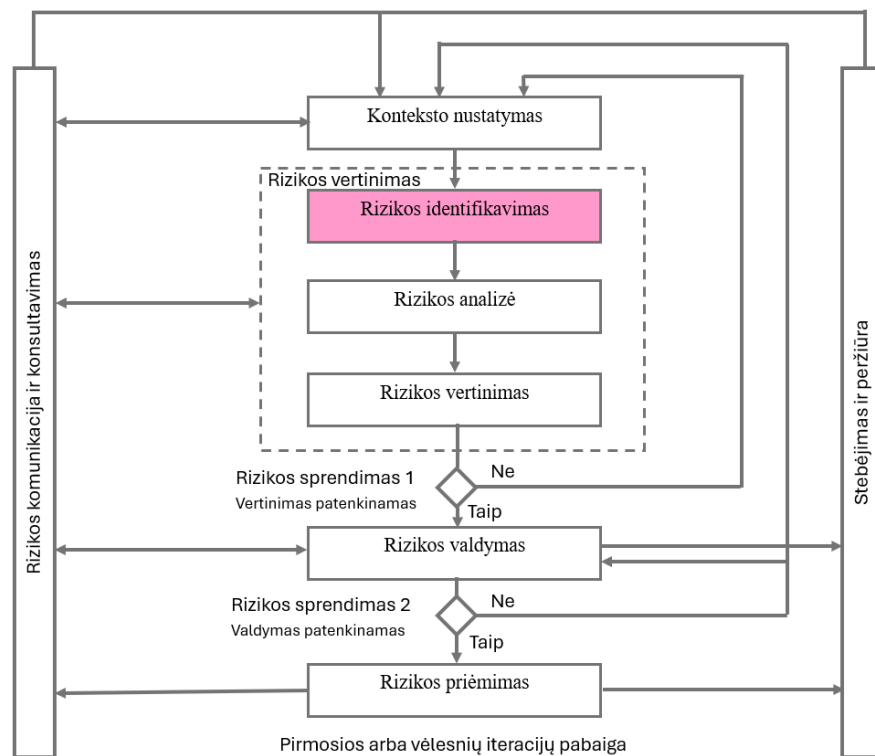
rizikos lygį. Rizikos valdymas susijęs su strategijų ir veiksmų planų kūrimu, kurie padeda sumažinti rizikas iki priimtino lygio. Rizikos stebėjimas ir peržiūra reikalauja nuolatinės rizikos valdymo priemonių efektyvumo kontrolės ir korekcinio veiksmų, jei tai būtina. Pagal 6 pav. matyti, kad rizikos saugumo valdymo procesas vyksta cikliškai, pradedant nuo konteksto nustatymo, pereinant per rizikos vertinimą ir valdymą, ir baigiant rizikos priėmimu. Jei vertinimas ar valdymas nėra patenkinamas, procesas grįžta į ankstesnius etapus, kol pasiekiamas patenkinamas rezultatas. Rizikos valdymo procesas yra nuoseklus ir pradedamas nuo konteksto nustatymo po kurio eina rizikos identifikavimo etapas, kuris apibrėžiamas kaip sistemingas procesas, kuriuo nustatomi potencialūs rizikos šaltiniai, grėsmės, pažeidžiamumai bei galimi poveikio scenarijai organizacijos sistemoms ir procesams. Pagrindinės rizikos identifikavimo etape veiklos (ISO 31000, 2018; ISO/IEC 27005, 2018; ISO/IEC 38507:2022, 2022):

- Konteksto analizė – apibrėžiamas DI sistemų veiklos kontekstas, nustatant vidinius ir išorinius veiksniai, darančius įtaką rizikos aplinkai .
- Grėsmių ir rizikos šaltinių nustatymas – identifikuojami visi galimi rizikos šaltiniai, tokie kaip techniniai trūkumai, duomenų pažeidžiamumas ar organizacinės klaidos .
- Pažeidžiamumų analizė – nustatomos DI sistemų technologinės ir procesinės silpnybės, kurios gali lemti saugumo pažeidimus.
- Poveikio scenarijų formavimas – sudaromi galimi rizikos scenarijai, apžvelgiant grėsmių poveikį organizacijai.

Mokslinėje literatūroje taip pat pabrėžiama, kad rizikos identifikavimo etapo rezultatai nulemia tolimesnio vertinimo ir valdymo priemonių veiksmingumą. Netinkamai atliktas identifikavimas gali sukelti netikslų rizikos vertinimą, dėl ko gali būti priimti neefektyvūs valdymo sprendimai (ISO/IEC 27005, 2018; Microsoft, 2024a). Todėl atsižvelgiant į rizikos identifikavimo etapo svarbą, magistriniame darbe toliau bus apsiribota ties šiuo etapu (6 pav.), kuriant kibernetinio saugumo rizikos scenarijų generatorių finansų įstaigoms. Magistrinis darbas apsiribos tik rizikos identifikavimo etapu dėl kelių esminių priežasčių:

1. Rizikos identifikavimas sudaro pagrindą tolesnėms veikloms, kur be tinkamos grėsmių ir pažeidžiamumų identifikacijos, tolimesnės rizikos vertinimo ir valdymo priemonės tampa neefektyvios (ISO 31000, 2018).
2. Finansų sektoriuje DI sistemos pasižymi specifinėmis grėsmėmis, kurių identifikavimas reikalauja nuoseklios ir sistemingos analizės (ISO/IEC27005, 2022; OECD, 2021a).
3. Dėl ribotų tyrimo laiko ir resursų, darbas koncentruojasi ties rizikos identifikavimo etapu, kuris yra kertinis etapas tolimesniems sprendimams priimti (ISO/IEC27005, 2022).

ISO/IEC 27005 (2018) standarte, taip pat pabrėžiama informacijos saugumo kultūros svarba organizacijoje, akcentuojant, kad visi darbuotojai turi būti įtraukti į rizikos valdymo procesą ir žinoti savo vaidmenį jame. Standarte pateikiamos gairės padeda organizacijoms ne tik sumažinti informacijos saugumo rizikas, bet ir užtikrinti, kad jų informacijos saugumo valdymo sistema būtų nuolat tobulinama ir prisitaikytų prie naujų grėsmių ir technologinių pokyčių.



Sudaryta autorės, pagal ISO/IEC 27005, Information technology — Security techniques — Information security risk management (2018)

6 pav. Rizikos saugumo valdymo procesas

ISO 31000 (2018) standartas pateikia išsamią rizikos valdymo metodiką, kuri gali būti taikoma visose organizacijos veiklos srityse. Šis standartas pabrėžia, kad rizikos valdymas yra nuolatinis procesas, kuris turi būti integruotas į organizacijos valdymo struktūras ir veiklos procesus. Pagrindiniai šio standarto principai apima rizikos valdymo suderinamumą su organizacijos tikslais ir strategija, vadovybės įsipareigojimą ir darbuotojų įsitraukimą, bei nuolatinį rizikos vertinimą ir peržiūrą. Standartas išskiria tris pagrindinius komponentus: principus, sistemą ir procesą. Rizikos valdymo principai nurodo, kad efektyvus rizikos valdymas turi būti sistemingas, struktūruotas ir integruotas, įgalinantis ir skatinantis nuolatinį tobulėjimą. Sistema apima rizikos valdymo integravimą, projektavimą, įgyvendinimą, vertinimą ir tobulinimą, kurie užtikrina, kad rizikos valdymas būtų integruotas į visas organizacijos veiklas. Procesas susideda iš rizikos identifikavimo, analizės, vertinimo ir apdorojimo etapų, taip pat nuolatinės rizikos stebėsenos ir komunikacijos. Standarte, taip pat pabrėžiama, kad rizikos valdymas turi būti pritaikytas prie konkrečios organizacijos konteksto ir aplinkybių, atsižvelgiant į jos specifinius poreikius ir tikslus, kuriame akcentuojama svarba suprasti ir įvertinti tiek vidinius, tiek išorinius veiksnius, galinčius turėti įtakos organizacijos tikslams. Šios gairės padeda organizacijoms kurti tvirtą rizikos valdymo kultūrą, kuri yra esminė siekiant sėkmingo ir tvaraus veiklos vykdymo ilgalaikėje perspektyvoje.

ISO/IEC TR 24028 (2020) pateikia specializuotas gaires rizikos valdymui, susijusiam su DI sistemomis. Šis standartas nagrinėja rizikos šaltinius ir valdymo priemones, pritaikytas DI gyvavimo ciklui, pradedant nuo pradinio projektavimo etapo iki diegimo, eksploatavimo ir nuolatinio tobulinimo. Taip pat, yra pateikiamas rizikos valdymo procesas DI sistemoms, atsižvelgiant į jų

unikalumą ir sudėtingumą. Vienas iš svarbiausių aspektų yra rizikos vertinimas ir kontrolė, ypač atsižvelgiant į DI sistemos sprendimų priėmimo proceso skaidrumą ir paaiškinamumą.

Nors kiekvienas iš aptartų standartų turi savo specifinį fokusą, jų tarpusavio sąveika ir integracija gali suteikti organizacijoms tvirtesnę pagrindą DI rizikos valdymui. Pavyzdžiui, ISO 31000:2018 bendrieji rizikos valdymo principai gali būti naudojami kaip bazė, ant kurios galima statyti specifinius DI rizikos valdymo procesus pagal ISO/IEC 38507:2022 ir ISO/IEC TR 24028:2020. Tuo tarpu, ISO/IEC 27005:2018 užtikrina, kad visos DI sistemos yra saugios ir atitinka informacijos saugumo reikalavimus, kurie yra būtini siekiant išvengti duomenų pažeidimų ir kibernetinių atakų. *Taigi, tarptautiniai standartai, tokie kaip ISO/IEC 38507:2022, ISO/IEC 27005:2018, ISO 31000:2018 ir ISO/IEC TR 24028:2020, suteikia tvirtą pagrindą DI rizikos valdymui. Jie padeda organizacijoms užtikrinti DI sistemų saugumą, efektyvumą ir etiškumą, prisitaikant prie nuolat kintančios technologinės aplinkos. Integruotas ir struktūruotas požiūris į rizikos valdymą leidžia organizacijoms veiksmingai valdyti rizikas ir pasiekti strateginius tikslus. Naudojant šiuos standartus kartu, galima pasiekti geresnių rezultatų, nes jie papildomai stiprina vienas kito teikiamas gaires ir rekomendacijas.* Sekančiame poskyryje bus aprašomi reikalavimai rizikos scenarijams.

2.3. Reikalavimai rizikos scenarijams

Rizikos scenarijai yra būtini veiksmingai rizikos valdymo strategijai, jie padeda identifikuoti galimus pavojus ir vertinti jų poveikį organizacijos veiklai. Tinkamai sukurti ir analizuoti rizikos scenarijai leidžia įmonėms ir organizacijoms pasiruošti galimiems iššūkiams, sumažinti neigiamų įvykių tikimybę, bei greičiau ir efektyviau reaguoti į incidentus. Rizikos scenarijai padeda identifikuoti ir įvertinti potencialias grėsmes, kurios gali turėti neigiamą poveikį jų veiklai. Pagal, MITRE (2016), efektyvūs rizikos scenarijai leidžia suprasti, kokie pavojai gali kilti ir kokios yra jų tikimybės, o tai padeda formuoti tinkamas rizikos valdymo strategijas .

Siekiant, kad rizikos scenarijai būtų veiksmingi, jie turi atitikti kelis svarbius reikalavimus, kurie užtikrina jų tikslumą, išsamumą ir praktiškumą. Pirmiausia, rizikos scenarijai turi būti kuriami pagal sistemingą ir struktūruotą metodiką. Pagal, ENISA, (2022) galima, teigti, kad rizikos scenarijų kūrimo procesas turi būti nuoseklus, įtraukiant visus galimus rizikos šaltinius – nuo techninių gedimų iki žmogiškųjų klaidų ir išorinių grėsmių, tokių kaip kibernetiniai išpuoliai ar gamtinės katastrofos. Ši struktūruota metodika užtikrina, kad nė viena rizikos sritis neliks neaptarta, o scenarijai bus išsamūs ir apims visas galimas grėsmes. Be to, rizikos scenarijai turi būti realistiški ir pagrįsti faktiniais duomenimis. Todėl, remiantis, NIST (2018) galima, teigti, kad efektyviems rizikos scenarijams būtina atlikti išsamų rizikos vertinimą, remiantis istorine incidentų analize ir patikimais informacijos šaltiniais.

Antra, rizikos scenarijai turi būti dinamiški ir nuolat atnaujinami. Kintančios technologinės, ekonominės ir politinės sąlygos gali reikšmingai pakeisti rizikos aplinką, todėl rizikos valdymo procesas turi būti lankstus ir prisitaikantis prie naujų iššūkių. Pagal, ISO 31000, (2018) standartą

galima daryti išvadas, kad yra rekomenduojama reguliariai peržiūrėti ir atnaujinti rizikos scenarijus, siekiant užtikrinti jų aktualumą ir efektyvumą. Tai apima naujų grėsmių identifikavimą, esamų rizikos vertinimų peržiūrą ir prevencinių, bei reagavimo veiksmų atnaujinimą. Taip pat, svarbu įtraukti tiek vidines, tiek išorines grėsmes. ISO/IEC 27005, (2018) standartas mini, kad rizikos scenarijai turėtų apimti visus organizacijos veiklos aspektus, įskaitant technines infrastruktūras ir žmogiškuosius išteklius. Tai reiškia, kad rizikos valdymo komanda turi glaudžiai bendradarbiauti su įvairiais departamentais, siekdama suprasti jų specifines grėsmes ir rizikas. DI aktas (Europos komisija, 2021), teigia, kad konteksto svarba turėtų būti, taip pat, labai svarbi, todėl rizikos scenarijų kūrimo turėtų būti apsibrėžtos specifinės sąlygos, kuriomis rizika gali pasireikšti. Tai apima scenarijaus ribų nustatymą, galimų įvykių seką ir jų pasekmes, padedant organizacijoms geriau suprasti, kaip konkrečios grėsmės gali paveikti jų veiklą ir kokių veiksmų reikia imtis, siekiant sumažinti riziką.

Galiausiai, rizikos scenarijai turi būti aiškiai dokumentuoti ir lengvai suprantami visiems organizacijos nariams, įskaitant ir ne techninius darbuotojus. Aiškumas ir suprantamumas yra būtini, kad visi organizacijos nariai galėtų efektyviai dalyvauti rizikos valdymo procese ir žinotų savo vaidmenį ir atsakomybes incidento atveju. NIST, (2018) mini, kad dokumentavimas turi būti atliktas taip, kad informacija būtų lengvai prieinama ir suprantama visiems suinteresuotiems asmenims. Ir, rizikos scenarijai turi būti suderinami su organizacijos bendrąja rizikos valdymo strategija ir integruoti į platesnį rizikos valdymo procesą. Tai užtikrina, kad scenarijai nebūtų izoliuoti dokumentai, bet aktyvi rizikos valdymo dalis, kuri prisideda prie bendrų organizacijos saugumo ir tęstinumo tikslų. ISO 31000, 2018 standarte, taip pat, yra aprašoma suderinamumo svarba, siekiant veiksmingo ir integruoto rizikos valdymo.

Rizikos scenarijus sudaro keli esminiai komponentai, kurie leidžia tiksliai įvertinti grėsmes ir planuoti veiksmingas atsakomąsias priemones. Pirmiausia, tai yra grėsmės aprašymas. Ši dalis turi būti itin detali, kuri apimtų ne tik grėsmių šaltinius, bet ir jų veikimo būdus. Toks aprašymas leidžia organizacijai geriau suprasti, kaip gali būti vykdomos atakos ir kokių priemonių reikėtų imtis siekiant apsisaugoti (NIST, 2012a). Pažeidžiamumų identifikavimas yra kita kritiškai svarbi scenarijaus dalis. Pažeidžiamumai gali būti technologiniai, organizaciniai ar net žmogiškieji. ISO 31000, (2018) rekomenduoja išsamiai analizuoti organizacijos sistemas, procesus ir personalą, siekiant nustatyti silpnąsias vietas, kurios gali būti išnaudotos. Pažeidžiamumų identifikavimas leidžia ne tik suprasti, kuriose vietose yra didžiausi pažeidžiamumai, bet ir padeda prioritetizuoti saugumo priemones.

Tikėtinų pasekmių analizė yra būtina, siekiant įvertinti, kokią poveikį grėsmės gali turėti organizacijai. Pagal NIST (2012), svarbu įvertinti tiek tiesioginius, tiek netiesioginius padarinius, įskaitant finansinius nuostolius, reputacijos pažeidimus, veiklos sutrikimus ir teisinius padarinius. Ši analizė padeda suvokti galimų incidentų mastą ir parengti tinkamas atsakomąsias priemones. Atsakomosios priemonės yra esminė rizikos scenarijaus dalis. Jos apima konkrečių veiksmų planavimą, kurių organizacija imsis įvykus incidentui. ENISA (2022) rekomenduoja, kad šios priemonės turi būti praktiškos ir lengvai įgyvendinamos, apimančios tiek techninius, tiek organizacinius veiksmus. Tai gali apimti incidentų aptikimo ir reagavimo sistemas, atsarginių kopijų kūrimą, avarinius planus ir komunikacijos strategijas. Atsakomosios priemonės turi būti reguliariai testuojamos ir atnaujinamos, siekiant užtikrinti jų veiksmingumą realiomis sąlygomis.

Apibendrinant, rizikos scenarijai yra esminė rizikos valdymo dalis, kuri leidžia efektyviai pasiruoti ir reaguoti į galimas grėsmes. Efektyvūs rizikos scenarijai turi būti realistiški, išsamūs, pagrįsti tiksliais duomenimis ir nuolat atnaujinami, kad atitiktų nuolat besikeičiančias grėsmes ir organizacijos pokyčius. Svarbiausios rizikos scenarijaus sudedamosios dalys apima grėsmių aprašymą, pažeidžiamumų identifikavimą, tikėtinų pasekmių analizę ir atsakomųjų priemonių planavimą. Kiekviena iš šių dalių turi būti integruota į bendrą rizikos valdymo strategiją ir reguliariai peržiūrima, siekiant užtikrinti veiksmingumą ir atitiktį naujausioms grėsmėms, bei pokyčiams. Tokiu būdu sukuriama veiksmingi rizikos scenarijai, padedantys apsisaugoti ir užtikrinti veiklos tęstinumą. Sekančiame skyriuje bus pateikiama dirbtinio intelekto sistemų rizikos scenarijų generavimo metodologija.

3. DIRBTINIO INTELEKTO SISTEMŲ RIZIKOS SCENARIJŲ GENERAVIMO METODOLOGIJA

Šiame skyriuje bus aprašomos rizikos scenarijaus sudedamosios dalys, rizikos scenarijaus elementų taksonomijos, dirbtinio intelekto įrankių palyginimas, užklausos konteksto formavimas, bei užklausos standarto apsibrėžimas dirbtinio intelekto rizikos scenarijams generuoti.

3.1. Rizikos scenarijaus struktūra ir elementų taksonomija

Rizikos scenarijus, tai rizikos valdymo komponentas, kuris aprašo galimus rizikos įvykius, apibūdinant nepageidaujamą situacinę aplinką ir pasekmes (Tony Martin-Vegue, 2021). Jų kūrimas padeda vadovautis rizikos vertinimo procesu, apibrėžiant ir išskaidant veiksmus, kurie prisideda prie nepageidaujamų reiškinių dažnumo ir dydžio. Kad būtų sukurtas išsamus rizikos scenarijus, jis susideda iš komponentų, tokių kaip: veikėjai, grėsmių įvykiai, motyvai, padariniai, turtas ir laikas (Morland-Austin IT GRC Knowledge Base, n.d.) (7 pav.). Toliau šiame poskyryje bus aprašomos rizikos scenarijų sudedamosios dalys ir apžvelgiamos jų taksonomijos.

Rizikos scenarijų sudedamosios dalys					
Veikėjas/Grėsmių bendruomenė (angl. Actor/Threat Community)	Grėsmių įvykiai (angl. Threat Event)	Ketinimai/Motyvai (angl. Intent/Motivation)	Padarinys (angl. Effect)	Turtas/Ištekliai (angl. Asset/Resource)	Laikas (angl. Timing)

Sudaryta autorės, pagal Morland-Austin IT GRC Knowledge Base, n.d., IT Risk Scenarios

7 pav. Rizikos scenarijaus sudedamosios dalys

Įvairūs veikėjai gali siekti pakenkti dirbtinio intelekto sistemoms naudojant kibernetines priemones. Kibernetiniai nusikaltėliai dažniausiai yra motyvuoti pelno siekimu, jie gali panaudoti DI vykdymui atakų, taip pat ir išnaudoti esamus DI sistemų pažeidžiamumus. Pavyzdžiui, jie gali bandyti nulaužti dirbtinio intelekto palaikomus pokalbių robotus, kad pavogtų kredito kortelių ar kitus duomenis, arba inicijuoti programinės įrangos ataką prieš dirbtinio intelekto pagrįstas tiekimo grandinės valdymo sistemas. Šie veikėjai yra išskirstomi į dvi grupes: vidinę (personala, rangovus) ir išorinę (konkurentai, pašaliniai asmenys, verslo partneriai, reguliavimo institucijos, rinkai). Vidinė grupė prieina prie viešai neatskleidžiamos informacijos, taip pat, turi prieigas prie vidinių tinklų. Tokie asmenys gali siekti pavogti ar sabotuoti duomenų rinkinius, naudojamus dirbtinio intelekto sistemose (ENISA, 2020).

Išorinių veikėjų grupė yra kur kas platesnė. Vieni iš jų yra valstybės remiami, kurie yra labiau pažengę, nei vidinės grupės veikėjai, jų pagrindinis tikslas yra panaudoti dirbtinio intelekto sistemas atakoms prieš kitas šalis (įskaitant pramonės šakas ir ypatingos svarbos infrastruktūros objektus), taip pat naudojant DI sistemas vidiniams tinklams apsaugoti. Šie veikėjai aktyviai ieško dirbtinio intelekto sistemų pažeidžiamumų, kuriais galėtų pasinaudoti kaip priemone pakenkti kitai šaliai. Kiti grėsmę keliantys veikėjai yra teroristai, kurie siekia sukelti fizinę žalą ar net gyvybių atėmimą. Pavyzdžiui, teroristai gali norėti įsilaužti į bepiločius automobilius, kad panaudotų juos kaip ginklą. Mažiau sudėtingi grėsmių veikėjai yra "script kiddies", kurie yra motyvuoti ideologiškais nusikaltimais. Tai dažniausiai būna nepatyrę asmenys, kurie naudoja kitų parašytus kodus ar scenarijus, kad atakuotų sistemas. Paprastai, šiems veikėjams trūksta kompetencijos patiems parašyti kodus. Ir konkurentai, kurie vis dažniau demonstruoja ketinimus atakuoti savo varžovus siekiant įgyti didesnę rinkos dalį (ENISA, 2020).

Grėsmės veikėjai naudoja taktikas: grėsmės įvykius, prieš objektus, kuriuos nori pažeisti. Šiuos grėsmių įvykius apima skirtingi neigiami įvykiai, kurie yra suskirstyti į priešiškus (žiūrėti 1 priedą) ir neprieiškus (žiūrėti 2 priedą) atvejus, kurių kiekvienas turi detalius aprašymus, padedančius identifikuoti specifinius rizikos veiksnius ir jų įtakas aplinkiniams. Tai gali apimti įvairias situacijas – nuo duomenų vagysčių ir kibernetinių atakų iki gamtinių nelaimių ir techninių klaidų. Visi šie grėsmių įvykiai yra sukeliama grėsmių šaltinių, kurie apibūdinami kaip: metodai, kuriais siekiama pasinaudoti arba situacijos, kurios gali būti netyčia panaudotos pažeidžiamumui (NIST, 2012b).

ENISA (2020) dokumente, taip pat yra pateikiama grėsmių klasifikacija siekiant identifikuoti ir analizuoti pagrindines grėsmes, susijusias su dirbtinio intelekto sistemomis (žr. 4 priedą). Ši taksonomija yra sudaryta pagal poveikio tipą klasifikuojant grėsmes ir grėsmių kilmę, leidžiant taikyti diferencijuotą požiūrį į riziką. Pavyzdžiui, neteisėtas duomenų naudojimas gresia duomenų konfidencialumo praradimu, kas yra itin aktualu finansų įstaigoms, saugančioms jautrią informaciją. Tokios grėsmės gali atsirasti dėl nepakankamai saugomų duomenų ar per daug liberalių prieigos teisių. Tuo tarpu, manipuliacijos DI sprendimų logika yra susijusi su neteisingu ar piktnaudžiavimu grįstu duomenų panaudojimu modelių mokyme, kas gali lemti klaidingus ar šališkus sprendimus, pavyzdžiui, kredito suteikimo ar sveikatos priežiūros sektoriuose. Fiziniai išpuoliai prieš DI

infrastruktūrą kelia tiesioginį pavojų įrangai ir duomenų bazėms, kas gali sutrikdyti ne tik kasdienę veiklą, bet ir ilgalaikę organizacijos strategiją.

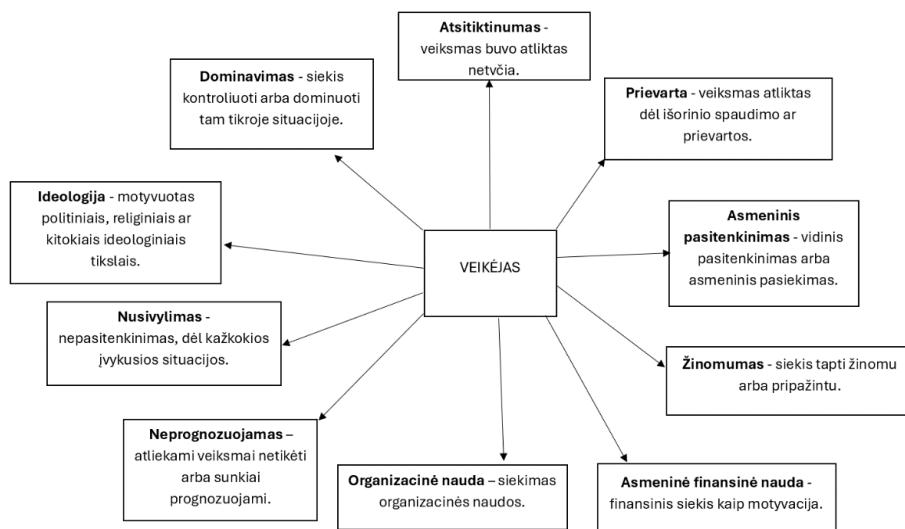
Grėsmių taksonomijos yra suskirstytos į kelias pagrindines kategorijas (ENISA, 2020):

- **Nedoras veiksmas/piktnaudžiavimas:** tai sąmoningi veiksmai, nukreipti prieš informacinių ir ryšių technologijų sistemas, siekiant pavogti, pakeisti ar sunaikinti duomenis ar įrangą.
- **Vengimas/įsibrovimas/įsilaužimas:** veiksmai, kurie apima neleistiną klausymąsi, trukdymą ar kontrolę be trečios šalies sutikimo.
- **Trečiųjų šalių fizinių turtų panaudojimas:** veiksmai, siekiant atskleisti, sunaikinti ar pavogti fizinius išteklius, tokius kaip: infrastruktūra, įranga ar neautorizuota prieiga.
- **Netyčinė žala:** atsitiktiniai veiksmai, kurie sukelia sugadinimą, žalą ar sužeidimus.
- **Sistemos gedimai:** iš dalies ar visiškai neveikiančio turto gedimai.
- **Trikdžiai:** netikėti paslaugų teikimo sutrikimai ar kokybės pablogėjimas, nukritęs žemiau reikalaujamo lygio.
- **Katastrofos:** staigūs įvykiai ar gamtinės katastrofos, sukeliančios žalą ir netektis.
- **Teisiniai reikalavimai:** trečiųjų šalių teisiniai veiksmai, siekiant kompensuoti nuostolius ar laikytis galiojančių teisės aktų.

Šios grėsmių taksonomijos kategorijos ne tik apibrėžia grėsmes, bet ir analizuoja jų poveikį DI sistemoms, jų turtams ir poveikį pagal kiekvieną grėsmės tipą. Patys grėsmių įvykiai taip pat gali būti aprašomi: labai bendrais terminais (pvz., sukčiavimas), labiau aprašomaisiais terminais, naudojant taktiką, metodus ir procedūras, arba labai konkrečiais terminais (pvz., konkrečių informacinių sistemų, technologijų, organizacijų, vaidmenų ar vietų pavadinimais). Taip pat, yra apsvarstomi reprezentatyvūs pavojingų įvykių rinkiniai, kurie gali būti, kaip atspirties taškas nustatant konkrečius pavojingus įvykius. Pavyzdžiui, organizacijos gali atsižvelgti tik į tuos grėsmės įvykius, kurie buvo pastebėti (viduje arba organizacijose, kurios yra lygiavertės / partnerės) arba visus galimus grėsmės įvykius. Šie grėsmių įvykiai yra nustatomi pagal įvykio tikimybes, kurios yra svertiniai rizikos veiksniai, pagrįsti tikimybėmis, kad tam tikra grėsmė gali pasinaudoti tam tikru pažeidžiamumu (arba pažeidžiamumų rinkiniu), analize. Tikimybės rizikos veiksnys apima tikimybes, kad grėsmės įvykis bus inicijuotas, įvertį su poveikio tikimybės įverčiu (t. y. tikimybė, kad grėsmės įvykis sukels neigiamą poveikį). Priešiškų grėsmių atveju įvykio tikimybės paprastai grindžiamos: priešininko ketinimais, priešininko pajėgumu ir priešininko taikymusi. Išskyrus priešiškos grėsmės įvykius, įvykio tikimybė apskaičiuojama naudojant istorinius įrodymus, empirinius duomenis ar kitus veiksnius. Grėsmės įvykio inicijavimas arba atsiradimo tikimybė aprašoma atsižvelgiant į konkretų laikotarpį (pvz., ateinančius šešis mėnesius, kitus metus arba laikotarpį, kol bus pasiektas konkretus orientyras) (NIST, 2012b).

Veikėjas prieš atliekant grėsmės įvykį turi turėti ketinimus/motyvacijas, dėl ko nori atlikti tam tikrus veiksmus, kurie gali potencialiai išnaudoti turtą. Motyvacija dažniausiai informuoja koks bus atakos intensyvumas ir atkaklumas, taip pat veikėjai paprastai veikia pagal savo emociją ar tam tikras situacijas, kurios paprastai informuoja apie atakos būdą. Yra apibrėžti 10 motyvacijos elementų, kurie apibūdina visus pagrindinius motyvus, kodėl veikėjas atlieka tam tikrus veiksmus (grėsmės įvykius) (8 pav.). Pavyzdžiui, ideologijos motyvuotas veikėjas gali turėti pakankamai kantrybės siekti

ilgalaikių tikslų ir tyliai dirbti ties šiuo tikslu metų metus, o kibervandalas ieško smūgių, kurie sukurtų intensyvų ir dėmesį patraukiantį išpuolį ir labai greitai praranda susidomėjimą ties tuo objektu. (Casey & Rosenquist, 2016)



Sudaryta autorės, pagal Casey & Rosenquist, Understanding Cyberthreat Motivations to Improve Defens (2016)

8 pav. Motyvacijos elementai

Padarinio komponentas nurodo pasekmes, kurios gali atsirasti dėl grėsmių įvykių ir yra skirstomos į tris pagrindines kategorijas: konfidencialumo pažeidimus, vientisumo sutrikimus ir prieinamumo sumažėjimus. Konfidencialumo pažeidimai dažniausiai siejami su duomenų praradimu, kurie yra vieni iš labiausiai paplitusių incidentų grėsmių. Vientisumo sutrikimai yra svarbūs nustatant finansinius sukčiavimo atvejus, kurie dažnai įvyksta dėl nepageidaujamo pakeitimo ar turinio manipuliavimo. Prieinamumo sumažėjimai aptinkami dažniausiai per pertraukimo scenarijus, kurie apima veiklų sutrikdymą. Todėl galima teigti, kad, padarinio komponentas leidžia geriau suprasti, kokią riziką kiekviena grėsmė kelia, ir yra svarbus renkant atsakymus, bei vertinant saugumo priemonių efektyvumą (Ellerhold et al., 2023).

Svarbiausias grėsmės scenarijaus elementas yra turtas, kuriam kyla grėsmė. Turtas apibrėžiamas kaip bet kas, kas turi vertę asmeniui ar organizacijai, todėl reikalauja apsaugos. Dirbtinio intelekto atveju turtas taip pat yra tas, kuris yra labai svarbus siekiant patenkinti poreikius, kuriems jis naudojamas, tai būtų: duomenys, programinė įranga, aparatinė įranga, ryšių tinklai, taip pat, dirbtinis intelektas turi turtą, tokį kaip: modeliai, procesoriai, artefaktai ir kiti (žr. 3 priedą), bei duomenys, modeliai, aktoriai (sistemos naudotojai ar priežiūros personalas), procesai ir įrankiai (žr. 5 priedą), kurie gali būti pažeisti ir (arba) sugadinti dėl tyčinių / netyčinių priežasčių (ENISA, 2020). Pavyzdžiui, duomenys yra vienas iš pagrindinių DI sistemų komponentų, ir jų kiekis bei kokybė gali žymiai paveikti grėsmių poveikį. Didelis duomenų kiekis gali sukelti sunkumų neteisėtai prieigai prie duomenų ar jų manipuliavimui, nes sudėtingesnė duomenų struktūra ir apimtis reikalauja sudėtingesnių metodų, kad būtų galima juos analizuoti ar klastoti. Kita vertus, dėl didelio duomenų

kiekio gali kilti problemų užtikrinant duomenų kokybę ir apsaugą, kadangi kiekvienas duomenų elementas turi būti tinkamai apsaugotas, o tai gali reikalauti papildomų išteklių ir sudėtingesnių duomenų tvarkymo procesų. Tuo tarpu aukštos kokybės duomenys gali žymiai padidinti DI sistemų efektyvumą, sumažinti klaidų tikimybę ir pagerinti sprendimų tikslumą. Tačiau aukštos kokybės, tiksli ir vertinga informacija taip pat gali tapti prioritetiniais taikiniais kibernetiniams nusikaltėliams, siekiantiems išnaudoti šiuos duomenis neteisėtam naudojimui.(ENISA, 2020)

Modeliai dirbtinio intelekto sistemose, kaip dar viena svarbi turto taksonomija, turinti didelę įtaką sistemos veikimui ir sprendimų kokybei, taip pat yra jautrūs įvairioms grėsmėms. Šios grėsmės apima tyčinį modelių klastojimą, kur modeliai gali būti modifikuojami taip, kad duotų klaidingus rezultatus ar būtų naudojami piktnaudžiavimo tikslais, pavyzdžiui, finansinėse sukčiavimo schemose ar dezinformacijos platinime. Kitas grėsmių šaltinis yra neteisingai parinkti ar manipuluoti mokymo duomenys, kurie gali būti sąmoningai arba netyčia suklaidinti modelio mokymo procese, dėl ko modelis įgyja klaidingą mokymo logiką ir generuoja neteisingus sprendimus, kas gali turėti tiesioginių pasekmių tokiose srityse kaip finansų valdyme. Be to, modeliai gali būti pažeidžiami dėl sistemos naudotojų ar priežiūros personalo klaidų – netinkamas modelio naudojimas, nepakankamas supratimas apie modelio veikimą ar netinkamas duomenų valdymas gali lemti netikslumus ar nukrypti nuo numatytos operacijų eigos. (ENISA, 2020)

Laiko komponentas rizikos scenarijuose, parodo, kaip rizika gali keistis priklausomai nuo laiko, ir kaip laiku įgyvendintos prevencinės priemonės gali ženkliai sumažinti ar neutralizuoti potencialią grėsmę. Ypatingas dėmesys skiriamas laikotarpiams, kuriuose apibrėžiami optimalūs laiko intervalai efektyviam grėsmių šalinimui. Pavyzdžiui, tam tikrų saugumo atnaujinimų ar patikros procedūrų vykdymas per tam tikrą laikotarpį gali padidinti jų efektyvumą, sumažinant riziką, kuri gali kilti dėl išorinių ar vidinių veiksmų. Taip pat, laiku atlikti veiksmai leidžia efektyviai valdyti ir mažinti su laiku susijusias rizikas. Tai apima grėsmių nustatymą, jautrumo periodų identifikavimą ir adekvačių atsako priemonių taikymą kritiniais momentais. Laiko aspektas taip pat susijęs su gebėjimu prognozuoti ir reaguoti į grėsmes taip, kad būtų išvengta ar sumažinta jų įtaka turtui (Young, 2019).

Apibendrinant, galima teigti, kad rizikos scenarijus yra struktūruotas pasakojimas, kuriame komponentai – veikėjai, grėsmės įvykiai, motyvai, padariniai, turtas ir laikas – sąveikauja tarpusavyje, sudarydami nuoseklią istoriją apie galimas rizikas ir jų valdymą. Veikėjai yra grėsmės šaltiniai, kurių identifikavimas leidžia nustatyti potencialius pavojus, o jų motyvacija – ideologinė, finansinė ar asmeninė – paaiškina veiksmų paskatas. Grėsmės įvykiai – tai konkretūs incidentai, atskleidžiantys saugumo spragas, kurios gali paveikti turtą, apimantį fizinę infrastruktūrą, informacines sistemas ir žmogiškuosius išteklius. Padariniai nusako pasekmes, tokias kaip finansiniai nuostoliai, reputacijos žala ar operacijų sutrikimai, tuo tarpu laikas pabrėžia grėsmės aktyvumo momentą ir reagavimo būtinumą. Šiuos elementus papildančios rizikos scenarijų taksonomijos yra viena iš esminių priemonių efektyviai identifikuoti, struktūrizuoti ir valdyti grėsmes. Taksonomijos padeda suskirstyti grėsmes pagal jų kilmę, poveikį ir pažeidžiamumus, taip užtikrinant tikslingą ir sistemingą rizikos valdymą. Toks požiūris yra ypač svarbus dirbtinio intelekto saugumo strategijose, kur būtina ne tik reaguoti į potencialias grėsmes, bet ir planuoti atsparumo stiprinimo veiksmus, apsaugant duomenis, sistemas ir vartotojus. Išsamios taksonomijos leidžia kurti saugesnes sistemas,

numatyti galimus incidentus ir užtikrinti efektyvų grėsmių mažinimą bei valdymą. Sekančiame poskyryje bus atliekama palyginamoji analizė dirbtinio intelekto įrankių.

3.2. Generatyvinių DI įrankių palyginimas

Kuriant finansinėms įstaigoms kibernetinio saugumo rizikos scenarijų generatorių, itin svarbu pasirinkti tinkamą generatyvinį dirbtinio intelekto įrankį, kuris užtikrintų aukštą saugumą, tikslumą ir pritaikomumą. Tokie įrankiai kaip: ChatGPT-4, Google Gemini, IBM Watson AI ir Microsoft Copilot pasižymi skirtingomis techninėmis savybėmis, įskaitant architektūrą, žetonų (angl. token) dydį, mokymo duomenų apimtį, kalbos supratimo tikslumą, integracijos galimybes, saugumo mechanizmus ir kainodarą. Ši analizė padės toliau įvertinti, kuris iš jų geriausiai atitinka finansų sektoriaus poreikius, ypatingą dėmesį skiriant duomenų apsaugai, integracijai su kitomis platformomis ir galimybei generuoti itin tikslius bei pritaikomus scenarijus.

3.2.1. ChatGPT-4

ChatGPT-4 architektūra yra pagrįsta transformatoriaus (angl. „Transformer“) modeliu, kuris yra laikomas vienu iš pažangiausių natūralios kalbos apdorojimo technologijų. Transformatoriuje naudojamas savęs dėmesio (angl. „self-attention“) mechanizmas, kuris analizuoja ryšius tarp teksto elementų bet kokių atstumu – siekiant apdoroti ilgalaikius kontekstus tekste. ChatGPT-4 sukurtas remiantis transformatoriaus dekoderio principu, kuris sudarytas iš daugybės sluoksnių, kurių kiekviename naudojamos daugiafunkcės galvutės, skirtos apdoroti skirtingus įvesties aspektus. Kiekvienas sluoksnis susideda iš savęs dėmesio mechanizmo ir užmaskuoto dėmesio (angl. „masked attention“), leidžiančio modeliui generuoti tekstą seka po sekos. Papildomai, naudojami pozicinės koduotės (angl. „positional encodings“) metodai užtikrina, kad modelis suprastų teksto eiliškumą. GPT-4 taip pat turi daugiau nei 1,75 trilijono parametrų, kas suteikia galimybę atlikti sudėtingesnes užduotis, tokiu būdu praplečiant natūralios kalbos apdorojimo galimybes. Ši architektūra leidžia efektyviai derinti didelio masto duomenų apdorojimą, tikslumą ir gebėjimą generuoti itin kokybišką tekstą, priartinant jį prie žmogaus rašymo stiliaus (Bhattacharya et al., 2024).

GPT-4 taip pat yra apmokytas su 240 TB tekstinių duomenų, kurie apima įvairius šaltinius, tokius kaip: moksliniai straipsniai, viešai prieinami duomenų rinkiniai ir licencijuotas turinys. ChatGPT-4 mokymo procesas yra išskaidytas į dvi pagrindines dalis – išankstinį mokymą (angl. „pre-training“) ir koregavimą (angl. „finetuning“), kur išankstiniame mokyme yra mokama suprasti bendrinės kalbos vartojimą ir kontekstą, o koregavimo dalyje vyksta tikslinimas konkrečiame duomenų rinkinyje, kad būtų optimizuotas GPT-4 našumas konkrečiose srityse, pvz., atsakymams į klausimus. Mokymo metodologija leidžia modeliui efektyviau apdoroti didelius duomenų rinkinius, įsisavinti kalbinius modelius ir generuoti teisingus atsakymus (Bhattacharya et al., 2024). GPT-4 išsiskiria dar vienu gebėjimu, tai apdoroti didelius tekstinius fragmentus, nes jo kontekstinis langas

palaiko iki 128 tūkst. žetonų (angl. „token“), o maksimalus išvesties žetonas gali būti net iki 16,384 (OpenAI, n.d.-a). Žetonas yra pagrindinis vienetas, kuriuo modelis apdoroja tekstą, apimantis žodžius, jų dalis ar net atskirus simbolius. Didesnis kontekstinis langas leidžia modeliui analizuoti ilgesnius tekstus, tokius kaip finansines ataskaitas, išsamias technines dokumentacijas, be būtinybės skaidyti juos į mažesnes dalis. Šis pranašumas suteikia didesnę lankstumą taikant GPT-4 realiose situacijose, kur svarbu išlaikyti turinio vientisumą ir kontekstą. Be to, didesnis žetonų limitas sumažina informacijos praradimo ar iškraipymo riziką ilgesniame tekste, todėl modelis gali geriau suprasti ir generuoti prasmingus atsakymus net ir sudėtinguose scenarijuose (OpenAI, 2024h).

GPT-4 pasižymi dideliu kalbos supratimo tikslumu ir lankstumu įvairiose užduotyse, įskaitant teksto generavimą, vertimus, santraukas ir techninius paaiškinimus. Modelio veikimas vertinamas naudojant standartinius vertinimo testus, tokius kaip MMLU (Massive Multitask Language Understanding), kur GPT-4 rezultatai siekia virš 86% tikslumo NLP užduotyse. Šios pažangos pagrindas – modelio gebėjimas geriau analizuoti sudėtingus lingvistinius priklausomumus, apdoroti ilgesnius tekstus ir tiksliau generuoti atsakymus, atitinkančius užduoties kontekstą (OpenAI, 2023b).

ChatGPT-4 gali būti integruotas į įvairias programas per OpenAI API, leidžiančią kūrėjams naudoti modelio galimybes savo produktuose ir paslaugose (kaip pavyzdžiui: debesų paslaugose (pvz., „Azure“), mobiliosiose aplikacijose ir kitose sistemose). Tai suteikia lankstumo pritaikant modelį skirtingoms užduotims ir pramonės šakoms (OpenAI, 2024c). Be to, GPT-4 palaiko daugiamodalinį mokymąsi, kuris leidžia apdoroti tekstinius ir vizualinius duomenis vienu metu, praplečiant jo taikymo sritį. Modelis taip pat turi išplėstines saugumo funkcijas, kurios mažina galimybę piktnaudžiauti technologija, įskaitant griežtas generuojamo turinio kontrolės priemones ir nuolatinį mokymą siekiant sumažinti klaidų ar šališkumo tikimybę naudojant RLHF (Reinforcement Learning with Human Feedback) metodą. Kur RLHF metodas apima žmogaus pateiktą įžvalgų integravimą į modelio mokymo procesą, leidžiant modeliui geriau atitikti etinius standartus, generuoti patikimesnę informaciją ir užtikrinti, kad atsakymai būtų kontekstiškai tinkami bei nepateiktų žalingos ar klaidinančios informacijos. Ir ChatGPT-4 gali būti sukonfigūruotas taip, kad vartotojo pateikti duomenys nebūtų saugomi ar naudojami modelio tolimesniam mokymui (OpenAI, 2023b). Taip pat ChatGPT-4 turi nemokamą versiją su ribotomis galimybėmis arba „Plus“ versija (~20 USD/mėn.), kur siūlomas pilnai prieinamas GPT-4 modelis (OpenAI, 2024i).

3.2.2. Google Gemini 1.5 Flash

Gemini 1.5 Flash taip pat sukurta remiantis transformatorių architektūra, išplėsta multimodalinėmis galimybėmis. Ši architektūra remiasi standartiniu transformatoriaus dekodierio principu ir yra pritaikyta efektyviai valdyti ir integruoti skirtingus duomenų tipus, įskaitant tekstą, vaizdus, garsus ir video. Modelis kartu su transformatorių architektūra naudoja ir mišrių ekspertų (angl. „Mixture of Experts“ - MoE) strategiją, kad dinamiškai nukreiptų skaičiavimo resursus į specifines užduotis, taip pagerindamas našumą ir sumažindamas skaičiavimo sąnaudas. Tokia architektūra taip pat leidžia greitesnę ir efektyvesnę mokymą bei prisitaikymą prie naujų užduočių, o jo lankstumas užtikrina sklandų integravimą į įvairius programų sprendimus, pavyzdžiui, debesijos

paslaugas ar lokalias DI sistemas (Gemini Team, 2024). Be to, Gemini 1.5 Flash modelis palaiko didelį konteksto langą – iki 32 tūkst. žetonų (mokamoje versijoje iki 1 milijono), kuris leidžia apdoroti didelio masto duomenis ir ilgas sekas, išplečiant jo panaudojimo galimybes nuo sudėtingų mokslinių tekstų analizės iki išplėstų dialogų apdorojimo (Gemini, 2024).

Kalbos supratimo srityje Gemini 1.5 Flash pasižymi aukštu kalbos supratimo tikslumu, gebėdamas atlikti sudėtingas semantines ir sintaksines užduotis, kurios buvo įvertintos naudojant įvairius standartinius testus. Vienas iš pagrindinių vertinimo testų buvo MMLU testas, apimantis 57 skirtingas disciplinas, įskaitant matematiką, istoriją ir teisę. Remiantis 2024 m. rugsėjo mėn. duomenimis, Gemini 1.5 Flash šiame teste pasiekė 67,3% tikslumą. Modelio treniravimas apima didžiulius duomenų rinkinius, apimančius tekstinius, vaizdinius ir kitus duomenų tipus, todėl jis gali efektyviai spręsti įvairias užduotis. Pavyzdžiui, modelis pasižymi gebėjimu generuoti logiškai struktūruotus tekstus, atpažinti kontekstinius niuansus ir atlikti programavimo kodo generavimo bei redagavimo užduotis (Google DeepMind, 2024).

Gemini 1.5 Flash yra suderinamas su įvairiomis platformomis, ypač su „Google Cloud“, kur jis gali būti integruotas per „Vertex AI“ įrankių rinkinį. Tai leidžia kūrėjams lengvai pritaikyti modelį savo poreikiams – nuo internetinių paslaugų kūrimo iki analitikos įrankių vystymo. Be to, atvira API prieiga užtikrina greitą integravimą į esamas verslo sistemas. Jo lankstumas ir galimybė prisitaikyti prie specifinių poreikių daro jį itin naudingą įvairiose industrijose, įskaitant sveikatos priežiūrą, švietimą ir klientų aptarnavimą (Google DeepMind, 2024).

Saugumas taip pat yra vienas iš pagrindinių Gemini 1.5 Flash elementų. Modelis sukurtas laikantis aukštų duomenų privatumo standartų, įskaitant vartotojo duomenų anonimiškumo užtikrinimą ir apsaugą nuo neteisėto naudojimo (Google Cloud, 2024). Taip pat modelis pasižymi ekonomišku: jis sukurtas taip, kad būtų prieinamas ir mažesnių organizacijų poreikiams, ypač dėl mažesnių skaičiavimo sąnaudų ir lanksčių kainodaros modelių „Google Cloud“ platformoje. Tai leidžia vartotojams naudotis pažangiomis DI galimybėmis, neperžengiant biudžeto ribų. Šiuo metu Gemini 1.5 Flash siūlo nemokamą versiją prisijungus prie Google paskyros, o Gemini Advanced planas siūlo platesnes galimybes su Gemini įrankiu už 21.99 eur (Gemini, 2024).

3.2.3. IBM Watson AI

Watsonx.ai architektūra yra atvira, modulinė ir pagrįsta dekodierio principu, pritaikyta tiek didelių organizacijų, tiek mažesnių įmonių poreikiams (IBM, 2024a). Platforma naudoja mikropaslaugų architektūrą, leidžiančią efektyviai valdyti atskiras funkcines dalis, tokias kaip modelių kūrimas, mokymas, dislokavimas ir valdymas. Ši struktūra palaiko sklandžią integraciją su trečiųjų šalių įrankiais ir kitais IBM produktais, pvz., duomenų valdymo ar saugumo sprendimais. Architektūra taip pat optimizuota dirbti su hibridiniais debesijos modeliais, suteikiant galimybę vykdyti DI modelius tiek vietinėje infrastruktūroje, tiek viešajame arba privačiame debesyje. Naudojant atvirojo kodo technologijas ir API, platforma užtikrina lankstumą bei suderinamumą su

įvairiomis sistemomis ir technologijomis, tuo pačiu palaikydama aukštą saugumo lygį ir našumą (Nikhil Gopal, 2024).

Watsonx.ai modeliai yra sukurti remiantis pažangiomis tokenizacijos technologijomis, kur vienas žetonas apima apie 4 simbolius arba 0,75 žodžio. Ši technika leidžia efektyviai analizuoti ir apdoroti didelius tekstinius duomenis (IBM watsonx, 2024b). Watsonx.ai suteikia mėnesiui laiko 50 tūkst. žetonų (IBM watsonx, 2024c). Patys mokymo duomenys kruopščiai atrenkami, siekiant užtikrinti aukštą tikslumą ir sumažinti šališkumo tikimybę. Naudojamos didelės apimties įvairių sričių duomenų rinkmenos, leidžiančios modeliams prisitaikyti prie skirtingų pramonės sektorių poreikių ir situacijų. Kalbos supratimo tikslumas yra vertinamas remiantis tokiais testais kaip MMLU, kurie leidžia įvertinti modelio gebėjimą atlikti įvairias užduotis. Pavyzdžiui, Watsonx.ai Llama 3 70B modelis yra pasiekęs 82,0% tikslumą MMLU teste, o Llama 3 8B modelis – 68,4% (IBM, 2024a; IBM AI community, 2024).

Integracija yra lengvai palaikoma su IBM paslaugomis ir trečiųjų šalių sprendimais, pavyzdžiui, duomenų valdymo ir analitikos įrankiais. Sauga yra viena iš platformos stipriųjų pusių – IBM taiko griežtas saugumo priemones, tokias kaip duomenų šifravimas, prieigos kontrolė ir reguliavimo atitiktis. Be pagrindinių funkcijų, watsonx.ai siūlo įvairias papildomas galimybes, tokias kaip duomenų paruošimas, modelių mokymas, diegimas ir valdymas (IBM watsonx, 2024a). Watsonx.ai taip pat suteikia prieigą prie iš anksto apmokytų modelių bibliotekos, įskaitant IBM sukurtus ir atvirojo kodo modelius, kuriuos galima pritaikyti specifiniams verslo poreikiams. Be to, watsonx.ai integruojasi su kitomis IBM paslaugomis, tokiomis kaip watsonx.data ir watsonx.governance, užtikrinant visapusišką AI sprendimų kūrimo ir valdymo ekosistemą (IBM Watsonx, 2024). Watsonx.ai siūlo kelis kainodaros planus, įskaitant nemokamą bandomąją versiją su ribotomis galimybėmis ir mokamus planus, pritaikytus skirtingiems verslo poreikiams. Kainodara grindžiama naudojamų išteklių kiekiu, įskaitant skaičiavimo vienetus (angl. „Capacity unit hours“) ir žetonus (angl. „Resource units“), kurie matuoja modelių mokymo sąnaudas (IBM watsonx, 2024c).

3.2.4. Microsoft Copilot

Microsoft Copilot architektūra remiasi dideliais kalbos modeliais (angl. „Large Language Models“, LLM), integruotais su Microsoft 365 ekosistema ir Microsoft Graph. Ši architektūra leidžia Copilot realiu laiku naudotis vartotojo duomenimis, tokiais kaip dokumentai, el. laišakai ir kalendoriaus įrašai, užtikrinant suasmenintą patirtį. Apdorojimo pajėgumai priklauso nuo Azure OpenAI infrastruktūros, kuri užtikrina didesnę greitį ir našumą. Tačiau, kaip ir visiems DI sprendimams, tam tikros sudėtingesnės užklauskos gali reikalauti daugiau resursų, ypač didelių duomenų rinkinių apdorojimo metu (Marius Sandbu, 2024). Be standartinių integracijų, Copilot Studio leidžia vartotojams jungtis prie išorinių duomenų šaltinių ir kurti agentus specifinėms užduotims atlikti. Šis lankstumas leidžia organizacijoms pritaikyti DI pagal savo poreikius, pavyzdžiui, automatizuoti ataskaitų kūrimą ar duomenų analizę. Papildomos galimybės apima duomenų vizualizaciją, turinio generavimą ir užduočių supaprastinimą (Microsoft Copilot Studio, 2024).

Copilot siūlo platų funkcionalumą, įskaitant turinio generavimą, santraukų kūrimą ir duomenų analizę. Integracijos galimybės su Word, Excel, PowerPoint, Outlook ir Teams suteikia vartotojams galimybę dirbti vienoje platformoje, išlaikant vientisą duomenų srautą. Copilot Studio dar labiau praplečia šias galimybes, leidžiant kūrėjams pritaikyti agentus specifinėms verslo reikmėms, integruojant trečiųjų šalių duomenų šaltinius ir paslaugas. Toks lankstumas stiprina Copilot kaip nepamainomą įrankį, skirtą ne tik individualiam darbui, bet ir visos organizacijos procesų optimizavimui. Copilot naudoja pažangius kalbos modelius, kurie leidžia užtikrinti aukštą kalbos supratimo tikslumą. Jis yra optimizuotas darbui su natūralia kalba, todėl generuoja atsakymus, kurie atitinka kontekstą ir vartotojo poreikius. Vis dėlto, vartotojams yra rekomenduojama patikrinti sugeneruotą turinį, kad būtų užtikrintas jo tikslumas ir atitiktis konkrečiai užduočiai (Microsoft 365 Copilot, 2024b). Microsoft Copilot efektyvumas grindžiamas žetonais – teksto apdorojimo vienetais. Kur kiekvienas Copilot užklauso įvesties žetonas sunaudoja 400 pajėgumo vienetų (angl. „Capacity Units“, CU) sekundžių, o išvesties žetonas – 1 200 CU sekundžių. Tikslus limitas žetonų oficialiose Microsoft svetainėse nenurodytas (Microsoft, 2024b).

Copilot laikosi aukščiausių saugumo standartų, integruotų į Microsoft 365 infrastruktūrą. Vartotojo duomenys nėra naudojami modelių mokymui, o visi procesai atitinka tokius reglamentus kaip: Bendrasis duomenų apsaugos reglamentas (GDPR). Be to, Copilot siūlo funkcijas, kurios užtikrina, kad vartotojo duomenys būtų saugomi ir naudojami tik tikslingai bei autorizuotai. Šios saugumo priemonės yra ypač svarbios organizacijoms, dirbančioms su jautria informacija, užtikrinant pasitikėjimą sistema (Microsoft 365 Copilot, 2024a). Copilot yra siūlomas kaip papildoma Microsoft 365 prenumeratos paslauga, kurios kaina priklauso nuo pasirinkto plano ir organizacijos poreikių ("Microsoft Copilot Pro" planas kainuoja \$20.00 vartotojui per mėnesį). Kainodara leidžia Copilot naudoti tiek mažoms komandoms, tiek didelėms korporacijoms, siekiančioms padidinti efektyvumą ir sumažinti darbo krūvį. Taip pat yra siūloma nemokama versija su ribotomis galimybėmis (Microsoft 365, 2024).

3.2.5. Generatyvinių DI įrankių palyginimas

Generatyviniai dirbtinio intelekto įrankiai vis sparčiau integruojami į įvairias sritis, siekiant automatizuoti procesus, didinti darbo efektyvumą ir pritaikyti inovatyvius sprendimus kasdienėse užduotyse. Ankstesniuose poskyriuose buvo išanalizuoti generatyviniai DI įrankiai, įvertinant jų architektūrą, mokymo duomenų apimtį, funkcionalumus ir kitas savybes. Siekiant sistemingai apibendrinti atliktą analizę, šiame poskyryje pateikiamas šių generatyvinių DI įrankių – „ChatGPT-4“, „Google Gemini“, „IBM Watson AI“ ir „Microsoft Copilot“ – palyginimas tarpusavyje (3 lentelė), kur būtų išskirtos pagrindinės įrankių savybės, pagal kurias toliau būtų apsibrėžta, kuris įrankis būtų tinkamiausias tolimesniam darbui, kuriant kibernetinio saugumo rizikos scenarijų generatorių finansų įstaigoms.

3 lentelė. Generatyvinių DI įrankių palyginamas

Kriterijus	ChatGPT-4	Google Gemini 1.5 Flash	IBM Watson AI (Watsonx.ai)	Microsoft Copilot
Architektūra	Pagrįstas transformatorių architektūra, dekoderio principu	Pagrįstas transformatorių architektūra, dekoderio principu	Modulinė sistema su galimybe pritaikyti modelius specifiniams poreikiams. Pagrįsta dekoderio principu.	Remiasi dideliais kalbos modeliais (LLM), integruotais su Microsoft 365 ekosistema ir Microsoft Graph
Žetonų dydis	iki 128 tūkst. žetonų	iki 32 tūkst. žetonų	50 tūkst. žetonų/mėn.	Tikslus dydis žetonų oficialiose Microsoft svetainėse nenurodytas
Mokymo duomenų apimtis	240 TB tekstinių duomenų	„Google“ oficialiai neskelbia tikslios „Gemini“ modelio mokymo duomenų apimtys. Tačiau „Google“ pabrėžia, kad modelis buvo apmokytas naudojant didelius ir įvairius duomenų rinkinius.	Tiksli mokymo duomenų apimtis nėra viešai atskleista, tačiau IBM pabrėžia, kad naudoja didelės apimtys duomenų rinkinius	„Microsoft“ oficialiai neskelbia tikslios „Copilot“ modelio mokymo duomenų apimtys.
Kalbos supratimo tikslumas	virš 86% tikslumo NLP užduotyse (MMLU testas)	67,3% tikslumas pasiektas teste MMLU	Pasižymi aukštu kalbos supratimo tikslumu, leidžiančiu efektyviai atlikti užduotis.	Naudoja pažangius kalbos modelius, kurie užtikrina aukštą kalbos supratimo tikslumą.
Funkcijos	Multi-modalinis, gebantis suprasti ir generuoti tekstą, vaizdus, kodą	Multi-modalinis, realaus laiko užduotys	Skirtas duomenų paruošimui, modelių mokymui, diegimui ir valdymui.	Tekstų analizė, duomenų sujungimas, dokumentų kūrimas ir valdymas realiu laiku.
Integracija	API, platus platformų palaikymas	API, Google Cloud ekosistema	Galima integruoti per API ar specifinius IBM įrankius	Integracija su įvairiomis Microsoft 365 programomis

3 lentelės tęsinys

Kriterijus	ChatGPT-4	Google Gemini 1.5 Flash	IBM Watson AI (Watsonx.ai)	Microsoft Copilot
Saugumas	Modelis palaiko serverio valdomą duomenų saugą, konfidencialumą, yra pritaikytos generuojamo turinio kontrolės priemonės, nuolatinis mokymas siekiant sumažinti klaidų ir šališkumo tikimybę	Aukšti duomenų privatumo standartai, įskaitant vartotojo duomenų anonimiškumo užtikrinimą ir apsaugą nuo neteisėto naudojimo	Duomenų šifravimas, prieigos kontrolė, reguliavimo atitiktys	Vartotojo duomenys nėra naudojami modelių mokymui, o visi procesai atitinka tokius reglamentus kaip: Bendrasis duomenų apsaugos reglamentas (GDPR)
Kaina	Nemokama versija su ribotomis galimybėmis arba „Plus“ versija (~20 USD/mėn.)	Nemokama versija su ribotomis galimybėmis arba „Advanced“ planas (21.99 EUR/mėn.)	Nemokama bandomoji versija su ribotomis galimybėmis ir mokami planai, pagal pritaikytus skirtingus verslo poreikius	Nemokama versija su ribotomis galimybėmis arba Microsoft 365 prenumeratos paslauga, kurios kaina priklauso nuo pasirinkto plano.

Šaltinis: sudaryta autorės

Palyginus ChatGPT-4, Google Gemini 1.5 Flash, IBM Watson AI (Watsonx.ai) ir Microsoft Copilot (3 lentelė), galima teigti, kad ChatGPT-4 yra tinkamiausias įrankis kuriant kibernetinio saugumo rizikos scenarijų generatorių finansų įstaigoms. Vienas iš pagrindinių pranašumų, kad ChatGPT-4 suteikia itin platų konteksto langą (iki 128 tūkst. žetonų), kuriame gali būti efektyviai apdoroti didelio masto ir ilgų sekų tekstiniai duomenys. Tai ypač svarbu kibernetinio saugumo srityje, kur reikia išlaikyti sudėtingų scenarijų kontekstą, analizuoti ilgas technines ataskaitas ir istorinius duomenis apie grėsmes. Be to, „ChatGPT-4“ gali pasiūlyti patogias integracijos galimybes su esamomis finansų įstaigų sistemomis per API, o multimodalinės funkcijos gali suteikti galimybę apdoroti ne tik tekstinę informaciją, bet ir kitas duomenų formas, pvz., sistemų veikimo istoriją ar kodą, o tai užtikrina įrankio universalumą.

Palyginus su Google Gemini 1.5 Flash, kuris taip pat palaiko multimodalinę funkcionalumą, ChatGPT-4 yra pranašesnis kalbos supratimo tikslumu (86 % MMLU testuose, kai 67,3 % Gemini), kas yra būtina finansų sektoriaus grėsmių prevencijai ir valdymui. Taip pat Gemini galimybės apdoroti sudėtingus ilgus tekstinius fragmentus yra ribotos, lyginant su ChatGPT-4. IBM Watson AI išsiskiria pažangiais saugumo ir duomenų apsaugos sprendimais, tačiau jo modulinė sistema

reikalauja didesnių sąnaudų ir sudėtingesnės integracijos, kas gali apsunkinti efektyvų naudojimą mažiau lanksčiose situacijose. Microsoft Copilot, savo ruožtu, yra labiau pritaikytas darbui Microsoft 365 ekosistemoje, tačiau jo funkcionalumas kibernetinio saugumo kontekste yra ribotas, ypač analizuojant sudėtingus scenarijus ar apdorojant didelės apimties duomenis.

Apibendrinant galima teigti, jog ChatGPT-4 yra optimalus pasirinkimas finansų įstaigų kibernetinio saugumo scenarijų generavimo srityje. Jo aukštas kalbos supratimo tikslumas, ilgo konteksto apdorojimo galimybės ir paprastas integravimas leidžia efektyviai spręsti kompleksines užduotis. Be to, konkurencinga kainodara ir multimodalinės funkcijos užtikrina, kad šis įrankis ne tik gali atitikti finansų sektoriaus poreikius, bet ir suteikti galimybę jį lanksčiai pritaikyti ateities uždaviniais. Sekančiame poskyryje bus aprašomas užklauso konteksto formavimas finansų įstaigų rizikos scenarijams generuoti.

3.3. Užklauso konteksto formavimas rizikos scenarijams generuoti

Kuriant kibernetinio saugumo rizikos scenarijų generatorių finansų įstaigoms, būtina atsižvelgti į specifinį šių organizacijų kontekstą, griežtus reguliavimo reikalavimus ir rizikos valdymo praktikas, kur užklauso konteksto formavimas yra esminė rizikos valdymo proceso dalis, leidžianti tinkamai identifikuoti, analizuoti ir valdyti galimas grėsmes. Remiantis tokiais standartais kaip ISO 31000:2018 ir ISO/IEC 23894:2023, rizikos valdymo kontekstas apibrėžiamas kaip sistemingas procesas, kuris apima tiek vidinius, tiek išorinius veiksnius (ISO 31000, 2018; ISO/IEC 23894, 2023). Toliau, šiame poskyryje bus analizuojama, kaip organizacijos vidinis ir išorinis kontekstas formuoja pagrindą rizikos scenarijams generuoti.

Rizikos scenarijų generavimas prasideda nuo tinkamai suformuoto užklauso konteksto. Jis apima organizacijos vidinių ir išorinių aplinkybių analizę, kurios pagrindu formuojamos užklauso dirbtinio intelekto modeliams. Tinkamai parengtas kontekstas leidžia (ISO 31000, 2018; ISO/IEC 23894, 2023):

- Identifikuoti kritinius organizacijos kibernetinio saugumo aspektus.
- Užtikrinti generuojamų scenarijų atitiktį realioms grėsmėms ir rizikos valdymo tikslams.
- Optimizuoti DI sistemų veikimą, pateikiant tikslingus duomenis.

Konteksto formavimas yra svarbus procesas, kurio metu yra pateikiama išsami informacija apie organizacijos vidinius ir išorinius aspektus. Konteksto formavimas finansų sektoriuje yra nuoseklus procesas, skirtas nustatyti organizacijos vidinius ir išorinius veiksnius, kurie turi įtakos kibernetinių grėsmių dinamikai. Remiantis ISO 31000 gairėmis, procesas apima tris pagrindinius etapus: vidinio konteksto analizę, išorinio konteksto analizę ir rizikos prioritetų nustatymą. Vidinio konteksto analizę pradedama nuo informacijos apie organizacijos struktūrą ir procesus rinkimo. Tai apima IT infrastruktūros apžvalgą, kibernetinio saugumo biudžeto dydį, esamas rizikos valdymo strategijas ir darbuotojų kompetencijas. (ISO 31000, 2018).

Išorinio konteksto analizė apima reguliavimo aplinkos, rinkos sąlygų ir išorinių grėsmių vertinimą. Finansų sektoriuje svarbu atsižvelgti į reguliacinius reikalavimus, tokius kaip DORA, GDPR, bei į išorines grėsmes, pavyzdžiui, sukčiavimo schemas ar geopolitinius veiksnius, kurie gali paveikti finansų įstaigas. Ir rizikos prioritetų nustatymas yra paskutinis konteksto formavimo etapas. Kur gavus konteksto rezultatus, identifikuojamos pagrindinės grėsmės, kurios turi būti įtrauktos į užklausas. (ISO 31000, 2018; ISO/IEC 23894, 2023).

Pagal standartus ISO 31000:2018 ir ISO/IEC 23894:2023 buvo sudarytas konteksto klausimynas iš penkių pagrindinių temų (šis klausimynas toliau bus naudojamas kibernetinio saugumo rizikos scenarijų generatoriaus kūrime, formuojant kontekstą užklausoms), kurios apima finansų įstaigų vidinį ir išorinį kontekstą, valdymo sistemas, grėsmių identifikavimą ir reakcijos strategijas (6 priedas):

1. **Organizacinis kontekstas:** šioje dalyje pateikiami klausimai, susiję su vadovybės įsipareigojimais, biudžeto paskirstymu, atsakomybių pasiskirstymu ir kultūriniais veiksniais. Pavyzdžiui, klausimas „Kaip jūsų organizacija nustato kritines kibernetinio saugumo rizikas, turinčias įtakos finansinėms operacijoms, ir nustato jų prioritetus?“ Kur kritinių rizikų nustatymas padeda užtikrinti, kad dėmesys būtų sutelkiamas į pagrindines grėsmes finansinių operacijų, pvz., sukčiavimą ir reguliavimo pažeidimus.
2. **Išorinis kontekstas:** nagrinėja reguliacinius ir rinkos veiksnius, tokius kaip GDPR ar PSD2 laikymasis. Pavyzdžiui, klausimas „Kaip jūsų organizacija stebi atsirandančias išorines grėsmes, būdingas finansų sistemoms?“ padeda suprasti išorines grėsmes, kaip: sukčiavimą ar nacionalinės valstybės kibernetines atakas, kurios yra svarbios finansinių organizacijų rizikai apsibrėžti.
3. **Valdymas ir strategijos:** nagrinėja, kaip organizacija integruoja kibernetinio saugumo valdymą į bendrą rizikos valdymo sistemą. Klausimai, tokie kaip: „Kaip dažnai peržiūrimos ir atnaujinamos jūsų kibernetinio saugumo politikos?“, leidžia įvertinti finansų įstaigos prisitaikymą prie besikeičiančių grėsmių.
4. **Grėsmių identifikavimas ir scenarijų kūrimas:** nagrinėjami kokie metodai naudojami grėsmių identifikavimui ir scenarijų validavimui. Pavyzdžiui, klausimas „Kokie metodai naudojami siekiant nustatyti kibernetinio saugumo grėsmes finansų sistemose?“ padeda užtikrinti veiksmingą pasirengimą prieš finansų sektoriui būdingą riziką.
5. **Incidentų valdymas ir atkūrimas:** apima klausimus apie organizacijos reagavimo į incidentus gebėjimus ir atkūrimo strategijas. Pavyzdžiui, klausimas „Ar išbandėte reagavimo į incidentus planus pagal realius finansinius kibernetinius scenarijus?“ leidžia nustatyti pasirengimą ekstremalioms situacijoms.

Apibendrinus, užklausos konteksto formavimas yra sudėtingas, tačiau esminis procesas, siekiant sukurti efektyvias rizikos scenarijų generavimo sistemas. Pritaikant ISO standartus ir išsamiai analizuojant organizacijos kontekstą, galima užtikrinti, kad generuojami scenarijai bus praktiškai naudingi ir atitiks finansų sektoriaus specifinius poreikius. Atsižvelgiant į ISO 31000:2018 ir ISO/IEC 23894:2023 standartų gaires, užklausos konteksto formavimas leidžia sistemingai identifikuoti rizikos šaltinius, vertinti galimas grėsmes ir kurti tikslingus rizikos scenarijus. Struktūruotas požiūris į konteksto analizę ne tik užtikrina efektyvių kibernetinio saugumo sprendimų

kūrimą, bet ir padeda finansų įstaigoms prisitaikyti prie kintančių technologinių ir reguliavimo aplinkybių. Sekančiame poskyryje bus pateikiamas užklausos standartas dirbtinio intelekto rizikos scenarijams generuoti.

3.4. Užklausos standartas dirbtinio intelekto rizikos scenarijams generuoti

Kibernetinio saugumo rizikos scenarijų kūrimas yra viena iš esminių kibernetinio saugumo rizikos valdymo veiklų, padedanti nustatyti galimus grėsmių įvykius ir jų poveikį organizacijos turtui. Dirbtinio intelekto įrankiai gali automatizuoti šį procesą, generuodami galimus rizikos scenarijus, kurie padėtų kibernetinio saugumo specialistams priimti informuotus sprendimus (Zacharis et al., 2024). Toliau šiame poskyryje bus nagrinėjamas užklausos inžinerijos (angl. prompt engineering) konceptas, taip pat pateikiami užklausos standartai, pagal kuriuos būtų kuriami kibernetinio saugumo rizikos scenarijai, kurie padėtų finansų įstaigoms sistemingai vertinti galimas grėsmes, užtikrinti atitiktį reguliaciniams reikalavimams ir stiprinti bendrą kibernetinį atsparumą.

Užklausų inžinerija yra disciplinuotas užklausų kūrimo procesas, kuriuo siekiama sukurti optimalias užklausas dirbtinio intelekto modeliams, siekiant išgauti tikslius ir naudotojui naudingus atsakymus. Tai yra esminis metodas, leidžiantis valdyti ir orientuoti DI modelių atsakymus pagal specifinius organizacijos poreikius, ypač finansų sektoriuje, kur kibernetinio saugumo rizikos yra itin reikšmingos. Tinkamai pritaikytas užklausų konstravimas užtikrina, kad modelis ne tik suprastų keliamą užduotį, bet ir pateiktų atsakymus, kurie atitinka reguliavimo reikalavimus ir organizacijos veiklos tikslus. Pagrindinis užklausų inžinerijos privalumas, yra tas, kad yra galimybė kontroliuoti modelio interpretaciją ir sumažinti galimų klaidų ar netikslumų riziką (IBM, 2024b; OECD, 2021b).

Finansų sektoriuje, kuriame kibernetinis saugumas yra kritinis veiksnys, užklausų inžinerija gali padėti suformuoti tikslingas užklausas, kurios gali būti orientuotos į tokius klausimus kaip sukčiavimo prevencija, duomenų apsauga ar atitiktis reguliavimo reikalavimams. Pavyzdžiui, vietoje bendrinės užklausos „Nurodykite finansų sektoriaus grėsmes,“ galima pateikti konkrečią ir struktūruotą užklausą: „Kaip gali būti išnaudojami duomenų saugumo trūkumai mokėjimo sistemose, ir kokias prevencijos priemones reikėtų taikyti pagal PCI DSS standartus?“ Tokia užklausa suteikia DI modeliui aiškesnį kontekstą, leidžiantį sugeneruoti naudingus scenarijus, ir kurie gali būti tiesiogiai pritaikomi praktikoje (ENISA, 2023a; IBM, 2024b). Efektyvių užklausų kūrimas yra pagrindinis užklausų inžinerijos elementas. Finansų sektoriuje užklausų formulavimas turi būti sistemingai suplanuotas ir apimti šiuos standartus (ISO 31000, 2018; ISO/IEC 23894, 2023; OpenAI, 2024f):

- **Tikslus klausimų formulavimas:** užklausos formuluotė turi būti aiški ir lengvai suprantama tiek žmogui, tiek DI modeliui. Užklausos turi būti aiškiai apibrėžtos ir atspindėti organizacijos poreikius. Pavyzdžiui, vietoje „Nurodykite galimas rizikas“ būtų efektyviau klausti „Kaip duomenų vagystė gali paveikti mažmeninės bankininkystės klientų pasitikėjimą pagal PCI DSS standartus?“

- **Aiškiai apibrėžtas kontekstas:** užklausa turėtų prasidėti nuo organizacijos specifikos ar situacijos aprašymo. Pavyzdžiui: „Organizacija veikia mažmeninės bankininkystės sektoriuje, naudojant PCI DSS suderinamas mokėjimo sistemas.“ Konteksto pateikimas padeda DI modeliui geriau suprasti problemą ir pritaikyti atsakymus.
- **Struktūruotas kontekstas:** užklausa turi įtraukti vidinę ir išorinę informaciją, pvz., įmonės IT infrastruktūros detales ir galimus išorinius veiksmus, tokius kaip naujos reguliavimo taisyklės.
- **Modelio supratimas:** kiekviena užklausa turi būti optimizuota pagal tai, kaip modelis interpretuoja pateiktą informaciją, siekiant užtikrinti tikslų atsakymą. Finansų srityje, kur rizikos labai įvairios, būtina užtikrinti, kad modelis tiksliai suprastų problemą. Pavyzdžiui, užklausa gali būti detalesnė: „Nurodykite, kaip sukčiavimo scenarijai gali būti susiję su neteisingu duomenų valdymu pagal GDPR reikalavimus“.
- **Aiškumas:** užklausa formuluoje turi būti vienareikšmė. Reiktų vengti bendrinių klausimų, tokių kaip: „Kokios yra grėsmės?“, ir naudoti aiškius terminus, pvz., „Kokie techniniai veiksniai padidina riziką kibernetiniam sukčiavimui?“
- **Tikslumas rezultatų reikalavimams:** užklausoje turi būti nurodyti reikalavimai rezultatams, pvz., „Sugeneruokite scenarijų, apimančią rizikos identifikavimą, jos poveikį ir rekomenduojamus sprendimus pagal ISO 31000.“
- **Tikslų apibrėžimas:** užklausoje turi būti aiškiai nurodyta, koks atsakymas reikalingas: ar tai būtų rizikos scenarijų rinkinys, rekomendacijos, ar duomenų analizė.

Sistemiškai generuojant kibernetinio saugumo rizikos scenarijus finansų įstaigoms, būtina atsižvelgti į tinkamai suformuluotas užklausas. Užklausų formos turi būti orientuotos į konkretų tikslą, tokį kaip atakos tipo identifikavimas, tikėtino poveikio aprašymas ar grėsmių valdymo strategijų pasiūlymai. Užklausa turi būti struktūrizuota taip, kad DI galėtų tiksliai identifikuoti potencialias grėsmes ir analizuoti galimus rizikos scenarijus (ISO 31000, 2018; ISO/IEC 23894, 2023). Toliau bus pateikiamos užklausų formos, sukurtos remiantis apsibrėžtais standartais. Šios užklausa bus orientuotos į finansų įstaigų kibernetinio saugumo iššūkius, pritaikytus jų specifiniams poreikiams ir grėsmių valdymo tikslams.

1. Užklausa forma grėsmių šaltinių nustatymui (angl. Actors)

Šios užklausa forma skirta identifikuoti grėsmių šaltinius (vidinius ir išorinius) bei jų motyvus, aprašant juos struktūrizuotai, remiantis tarptautinėmis rizikos valdymo gairėmis (9 pav.).

[Instrukcija] Atsakykite į pateiktus klausimus, suskirstydami informaciją pagal nurodytas kategorijas:

1. Organizacijos apibūdinimas (finansų įstaigos rūšis, įmonės dydis).
2. Išvardinkite galimus vidinius ir išorinius grėsmių šaltinius, apibūdindami jų pagrindines savybes.
3. Paaiškinkite, kokie yra jų tikėtini motyvai (pvz., finansinė nauda, sabotžas, kerštas).
4. Nurodykite, kaip šie šaltiniai gali paveikti jūsų organizacijos turtą ar operacijas.

[Pateikimo formatas] Kiekvieną atsakymą struktūrizuokite į kategorijas:

- Įmonė
- Grėsmės šaltinis (vidinis arba išorinis)
- Tikėtini veiksmai (aprašykite konkrečius veiksmus)
- Motyvai (nurodykite tikėtinus tikslus)

Šaltinis: sudaryta autorės

9 pav. Užklauso form grėsmių šaltinių nustatymui (angl. Actors)

2. Užklauso form grėsmių įvykių nustatymui (angl. Threat Events)

Šios užklauso form skirta sukurti grėsmių scenarijus, kurie atitiktų tarptautines gerąsias praktikas (10 pav.).

[Instrukcija] Sukurkite galimus kibernetinių grėsmių įvykius:

1. Organizacijos apibūdinimas (finansų įstaigos rūšis, įmonės dydis).
2. Aprašykite kiekvieną grėsmės įvykį, nurodydami šiuos elementus:
 - Potencialūs grėsmės veiksmai (pvz., duomenų vagystė, DDoS ataka).
 - Pažeidžiamumo tipai, kuriais gali pasinaudoti grėsmė.
 - Tikėtinos pasekmės organizacijai.
3. Priskirkite kiekvienam įvykiui kategoriją (pvz., technologinė, organizacinė, žmogiškoji).

[Pateikimo formatas] Naudokite sąrašą, kuriame aiškiai būtų išskiriami kiekvieno įvykio elementai.

Šaltinis: sudaryta autorės

10 pav. Užklauso form grėsmių įvykių nustatymui (angl. Threat Events)

3. Užklauso form poveikio analizei (angl. Impact)

Šios užklausos forma skirta įvertinti, kaip grėsmės gali paveikti organizacijos tikslus ir veiklą, pateikiant išmatuojamą poveikį (11 pav.).

[Instrukcija] Įvertinkite galimų grėsmių poveikį:

1. Organizacijos apibūdinimas (finansų įstaigos rūšis, įmonės dydis).
2. Kokios organizacijos veiklos sritys būtų paveiktos (pvz., finansai, reputacija, operacijos)?
3. Kaip šis poveikis gali būti išmatuotas (pvz., finansiniai nuostoliai, veiklos pertraukimai)?
4. Nurodykite, kokių veiksmų reikėtų imtis siekiant sumažinti šį poveikį.

[Pateikimo formatas] Kiekvieną poveikio vertinimą pateikite kaip:

- Įmonė
- Poveikio sritis
- Pasekmių tipas
- Galimos priemonės rizikai sumažinti

Šaltinis: sudaryta autorės

11 pav. Užklausos forma poveikio analizei (angl. Impact)

4. Užklausos forma motyvų analizei (angl. Intentions/Motives)

Šios užklausos forma skirta identifikuoti galimus grėsmių veikėjų tikslus ir juos apibūdinti (12 pav.).

[Instrukcija] Apibūdinkite galimus grėsmių veikėjų tikslus:

1. Organizacijos apibūdinimas (finansų įstaigos rūšis, įmonės dydis).
2. Kokie tikslai galėtų paskatinti šiuos veiksmus (pvz., ekonominė nauda, politinis spaudimas)?
3. Kaip šie tikslai gali paveikti jūsų organizaciją?
4. Kokie prevencijos metodai galėtų sumažinti šių motyvų įgyvendinimo tikimybę?

[Pateikimo formatas] Informaciją pateikite kaip:

- Įmonė
- Tikslas
- Potencialus poveikis
- Prevencijos metodas

Šaltinis: sudaryta autorės

12 pav. Užklausos forma motyvų analizei (angl. Intentions/Motives)

5. Užklauso forma turto analizei (angl. Assets)

Šios užklauso forma skirta nustatyti ir įvertinti pažeidžiamiausius organizacijos išteklius (13 pav.).

[Instrukcija] Nurodykite organizacijos svarbiausius turtus ir jų pažeidžiamumą:

1. Organizacijos apibūdinimas (finansų įstaigos rūšis, įmonės dydis).
2. Kokie organizacijos turtai yra labiausiai pažeidžiami?
3. Kaip jie gali būti paveikti konkrečių grėsmių scenarijų?
4. Kokius apsaugos metodus taikote?

[Pateikimo formatas] Kiekvieną atsakymą pateikite struktūrizuotai:

- Įmonė
- Turtas
- Pažeidžiamumo tipas
- Apsaugos metodai

Šaltinis: sudaryta autorės

13 pav. Užklauso forma turto analizei (angl. Assets)

6. Užklauso forma laiko aspektams (angl. Time)

Šios užklauso forma skirta nustatyti grėsmių pasireiškimo laikotarpį ar trukmę (14 pav.).

[Instrukcija] Įvertinkite grėsmių atsiradimo ir poveikio laikotarpį:

1. Organizacijos apibūdinimas (finansų įstaigos rūšis, įmonės dydis).
2. Kada labiausiai tikėtina, kad grėsmė gali pasireikšti?
3. Kiek laiko gali užtrukti grėsmės poveikis organizacijai?
4. Kaip galite efektyviai reaguoti per šį laikotarpį?

[Pateikimo formatas] Informaciją pateikite aiškiai:

- Įmonė
- Laiko tarpas (tikėtina pradžia ir trukmė)
- Tikėtinos pasekmės
- Reagavimo priemonės

Šaltinis: sudaryta autorės

14 pav. Užklauso forma laiko aspektams (angl. Time)

7. Bendra užklauso form

Šios užklauso forma skirta integruoti visus kriterijus į vieną nuoseklų procesą (15 pav.).

[Instrukcija] Atlikite visapusišką rizikos analizę, apimančią šiuos elementus:

1. Organizacijos apibūdinimas (finansų įstaigos rūšis, įmonės dydis).
2. Grėsmių šaltinių identifikavimas.
3. Galimų grėsmių įvykių sudarymas ir poveikio vertinimas.
4. Pažeidžiamiausių turtų identifikavimas ir apsaugos priemonių nustatymas.
5. Laiko aspektų įvertinimas ir reakcijos planavimas.

[Pateikimo formatas] Kiekvieną dalį pateikite aiškiai ir struktūrizuotai, laikydamiesi šių sekcijų:

- Įmonė
- Grėsmės
- Grėsmių įvykiai
- Poveikis

Šaltinis: sudaryta autorės

15 pav. Bendra užklauso forma

Visos šios užklauso formos buvo pagrįstos remiantis ENISA publikacijomis (ENISA, 2020, 2022), kurios buvo naudojamos kaip teorinė pagrindimo bazė, siekiant užtikrinti mokslinį užklauso struktūros pagrįstumą. ISO standartais, kur struktūra remiasi ISO 31000:2018 ir ISO/IEC 23894 gairėmis, kad būtų užtikrintas nuoseklumas (ISO 31000, 2018; ISO/IEC 23894, 2023). Ir OpenAI dokumentacija (OpenAI, 2024g), kurioje pateikiami standartai kaip turėtų būti suformuotos užklauso, kad jos būtų aiškios, konkrečios ir leistų gauti tikslius atsakymus (sugeneruotus scenarijus).

Atsakymų validacija padeda užtikrinti, kad dirbtinio intelekto modelio sugeneruoti rezultatai atitiktų užklauso kontekstą, pateikiamą pačią užklausą, finansų sektoriaus reikalavimus ir praktinius poreikius. Finansų sektoriuje, kur rizikos valdymas yra griežtai reglamentuotas ir priklauso nuo tikslų duomenų, atsakymų kokybės patikrinimas tampa būtinu procesu. Vienas iš pagrindinių atsakymų validavimo etapų yra ekspertų vertinimas. Kur sugeneruoti scenarijai pateikiami kibernetinio saugumo specialistams, kurie įvertina jų tikslumą, naudingumą ir atitiktį organizacijos poreikiams. Pavyzdžiui, jei DI modelis sugeneruoja scenarijų, kaip apsaugoti klientų asmens duomenis nuo neteisėtos prieigos, ekspertai gali peržiūrėti, ar jame pateikiamos rekomendacijos atitinka GDPR 33 straipsnį (pranešimai apie duomenų pažeidimus) ir 35 straipsnį (poveikio duomenų apsaugai vertinimas). Šis procesas padeda užtikrinti, kad atsakymai būtų teisiškai suderinti ir praktiškai taikytini (DAIR.AI, 2024; OpenAI, 2024g).

Simuliacijos yra kitas svarbus validavimo etapas, leidžiantis praktiškai išbandyti sugeneruotus scenarijus realiose ar modeliuotose situacijose. Finansų sektoriuje simuliacijos gali būti naudojamos

testuojant organizacijos pasirengimą į kibernetines grėsmes. Pavyzdžiui, sugeneruotas scenarijus apie mokėjimo sukčiavimo atvejį gali būti išbandytas simuliuojant realų incidentą, kad būtų įvertintas organizacijos incidentų valdymo planas. Tokios simuliacijos gali padėti nustatyti, ar DI pateiktos rekomendacijos yra praktiškai įgyvendinamos ir ar jos efektyviai mažina grėsmes. Atitikties analizės validavimas užtikrina, kad atsakymai atitiktų ne tik organizacijos poreikius, bet ir tarptautinius standartus bei reguliacinius reikalavimus. Tai ypač svarbu finansų sektoriuje, kur kibernetinis saugumas glaudžiai susijęs su tokiais reglamentais kaip DORA, GDPR ar PCI DSS. Dirbtinio intelekto modelio sugeneruoti atsakymai turi būti kruopščiai peržiūrimi, siekiant užtikrinti, kad jie būtų suderinti su užklausoje pateiktais reguliaciniais standartais. Pavyzdžiui, jei modelis pateikia rekomendaciją apie duomenų apsaugą, būtina patikrinti, ar ji visiškai atitinka GDPR reikalavimus, susijusius su asmens duomenų saugojimu ir pranešimų apie pažeidimus tvarka (DAIR.AI, 2024; OpenAI, 2024g).

Validacija taip pat apima modelio veikimo analizę, siekiant įvertinti, ar dirbtinio intelekto modelis tinkamai interpretavo užklausą ir ar jis teisingai pritaikė kontekstą. Remiantis užklauso inžinerijos dokumentacijomis, validuojant atsakymus svarbu įvertinti šiuos aspektus (DAIR.AI, 2024; OpenAI, 2024g):

- Aiškumas: ar sugeneruotas atsakymas yra aiškus ir suprantamas vartotojui.
- Konteksto atitikimas: ar gauti rezultatai tiesiogiai susiję su užklauso kontekstu.
- Nuoseklumas: ar atsakymas yra logiškai suderintas ir struktūruotas.

Apibendrinant, galima teigti, kad aiškiai struktūruotos užklauso, apibrėžti reikalavimai ir nuosekli atsakymų validacija užtikrina, kad dirbtinio intelekto sprendimai gali padėti efektyviai valdyti kibernetines grėsmes, kylančias finansų sektoriuje. Užklauso formos ir tinkamai suformuluotas užklauso kontekstas yra esminiai elementai, užtikrinantys, kad dirbtinio intelekto modeliai finansų sektoriuje generuotų kokybiškus ir praktikoje pritaikomus scenarijus. Atsižvelgiant į tarptautinius standartus, tokius kaip ISO 31000 ir ISO/IEC 23894, užklauso turi būti aiškiai struktūruotos ir orientuotos į organizacijos poreikius. Toks požiūris ne tik didina finansų įstaigų efektyvumą kibernetinių grėsmių valdyme, bet ir skatina DI naudojimą atsakingai ir skaidriai. Taigi, suformuotos septynios užklauso formos yra svarbios, kaip metodinis instrumentas, kuris ne tik gali supaprastinti grėsmių identifikavimą ir klasifikavimą, bet ir leisti kurti išsamius rizikos scenarijus, būtinus kibernetinio saugumo strategijų tobulinimui. Sekančiame skyriuje bus aprašomas sukurtas kibernetinio saugumo rizikos scenarijų generatorius finansų įstaigoms.

4. DI SISTEMOS RIZIKOS SCENARIJŲ GENERATORIUS

Šiame skyriuje bus aprašytas generatoriaus ([nuorodą į generatorių](#)) kūrimo ir apmokymo procesas, vartotojo sąsaja, saugumo ir privatumo sprendimai.

4.1. Generatoriaus kūrimas

Generatoriaus kūrimo procesas buvo sudarytas iš kelių svarbių etapų, kurie leido sukurti specializuotą įrankį finansų įstaigų kibernetinio saugumo rizikos scenarijų generavimui. Šie etapai apima duomenų surinkimą, modelio apmokymą, testavimą, validaciją bei optimizaciją, siekiant užtikrinti tikslumą ir pritaikomumą realioms situacijoms. Pats generatorius buvo įgyvendintas naudojant „ChatGPT Editor“ įrankį. Toliau, šiame poskyryje bus aprašomas generatoriaus kūrimo procesas.

Pagrindinis generatoriaus tikslas buvo suteikti finansų įstaigoms įrankį, leidžiantį generuoti, analizuoti ir valdyti jų naudojamų DI sistemų kibernetinių grėsmių scenarijus. Šis tikslas buvo apibrėžtas remiantis šiais kriterijais:

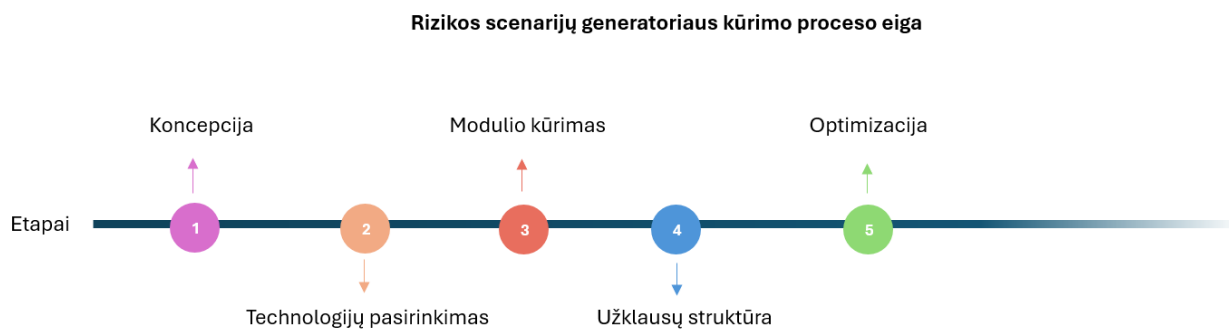
- Rizikų valdymo svarba: finansų įstaigos yra viena iš labiausiai pažeidžiamiausių sektorių dėl darbo su jautria informacija ir priklausomybės nuo skaitmeninių sistemų.
- Reguliavimo poreikiai: integruoti tokius reglamentus kaip DORA, GDPR ir kitus finansų sektoriaus reguliavimo dokumentus.
- Lankstumas: suteikti galimybę individualizuoti scenarijus pagal organizacijos dydį, sektorių ir pasirengimo lygį.

Generatorius buvo sukurtas siekiant generuoti finansų įstaigoms realistiškus kibernetinių rizikų scenarijus, kuris leistų vartotojams suformuluoti užklausas, pagrįstas įvairiomis kibernetinėmis grėsmėmis, pažeidžiamumais ar specifinėmis atakomis, pavyzdžiui, DDoS atakomis ar duomenų nutekėjimu. Pagrindinis koncepcijos tikslas buvo suteikti vartotojams galimybę lengvai kurti scenarijus be būtinybės turėti galias technines žinias apie kibernetinį saugumą. Pačio generatoriaus vartotojo sąsaja suteikia greitą ir paprastą prieigą prie generavimo proceso, leisdamą vartotojui pačiam suvesti poreikį užklausos forma ir pateikti generatoriui, kuris sugeneruoja pagal užklausą galimus rizikos scenarijus, arba atsakyti į klausimą, pagal kurį generatoriui yra pateikiamas įmonės kontekstas ir pagal gautus atsakymus yra sugeneruojami scenarijai. Pats kūrimo procesas apėmė įvairių finansinių įstaigų kibernetinių grėsmių duomenų analizę, naudojamą scenarijų generavimui. Generuojant scenarijus, sistema naudoja finansų įstaigų reguliacinės aplinkos ir rizikos valdymo duomenų bazę (DORA, ENISA, MITRE, ISO standartai, NIST direktyvos ir kitas publikacijas),

kurioje yra apibrėžiama finansų įstaigų reguliacinė aplinka ir kibernetinių atakų metodai, bei jų potencialios pasekmės, kurios yra susietos su realiais saugumo incidentais, pavyzdžiui, pažeidžiamumais informacinėse sistemose.

Šio generatoriaus kūrimui buvo naudojamas „ChatGPT Editor“, kuris leidžia modeliuoti realias kibernetines grėsmes ir kurti rizikos scenarijus. „ChatGPT Editor“ suteikė pagrindines galimybes integruoti scenarijų kūrimo logiką į GPT-4 modelį, leidžiant vartotojams atsakyti į klausimą ar formuoti tiesiogiai užklausas, o generatoriui – atsakyti sukurtais scenarijais. Technologinis pagrindas remiasi dirbtinio intelekto ir natūralios kalbos apdorojimo metodais, kurie leidžia apdoroti ir generuoti prasmingus bei realius kibernetinių atakų scenarijus. GPT-4 modelis naudojamas siekiant suprasti vartotojo atsakymą (į klausimą pateiktą atsakymą arba pateiktą tiesioginę užklausą), identifikuoti esminius saugumo elementus ir pateikti realistiškus atsakymus. Pavyzdžiui, kai vartotojas užduoda klausimą apie DDoS ataką, modelis analizuoja grėsmes, remdamasis istoriniais duomenimis, ir pateikia tinkamus scenarijus, apibūdinančius galimas pažeidžiamas vietas (turtą) bei pasekmes. Be to, buvo naudojami įvairūs standartai, direktyvos, reglamentai ir aktai kibernetinio saugumo ir finansų įstaigų reguliacinės aplinkos, skirti užtikrinti, kad atsakymai būtų tikslūs ir atitiktų vartotojų lūkesčius dėl rizikos valdymo finansų įstaigose.

Kūrimo procesas buvo padalintas į kelis etapus (16 pav.). Pirmasis etapas buvo idėjos vystymas – siekis sukurti įrankį, kuris galėtų generuoti kibernetinių rizikų scenarijus finansų įstaigoms, remdamasis finansų įstaigų reguliacine aplinka, grėsmėmis, ketinimais, padariniais ir pažeidžiamumais. Vėlesniuose etapuose buvo nuspręsta, kokios technologijos ir duomenų bazės bus naudojamos, siekiant užtikrinti, kad scenarijai būtų tikslūs ir pagrįsti realiais duomenimis. Po to buvo sukurta pagrindinė modulio struktūra, kurioje užklausos paverčiamos scenarijais, atsižvelgiant į vartotojo pateiktą informaciją ir buvo parengta klausimyno forma, leidžianti vartotojams pateikti svarbiausią informaciją apie jų organizacijos rizikos valdymo situaciją. Užklausos struktūra, taip pat buvo optimizuota, kad vartotojai galėtų lengvai suformuluoti klausimus, susijusius su įvairiomis grėsmėmis ar atakų tipais. Galiausiai, buvo atlikta sistemos optimizacija, siekiant pagerinti našumą ir užtikrinti greitesnį užklausų apdorojimą.

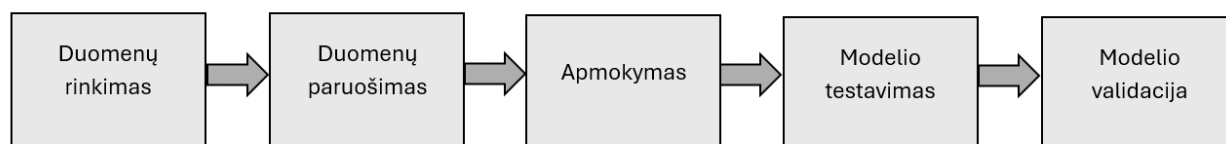


Šaltinis: sudaryta autorės

16 pav. Rizikos scenarijų generatoriaus kūrimo proceso eiga

Siekiant ChatGPT-4 modelį pritaikyti kibernetinių rizikų scenarijų generavimui, buvo atliktas specializuotas apmokymo procesas, apimantis kelis svarbius etapus (17 pav.):

- **Duomenų rinkimo ir paruošimo:** DI apmokymo procesas prasidėjo duomenų surinkimu, orientuotu į finansų įstaigų kibernetinio saugumo sritį. Šie duomenys apėmė finansų įstaigų reguliacinę aplinką ir didelį grėsmių tipų bei pažeidžiamumų duomenų rinkinį iš įvairių šaltinių, įskaitant viešąsias finansų įstaigų ir kibernetinio saugumo duomenų bazes bei mokslinius straipsnius.
- **Duomenų žymėjimo:** surinkti duomenys buvo suskirstyti pagal tam tikras kibernetinio saugumo kategorijas, tokias kaip: grėsmių tipai, atakos ar galimi pažeidžiamumai. Žymėjimo procesas leido modelio apmokymą nukreipti tikslinėmis kryptimis, siekiant, kad modelis galėtų greičiau atpažinti ir diferencijuoti skirtingus kibernetinių grėsmių aspektus. Šis žingsnis buvo svarbus, kad modelis galėtų identifikuoti ne tik pažeidžiamumus, bet ir galimus jų eksploatavimo scenarijus.
- **Modelio apmokymo:** kitas etapas buvo modelio treniravimas, kuris apėmė scenarijų struktūros supratimo apmokymą, užklausų interpretacijos tobulinimą, individualizuotų scenarijų generavimą ir taisyklių bei pavyzdinių scenarijų pateikimą generatoriui (žr. 4.2 poskyrį).
- **Vertinimo mechanizmų naudojimą:** po pirminio treniravimo generatorius buvo testuojamas naudojant įvairius vertinimo kriterijus (7 priedas), kad būtų patikrinta generuojamų scenarijų kokybė ir atitikimas realiam kibernetinio saugumo kontekstui.



Šaltinis: sudaryta autorės

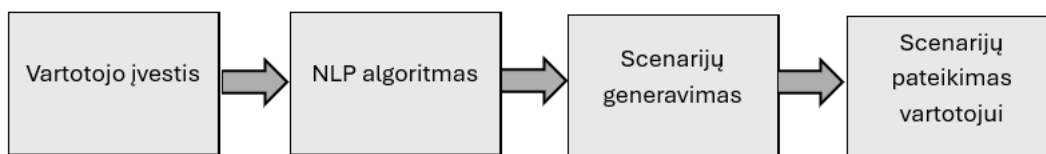
17 pav. ChatGPT apmokymo schema

Scenarijų generavimo logika taip pat buvo viena iš svarbesnių generatoriaus elementų, kuri leido generatoriui ne tik apdoroti vartotojo užklausą, bet ir sukurti struktūruotą ir logišką atsakymą, kuris atitiktų pateiktos užklausos kontekstą ir tikslą. Pirmasis šio proceso žingsnis prasideda nuo užklausos analizės, kur identifikuojamas kontekstas pateiktos užklausos. Toliau užklausa vertinama atsižvelgiant į jos specifiką – ar vartotojas klausia apie bendras kibernetines grėsmes, ar domisi konkrečiu scenarijumi, pavyzdžiui, DDoS atakomis, duomenų nutekėjimu ar pažeidžiamumų įvertinimu. Šiame etape generatorius taip pat naudoja natūralios kalbos apdorojimo metodus, kurie

padeda nustatyti užklausoje tematiką ir grėsmių tipą, siekiant užtikrinti tikslų tolesnį duomenų apdorojimą.

Kai užklausoje kontekstas yra identifikuotas, generatorius pereina prie antrojo scenarijų generavimo žingsnio, kuris apima kibernetinių grėsmių ir rizikos faktorių identifikavimą. Kibernetinių grėsmių identifikavimo procesas vykdomas atpažinti grėsmių tipus ir jų charakteristikas pagal pateiktą informaciją. Tai leidžia generatoriui įvertinti, kokie pažeidžiamumai ar rizikos veiksniai yra susiję su vartotojo užklausoje nurodytu scenarijumi. Pavyzdžiui, jei vartotojas pateikia užklausoje apie galimą sukčiavimo ataką (angl. „phishing“), sistema automatiškai identifikuoja šios grėsmės požymius ir galimus pažeidžiamumus, kuriuos ši grėsmė gali paveikti.

Sekančiame etape vyksta konkretaus scenarijaus struktūravimas. Šis etapas svarbus, nes generatorius ne tik surenka identifikuotus duomenis apie grėsmę ir pažeidžiamumus, bet ir struktūrizuoja atsakymą pagal kibernetinio saugumo scenarijų kūrimo komponentus. Kur kiekvienas scenarijus yra suskaidomas į esminius komponentus: grėsmės aprašymą, galimų pažeidžiamumų sąrašą, atakas ir švelninimo strategijas. Ši struktūra leidžia vartotojui aiškiai suprasti, kokia grėsmė aptikta, kokie gali būti jos padariniai ir kaip ji gali būti suvaldyta. Tokia struktūruota informacija yra ypač svarbi finansų organizacijoms, siekiančioms įvertinti riziką ir planuoti kibernetinio saugumo veiksmus. Ir paskutinis etapas yra galutinio atsakymo pateikimas vartotojui, kur atsižvelgiant į užklausoje sudėtingumą ir pateiktą kontekstą, generatorius pateikia struktūruotą, suprantamą ir patikimą atsakymą, kuriame aiškiai išdėstytos kibernetinės grėsmės, pažeidžiamumai ir apsaugos metodai. 18 pav. pateikta schema, su pagrindiniais generatoriaus komponentais ir kaip jie tarpusavyje sąveikauja.



Šaltinis: sudaryta autorės

18 pav. Pagrindiniai generatoriaus komponentai ir jų tarpusavio sąveika

Apibendrinant galima teigti, kad generatoriaus kūrimo procesas buvo kompleksinis ir orientuotas į praktinį pritaikomumą, kuris užtikrino sistemos patikimumą ir pritaikomumą finansų įstaigų kibernetinio saugumo srityje. Kiekvienas etapas nuo duomenų rinkimo iki rezultato pateikimo padėjo užtikrinti, kad generatorius tiksliai identifikuotų grėsmes, pateiktų struktūruotus atsakymus ir atitiktų finansų įstaigų poreikius bei standartus. Sukurtas generatorius gali padėti finansų įstaigoms sumažinti riziką, priimti informuotus sprendimus ir stiprinti organizacinį atsparumą prieš kibernetines atakas. Sekančiame poskyryje bus aprašomas generatoriaus apmokymo procesas.

4.2. Generatoriaus apmokymo procesas

Kibernetinių rizikų valdymo srityje itin svarbu kurti automatizuotus įrankius, galinčius tiksliai ir greitai generuoti realistiškus scenarijus, pagrįstus kibernetinėmis grėsmėmis ir pažeidžiamumais. Generatorius buvo sukurtas siekiant pateikti tikslius ir veiksmingus rizikos vertinimo scenarijus finansų įmonėms, pagrįstus naujausiomis saugumo grėsmių ir atakų taksonomijomis. Toliau šiame poskyryje bus aprašomas generatoriaus apmokymo procesas.

Kibernetinių rizikų scenarijų generatoriaus kūrimo procesas reikalavo ne tik technologinių sprendimų, bet ir specifinio apmokymo, siekiant užtikrinti, kad įrankis tinkamai suprastų vartotojų pateiktas užklaudas ir pateiktų kokybiškus scenarijus. Pagrindinė apmokymo dalis buvo vykdoma naudojant generatoriaus "Configure" -> "Instructions" lauką, kuriame buvo pateikiamos detalios instrukcijos, kaip generatorius turi elgtis reaguodamas į įvairias užklaudas (8 priedas). Apmokymo procesas buvo pagrįstas trimis pagrindinėmis kryptimis:

1. Scenarijų struktūros supratimo mokymu;
2. Užklausių interpretacijos tobulinimu;
3. Individualizuotų scenarijų generavimu;
4. Taisyklių ir pavyzdinių scenarijų pateikimu generatoriui.

Užklausių valdymas šiame generatoriuje buvo pagrįstas natūralios kalbos apdorojimo principais, kurie leidžia pačiam generatoriui suvokti užklaudas ir pateikti tinkamus atsakymus. Pavyzdžiui, jei vartotojas pateikia užklausa, susijusią su DDoS ataka, generatorius analizuoja užklaustos tekstą, identifikuoja pagrindines atakos charakteristikas, ir, remdamasi turimais duomenimis, sugeneruoja scenarijų. Pats generatorius per "Instructions" lauką buvo apmokinamas per kelis pagrindinius etapus. Pirmas etapas buvo scenarijaus šablono struktūrizavimas, kur kiekvieno scenarijaus pagrindą sudaro šeši pagrindiniai elementai:

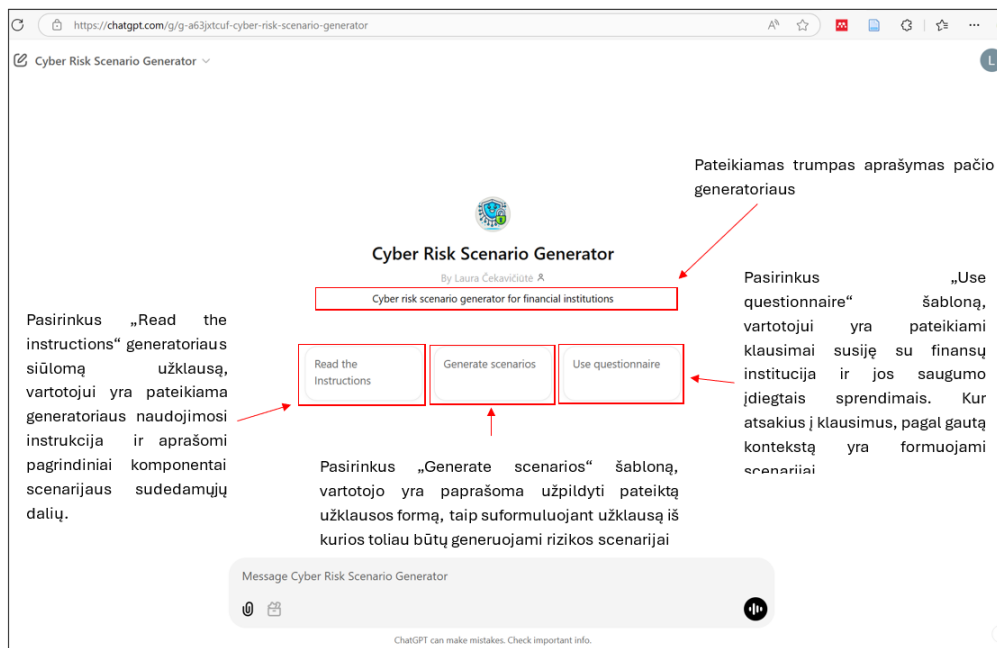
- **Veikėjas:** galimas kibernetinės grėsmės šaltinis. Tai gali būti žmogus ar sistema, pavyzdžiui, vidiniai darbuotojai, išoriniai įsilaužėliai arba kenksmingi DI agentai ar automatizuoti botai. Veikėjų identifikavimas leidžia tiksliai suprasti, kokios galimos kibernetinės grėsmės kyla iš vidinių ar išorinių šaltinių.
- **Grėsmių įvykiai:** identifikuojama kokia grėsmė gali kilti sistemai, remiantis grėsmių taksonomija. Tai gali būti duomenų vagystė, DDoS ataka ar kita kibernetinė ataka, kuri gali turėti kritinį poveikį finansų organizacijos informacinėms sistemoms.
- **Ketiniai/motyvai:** apibūdinamas grėsmės veikėjo motyvas. Nurodomi, kokie gali būti ketiniai, pavyzdžiui, finansinė nauda, politinis spaudimas, šnipinėjimas, kerštas ar informacijos kompromitavimas. Tai padeda tiksliau įvertinti grėsmes, atsižvelgiant į veikėjo tikslus.
- **Padariniai:** šiame etape pateikiami galimi padariniai, kuriuos grėsmė gali turėti sistemai ar turtui. Pateikiami galimi neigiami padariniai, tokie kaip duomenų nutekėjimas, reputacijos praradimas ar finansinė žala.

- **Turtas/ištekliai:** aprašoma, kokie turtai ar ištekliai yra pažeidžiami dėl grėsmės, pvz., finansiniai duomenys, klientų informacija ar intelektinė nuosavybė, kurie gali būti pažeisti atakos metu. Tai padeda nustatyti, kurioms sistemoms gali būti reikalingos papildomos apsaugos priemonės.
- **Laikas:** apibūdinamas grėsmių įvykių laikas arba laiko intervalas. Tai gali būti specifinis laiko intervalas ar laikotarpis, per kurį įvyksta grėsmė, pvz., per kelias minutes ar kelias savaites trunkanti ataka. Laikas padeda suprasti, per kiek laiko gali įvykti konkretus kibernetinis įvykis ir kokią poveikį tai gali turėti verslui.

Taigi, ši struktūrizuota scenarijaus forma suteikia galimybę išsamiai aprašyti kibernetinę grėsmę ir ją susieti su galimomis rizikomis finansų organizacijos infrastruktūrai. Sekantis apmokymo etapas buvo sukurti kelis naudojimo režimus, kurie buvo suskirstyti į kelis užklausų šablonus (19 pav.):

1. „Read the Instructions“: kur buvo orientuojamasi į vartotojų supažindinimą su generatoriaus galimybėmis ir veikimo logika. Šio užklausos šablono tikslas buvo pateikti aiškias ir nuoseklias instrukcijas, kaip vartotojas gali naudotis generatoriumi. Apmokymo metu šis šablonas buvo kuriamas taip, kad vartotojas galėtų gauti informaciją apie scenarijų struktūrą (pvz., veikėjas, grėsmių įvykiai, ketinimai, padariniai, turtas/ištekliai, laikas). Instrukcijose taip pat buvo pabrėžti pagrindiniai generatoriaus principai, tokie kaip vartotojo užklausų formavimas, tikslinės informacijos pateikimas ir užklausų tikslumas. Šis šablonas buvo naudojamas ne tik kaip instrukcijų pateikimas vartotojui paspaudus ant šablono „Read the Instructions“, bet ir generatoriaus apmokymui atpažinti, kokie elementai yra būtini kuriant pilnavertį scenarijų. Pavyzdžiui, buvo apibrėžti tokie elementai kaip veikėjas (angl. *Actor*), grėsmių įvykiai (angl. *Threat Events*), ketinimai (angl. *Intentions*), poveikis (angl. *Effect*), pažeidžiami turtai/ištekliai (angl. *Assets*) ir laikas (angl. *Time*). Generatorius buvo apmokytas interpretuoti šiuos elementus kaip užklausos komponentus ir pateikti struktūrizuotą atsakymą, užtikrinant, kad visi šie punktai būtų aptarti generuojamame scenarijuje. Šio režimo apmokymo metu buvo įtraukti ir pavyzdžiai, kaip kiekvienas elementas turėtų būti aprašomas, pvz., „Grėsmės įvykiai: DDoS ataka prieš pagrindinį įmonės tinklalapį.“
2. „Generate Scenarios“: šis užklausos šablonas buvo skirtas apmokyti generatorių suprasti vartotojų užklausas ir konvertuoti jas į individualizuotus kibernetinių grėsmių scenarijus. Šis šablonas buvo apmokomas taip, kad generatorius galėtų apdoroti įvairių sudėtingumo lygių užklausas. Apmokymo metu buvo naudojami konkretūs užklausų šablonai, kurie buvo apsirašyti metodologinėje dalyje. Modelis buvo mokomas atpažinti pagrindines užklausos dalis, kaip grėsmės tipą, veikėją, pažeidžiamumą ir poveikį. Be to, buvo įtraukti įvairūs hipotetiniai ir realūs grėsmių pavyzdžiai, kurie leido generatoriui mokytis adaptuoti scenarijus pagal vartotojo kontekstą. Kiekvienas sugeneruotas scenarijus buvo apmokintas remtis tokiomis dokumentacijomis kaip įvairūs reguliavimo reikalavimai (pvz., DORA, ISO standartai) ir pateikti praktines rekomendacijas rizikos mažinimui.
3. „Use Questionnaire“: režimas buvo skirtas detalizuotiems vartotojo atsakymams rinkti ir pagal juos generuoti scenarijus. Šio šablono apmokymo metu pagrindinis dėmesys buvo skiriamas integruotam klausimynui, kuriame pateikti klausimai apėmė finansų organizacijos valdymą, infrastruktūros pažeidžiamumus, reguliavimo reikalavimus ir incidentų valdymo

praktiką. Kiekvienas klausimas buvo susietas su tam tikra scenarijaus dalimi, pavyzdžiui, klausimai apie organizacijos IT infrastruktūrą leido sugeneruoti išsamius scenarijus apie galimus grėsmių veikėjus ir poveikį. Apmokant modelį, buvo išbandomos įvairios vartotojų atsakymų kombinacijos, kad būtų užtikrintas scenarijų pritaikomumas skirtingose situacijose. Taip pat buvo įtraukti aiškūs nurodymai, kad visi klausimai turi būti atsakyti nuosekliai, o neatsakytų klausimų atveju generatorius turėtų prašyti papildomos informacijos. Šis režimas užtikrino, kad sugeneruoti scenarijai būtų ne tik tikslūs, bet ir pritaikomi prie individualių organizacijos poreikių.



19 pav. Generatoriaus pagrindinis langas

Taip pat prie scenarijaus kriterijų papildomai generatorius buvo apmokintas generuoti švelninimo strategijas, kur būtų pateikiamos rekomendacijos, kaip būtų galimybė išvengti sugeneruoto scenarijaus (rekomendacijos formuluojamos remiantis standartais, direktyvomis, publikacijomis, tokiomis kaip: ENISA, ISO, NIST, pagal kurias pateikiami tinkamiausi sprendimo būdai apsisaugoti nuo pateikto rizikos scenarijaus). Ir galiausiai, generatoriaus kūrimo metu buvo išjungtos „Capabilities“ funkcijos: „Web Search“, „DALL-E Image Generation“ ir „Code Interpreter & Data Analysis“. Šis sprendimas buvo priimtas siekiant užtikrinti maksimalų generatoriaus efektyvumą ir saugumą, laikantis „ChatGPT“ dokumentacijos gairių (OpenAI, 2024a). Išjungus „Web Search“, buvo išvengta priklausomybės nuo išorinių duomenų šaltinių, užtikrinant, kad generatorius veiktų tik su vidiniais, iš anksto įkeltais dokumentais ir žinių baze, įskaitant ISO, NIST, DORA standartus bei finansų sektoriaus ataskaitas. Tai taip pat sumažino riziką, susijusią su duomenų nutekėjimu ar netinkamu informacijos pateikimu iš išorinių šaltinių. Išjungus „DALL-E Image Generation“, buvo išlaikytas generatoriaus funkcionalumo aiškumas, nes vaizdiniai komponentai nebuvo esminiai kibernetinių grėsmių scenarijų generavimo procese. Tuo tarpu atsisakymas naudoti „Code Interpreter & Data Analysis“ užtikrino, kad vartotojo pateikti duomenys būtų apdorojami tik naudojant tekstinius modelius, taip pagerinant užklausų apdorojimo tikslumą ir greitį.

Kuriant generatorių taip pat buvo susidurta su iššūkiais, vienas iš pagrindinių iššūkių buvo užtikrinti, kad sugeneruoti scenarijai būtų tiek realistiški, tiek naudingi vartotojams. Dirbtinio intelekto modelio treniravimas pareikalavo optimizacijos, siekiant sumažinti klaidų skaičių ir padidinti atsako greitį. Pradiniuose kūrimo etapuose buvo pastebėta, kad kai kurios užklausos nebuvo tinkamai suprantamos dėl natūralios kalbos apdorojimo niuansų. Šie iššūkiai buvo sprendžiami tobulinant duomenų bazę, pridėjus daugiau grėsmių ir pažeidžiamumų pavyzdžių, taip gerinant sistemos gebėjimą atpažinti ir apdoroti užklausas. Taip pat buvo integruotos taksonomijos, kaip grėsmių ir turto, kurios padėjo standartizuoti grėsmių ir atakų klasifikavimą bei sisteminimą. Tai suteikė aiškią struktūrą grėsmių analizei, o vartotojams buvo sudaryta galimybė pateikti nuoseklias užklausas, kurios būtų tinkamai apdorojamos.

Kitas reikšmingas iššūkis buvo natūralios kalbos apdorojimo (sutrump. NLP) sudėtingumas. Kadangi vartotojai turi galimybę pateikti užklausas naudodami laisvą formą (ne struktūrizuotą kalbą), generatorius turėjo mokėti tiksliai suprasti įvairias grėsmių aprašymo frazes ir grįžti su tinkamais rizikos scenarijais. NLP modeliai kartais netiksliai apdorojo įvestį, ypač esant sudėtingoms arba labai specifinėms užklausoms. Šio iššūkio sprendimui buvo pritaikytos papildomos mokymo duomenų bazės, kurios padėjo pagerinti NLP modelio gebėjimą atpažinti sudėtingas grėsmių sąsajas. Taip pat buvo įgyvendinti papildomi algoritmai, skirti tikslinti užklausas pagal įvesties specifiką. Taigi, generatoriaus tikslas buvo sukurti ne tik tikslų, bet ir realistišką scenarijų, kuris būtų pritaikomas realiame pasaulyje. Tačiau iššūkis buvo surasti balansą tarp scenarijų detalumo ir jų praktinio pritaikymo. Pradinės versijos scenarijai kartais buvo per abstraktūs, o tai ribojė jų naudingumą praktikoje. Sprendžiant šią problemą, generatorius buvo papildomai apmokintas prisitaikyti pagal finansų sektorius ir specifinius verslo poreikius. Tai leido vartotojams generuoti ne tik universalius scenarijus, bet ir pritaikytus konkrečioms finansų sritims, pavyzdžiui, bankams, draudimo įmonėms ar elektroninių pinigų įstaigoms.

Taip pat buvo susidurta su simbolių limitu „Instructions“ lauke ribojimu iki 8000 simbolių, dėl ko teko optimizuoti pateikiamas instrukcijas. Kiekvienas simbolis buvo kruopščiai įvertintas, o perteklinė informacija pašalinta arba supaprastinta, kad būtų pateikti tik svarbiausi apmokymo elementai. Kitaip tariant, vietoj išsamių instrukcijų turinio buvo integruoti tik esminiai punktai. Šis ribojimas privertė sudėlioti informaciją taip, kad ji būtų glausta, bet kartu užtikrintų, jog generatorius išlaikytų aukštą atsakymų kokybę ir tikslumą. Be to buvo limituotas ir generatoriaus užklausių naudojimas vartotojams. Kur sukurtam generatoriui leido įvesti tik tam tikrą užklausių skaičių per nustatytą laiką, po kurio sistema automatiškai nustatydavo apribojimą, neleidžiant naudoti įrankio tol, kol nepraeis kelių valandų pauzė. Šis apribojimas kilo dėl technologinių resursų optimizavimo ir modelio apkrovos valdymo. Nors šie ribojimai padėjo išlaikyti sistemos stabilumą, jie taip pat ribojė vartotojų galimybę naudoti generatorių intensyviai per trumpą laikotarpį. Tam tikslui vartotojams buvo sukurtas naudojimosi instrukcijų šablonas, kuriame pateikiamos gairės kaip reiktų naudotis generatoriumi ir kurti užklausas struktūruotas.

Apibendrinant, generatoriaus apmokymo procesas buvo sudėtingas ir daugialypis, apimantis tiek techninių sprendimų įgyvendinimą, tiek vartotojų poreikių analizę. Šio įrankio tikslas buvo sukurti generatorių, kuris generuotų realistiškus kibernetinių rizikų scenarijus finansų įstaigoms. Generatoriaus apmokymas per „Configure“ -> „Instructions“ lauką buvo pagrindinis kūrimo etapas,

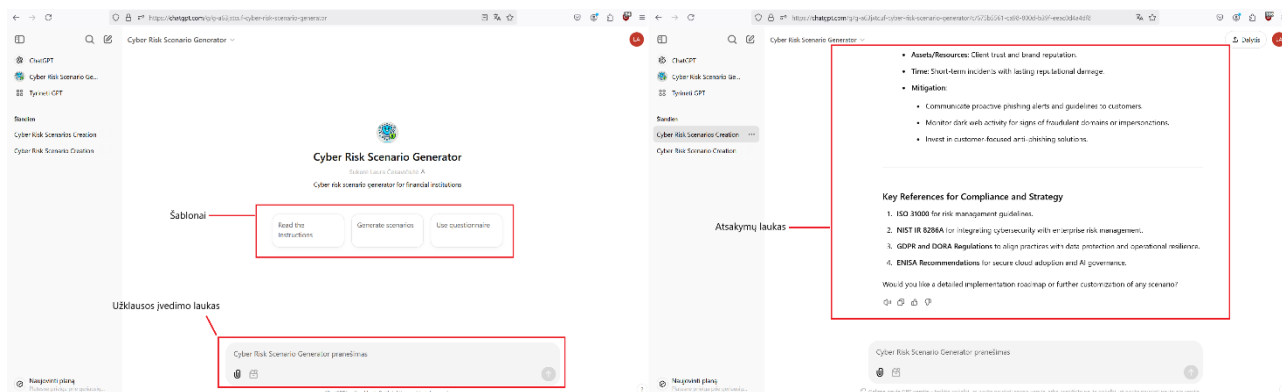
kuris leido sistemingai apmokyti generatorių suprasti vartotojų užklausas, generuoti individualizuotus scenarijus ir pritaikyti juos pagal finansų įstaigų poreikius. Naudojant iteracinį metodą ir pavyzdinius scenarijus, generatorius tapo ne tik tikslus, bet ir efektyvus, užtikrinantis vartotojų poreikių patenkinimą ir atitiktį kibernetinio saugumo standartams. Nors apmokymo metu buvo susidurta su iššūkiais, tokiais kaip sudėtingų kibernetinių grėsmių modeliavimas, natūralios kalbos apdorojimo tobulinimas ar tam tikrų limitų apribojimas, šie iššūkiai buvo įveikti tobulinant generatorių ir pateikiant vis daugiau informacijos ir duomenų, kaip taksonomijos grėsmių ir turto, įvairios direktyvos, standartai ir kita. Sekančiame poskyryje bus aprašoma generatoriaus vartotojo sąsaja.

4.3. Generatoriaus vartotojo sąsaja

Vartotojo sąsaja yra svarbi kiekvienos sistemos dalis, nes ji apibrėžia, kaip vartotojas sąveikauja su sistema. Rizikos scenarijų generatoriaus vartotojo sąsaja, ChatGPT platformoje, pasižymi paprastumu ir efektyvumu. Kadangi ji nebuvo kuriama nuo pagrindų, o pritaikyta ChatGPT teikiamame sąsajos rėme, ji išsiskiria minimalizmu ir patogumu, kas leidžia vartotojui lengvai ir greitai pateikti užklausas bei gauti atsakymus. Šiame poskyryje detaliai nagrinėjama vartotojo sąsaja, jos elementai, veikimo principai ir tai, kaip sąsaja prisideda prie rizikos scenarijų generavimo proceso.

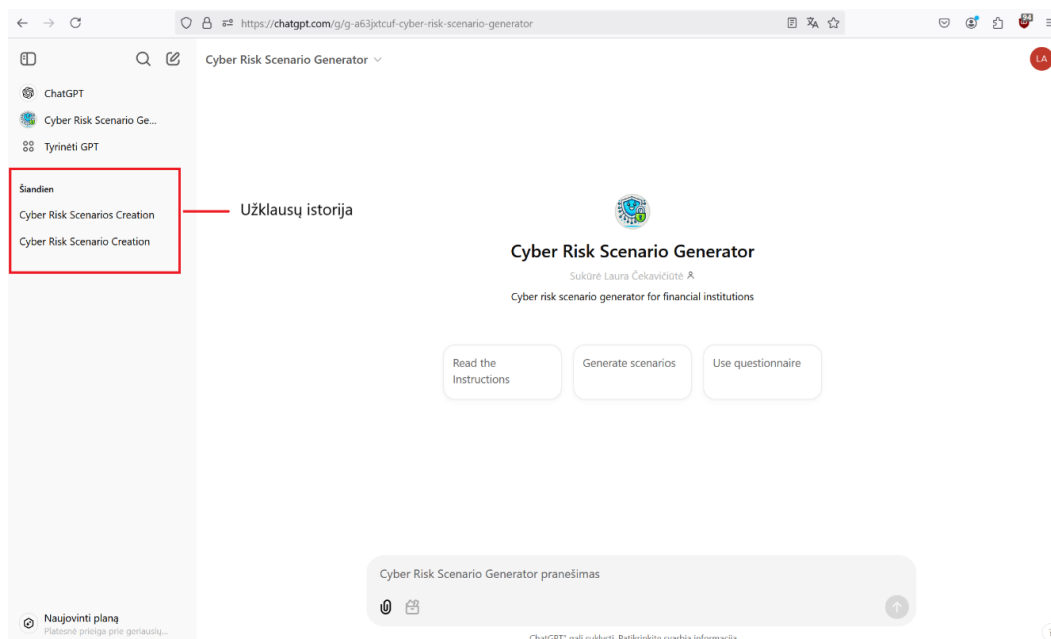
Rizikos scenarijų generatoriaus vartotojo sąsaja yra paremta ChatGPT platformos struktūra, kuri užtikrina intuityvų ir patogų naudojimą. Vartotojo sąsaja susideda iš kelių pagrindinių elementų, kurie sudaro paprastą ir aiškią sąveikos sistemą. Vartotojas turi prieigą prie tekstinio įvedimo lauko, kur gali pateikti savo užklausas, ir atsakymų lauką, kuriame rodomi gražinti rezultatai (20 pav.). Sąsaja pasižymi minimaliu dizainu, kuris orientuojasi į funkcionalumą ir neapkrauna vartotojo vizualiniais trikdžiais. Tokiu būdu ChatGPT aplinkoje naudojama sąsaja leidžia vartotojui koncentruotis tik į pagrindines užduotis ir nesiblaškyti.

Vartotojo sąsaja taip pat pasižymi iš anksto numatytų instrukcijų, scenarijų kūrimo ir klausimyno šablonais, kurie yra integruoti į ChatGPT aplinką (20 pav.). Šie šablonai leidžia greitai pradėti darbą ir pateikia kontekstus, pagal kuriuos toliau yra formuojamos užklausos ir iš jų rizikos scenarijai. Toks šablonų naudojimas užtikrina, kad net ir pradedantys vartotojai lengvai įsisavintų, kaip formuoti užklausas bei naudotis sistema, taip pagerindami naudotojo patirtį.



20 pav. Generatoriaus vartotojo sąsaja.

Be to, sąsaja sukurta taip, kad vartotojas galėtų lengvai naviguoti tarp skirtingų užklausių ir grįžti prie anksčiau pateiktų scenarijų. Kairėje sąsajos dalyje yra ankstesnių užklausių istorija (21 pav.), leidžianti patogiai peržiūrėti arba redaguoti jau sukurtus scenarijus. Tai ypač naudinga kuriant kompleksinius rizikos scenarijus, kurie gali reikalausti kelių skirtingų žingsnių arba papildomų paaiškinimų. Galimybė grįžti prie ankstesnių atsakymų ir scenarijų suteikia vartotojui daugiau kontrolės ir lankstumo dirbant su sistema. Sąsajos minimalumas, taip pat leidžia vartotojui greitai pradėti naudotis generatoriumi be papildomų mokymų. Kadangi sąsaja yra intuityvi, nereikia jokių papildomų instrukcijų ar sudėtingų nustatymų, kad būtų galima pradėti pateikti užklausas. Vartotojas tiesiogiai rašo savo užklausą tekstiname įvedimo lauke, o sistema grąžina išsamius, suformuotus atsakymus pagal pateiktą informaciją. Tokia paprasta ir aiški sąsaja leidžia generatoriumi naudotis efektyviai ir greitai, nes vartotojas nedelsdamas gali pradėti darbą ir gauti reikiamus rezultatus.



21 pav. Generatoriaus sukurtų užklausų istorija

Pati vizualinė dalis yra itin minimali: vienintelis papildomas grafikos elementas – logotipas, sukurtas specialiai šiam generatoriui. Logotipas yra vienintelis vizualinis identifikatorius, atspindintis sistemos specifiką ir išskiriantis ją iš kitų ChatGPT aplinkos sąsajų. Logotipas matomas viršuje, šalia pavadinimo, todėl jis nesukuria vizualinės apkrovos, o padeda lengviau atpažinti generatorių. Tokio dizaino pasirinkimas buvo strateginis sprendimas, siekiant išlaikyti švarų ir funkcionalų vartotojo sąsajos dizainą, kuris atitiktų ChatGPT platformos estetiką.

Apibendrinant, vartotojo sąsaja yra patogi rizikos scenarijų generavimo užduotims atlikti ChatGPT platformoje. Minimali, tačiau funkcionali sąsaja leidžia vartotojams greitai ir paprastai pateikti užklausas, gauti atsakymus bei pritaikyti iš anksto nustatytus šablonus. Sąsaja yra lengvai suprantama ir nereikalauja papildomų mokymų, o logotipas suteikia sistemai atpažįstamumo bei perteklinės vizualinės informacijos. Šis dizainas užtikrina efektyvią ir intuityvią vartotojo patirtį, leidžiančią vartotojams sklandžiai naudotis rizikos scenarijų generatoriumi ir susikonscentruoti ties pagrindine jo funkcija – kibernetinių rizikos scenarijų kūrimu. Sekančiame poskyryje bus aprašoma generatoriaus saugumo ir privatumo apsauga.

4.4. Generatoriaus saugumo ir privatumo apsauga

Šiuolaikinėje skaitmeninėje erdvėje kuriant ir naudojant dirbtinio intelekto sistemas, saugumo ir privatumo apsaugos klausimai tampa itin aktualūs. Kibernetinio rizikos scenarijų generavimo įrankis finansų įstaigoms, sukurtas naudojant ChatGPT Editor įrankį, taip pat susiduria su įvairiais iššūkiais, susijusiais su informacijos saugumu ir privatumu. Atsižvelgiant į tai, kad sukurtas generatorius naudoja ChatGPT platformos infrastruktūrą ir jos taikomas privatumo bei saugumo metodus, itin svarbu išanalizuoti, kokios technologinės ir organizacinės priemonės taikomos duomenų apsaugai užtikrinti. Šiame poskyryje aptariami saugumo ir privatumo saugos aspektai, taip pat vartotojo privatumą apsaugantys procesai ir kokios priemonės taikomos siekiant išvengti informacijos nutekėjimo ir apsaugoti generuojamus scenarijus nuo galimų kibernetinių grėsmių.

Vienas svarbiausių aspektų duomenų saugumo srityje yra šifravimo metodai, kurie naudojami apsaugoti tiek įvesties, tiek išvesties informaciją tarp vartotojo ir ChatGPT sistemos. Remiantis OpenAI saugumo gairėmis, duomenų šifravimas atliekamas tiek perdavimo, tiek saugojimo metu, siekiant užtikrinti, kad jautri informacija būtų apsaugota nuo neteisėtos prieigos. Taip pat OpenAI riboja prieigą prie duomenų, siekiant apsaugoti vidinius ir išorinius duomenis, o neautorizuotiems vartotojams prieiga yra draudžiama. Šie metodai yra ypatingai svarbūs, nes kibernetinių rizikos scenarijų generavimo įrankyje gali būti naudojami konfidencialūs duomenys, susiję su įmonės kibernetine sauga ar potencialiomis grėsmėmis (OpenAI, 2024e).

Kuriant ir naudojant dirbtinio intelekto sistemas, privatumo užtikrinimas tampa taip pat svarbus, nes vartotojų įvesti duomenys gali būti jautrūs. Šiuo atžvilgiu, OpenAI privatumo politika numato, kad vartotojų duomenys nėra naudojami sistemos mokymui, nebent vartotojas duoda aiškų sutikimą. Tai užtikrina, kad vartotojų pateikta informacija scenarijų generavimo įrankiui būtų naudojama tik tam konkrečiam tikslui ir nebūtų toliau panaudota be jų sutikimo. Siekiant dar labiau užtikrinti privatumo apsaugą, OpenAI taip pat taiko duomenų anonimizavimo metodus, kurie leidžia deidentifikuoti pateiktą informaciją, taip išsaugant asmenų privatumą. Ši priemonė užtikrina, kad duomenys, kurie gali būti jautrūs ir atskleisti privatų turinį, būtų apsaugoti nuo neteisėto panaudojimo ar analizuojami be vartotojo sutikimo (OpenAI, 2023a).

Prieigos kontrolė ir autentifikacija yra būtina siekiant užtikrinti saugų ir patikimą vartotojų naudojimąsi dirbtinio intelekto sistema. OpenAI įgyvendina prieigos kontrolės sistemą, kuri leidžia reguliuoti, kas ir kokiomis teisėmis gali naudotis rizikos scenarijų generatoriumi, taip pat kitomis ChatGPT funkcijomis (OpenAI, 2024f). Autentifikacijos sistema reikalauja vartotojų identifikacijos prieš suteikiant prieigą prie įrankio, o autorizacija nustato vartotojo teises ir prieigos lygius. Taip pat, patį generatorių galima pasiekti tik turint nuorodą, viešai generatorius nėra pasiekiamas. Be to, OpenAI stengiasi užtikrinti, kad autentifikacija ir autorizacija būtų vykdoma atsižvelgiant į naujausius saugumo standartus ir technologijas, siekiant sumažinti kibernetinių grėsmių poveikį (OpenAI, 2024d).

ChatGPT atitiktis teisiniams reikalavimams, ypač Europos Sąjungoje galiojančiam Bendrojo duomenų apsaugos reglamentui (GDPR), yra taip pat itin svarbi duomenų apsaugos ir vartotojų privatumo užtikrinimo dalis. OpenAI laikosi GDPR reikalavimų, kurie reglamentuoja asmens duomenų tvarkymą ir suteikia vartotojams teisę valdyti savo asmeninę informaciją (OpenAI, 2024b). Pavyzdžiui, vartotojai gali prašyti peržiūrėti arba ištrinti savo asmens duomenis, taip užtikrinant, kad jų informacija bus tvarkoma pagal teisinius reikalavimus. Be to, atitiktis reikalavimams apima asmens duomenų minimalizavimą ir skaidrumą, kurių laikantis ribojamas rinktų duomenų kiekis tik tam tikslui, kuriam vartotojai sutiko juos pateikti (OpenAI, n.d.-a).

Galimų kibernetinių grėsmių, tokių kaip DDoS atakos ar neteisėta prieiga prie duomenų, prevencija ir atsako mechanizmai yra svarbi saugumo dalis. OpenAI yra sukūrusi grėsmių stebėjimo ir reagavimo sistemas, kurios realiu laiku analizuoja ChatGPT sistemos veiklą, stebėdamos galimus kenkėjiškus veiksmus (OpenAI, 2023c). Be to, siekiant neutralizuoti potencialias grėsmes, incidentų valdymo komanda vykdo nuolatinį stebėjimą ir greitą reagavimą į grėsmes. Jei yra aptinkamas neteisėtas prisijungimas arba įtartina veikla, sistema iš karto blokuoja galimai kenkėjišką veiksmą (OpenAI, n.d.-b). Tokie atsako mechanizmai ne tik užtikrina saugumą, bet ir leidžia palaikyti nuolatinę sistemos veikimą be didelių trukdžių.

Apibendrinant, galima teigti, kad kibernetinio rizikos scenarijų generatorius taiko išsamias saugumo ir privatumo apsaugos priemones, įskaitant duomenų šifravimą, anonimizavimą, prieigos kontrolę, griežtus autentifikacijos reikalavimus ir atitiktį GDPR reikalavimams. Šie metodai užtikrina, kad vartotojų pateikta informacija būtų apsaugota ir tvarkoma saugiai bei atsakingai. Taip pat nuolatinė grėsmių stebėsena ir reagavimo priemonės leidžia sumažinti kibernetinių atakų riziką bei suteikia vartotojams didesnę pasitikėjimą naudojantis šia dirbtinio intelekto sistema ir joje sukurtais įrankiais. Sekančiame skyriuje bus atliekamas eksperimentas su sukurtu rizikos scenarijų generatoriumi.

5. EKSPERIMENTAS SU SUKURTU GENERATORIUMI

Šiame skyriuje bus pateikiamos keturios skirtingų finansų įstaigų rinkų hipotetinės įmonės su kuriomis toliau bus atliekami eksperimentai naudojantis kibernetinio saugumo rizikos scenarijų generatoriumi. Kad įsitikinti sukurto generatoriaus veiksmingumu, papildomai bus atlikti dar du eksperimentai su pasirinktomis hipotetinėmis finansų įmonėmis, pateikiant įmonių kontekstus su užklausomis kibernetinio saugumo ekspertams.

5.1. Hipotetinės įmonės

Eksperimento metu buvo pasirinktos keturios skirtingos finansų sektoriaus hipotetinės įmonės. Šios įmonės reprezentuoja svarbiausias finansų rinkas, remiantis sisteminiu požiūriu į kibernetinio saugumo riziką. Buvo pasirinktos tokios finansų įstaigos, kuriose atsispindi didžiausias dirbtinio intelekto naudojimo mastas. Tokiu būdu buvo užtikrinta, kad eksperimentas apimtų aktualiausias ir plačiausiai atstovaujamas kibernetinio saugumo problemas. Hipotetinės finansų įmonės buvo pasirinktos pagal šiuos kriterijus:

1. DI technologijų naudojimo mastas veiklos srityse – įmonės, kurios plačiai diegia DI sistemas rizikos vertinimui, sukčiavimo prevencijai, rinkos analizei ir kitoms veikloms (McKinsey, 2023).
2. Pažeidžiamumas kibernetinėms grėsmėms – pasirinktos įmonės, kurių IT infrastruktūra, duomenų valdymo ypatumai ir paslaugų pobūdis sukuria didesnę kibernetinio saugumo rizikos potencialą (World Economic Forum, 2023).
3. Įmonių skaičius ir reikšmė rinkoje – buvo atsižvelgta į tai, kurios finansų sektoriaus įmonės yra plačiausiai atstovaujamos rinkoje, kad eksperimentas būtų reprezentatyvus ir aktualus (OECD, 2024b).

Šie kriterijai leido atrinkti tokias įmones, kurios ne tik skirtingai naudoja DI ir kitas technologijas, bet taip pat susiduria su skirtingais kibernetinio saugumo iššūkiais:

- Kredito ir mokėjimų rinkos hipotetinė įmonė buvo pasirinkta, nes atitinka DI technologijų naudojimo masto kriterijų – plačiai diegia automatizuotas sistemas sukčiavimo prevencijai ir mokėjimų apdorojimui. Be to, ši rinka pasižymi aukštu pažeidžiamumu, kadangi apima jautrių klientų duomenų valdymą ir realiuoju laiku vykdomas finansines operacijas (Mark Cankett, 2022; McKinsey, 2023).
- Draudimo rinkos hipotetinė įmonė buvo pasirinkta dėl savo aukšto pažeidžiamumo kriterijaus, kadangi ji valdo didelius klientų duomenų kiekius ir naudoja pažangias DI platformas rizikos vertinimui. Be to, draudimo įmonės yra vienos iš plačiausiai atstovaujamų finansų rinkoje, todėl jų analizė užtikrina reprezentatyvumą (OECD, 2020, 2024b).

- Finansinių priemonių rinkos hipotetinė įmonė buvo pasirinkta dėl didelio DI technologijų naudojimo algoritmo pagrindu veikiančiose prekybos sistemose ir portfelių valdyme. Ši įmonė taip pat pasižymi aukštu technologiniu pažeidžiamumu, kadangi algoritminė prekyba yra itin jautri kibernetinėms grėsmėms ir technologiniams trikdžiams (Deloitte, 2024b; Dr Florian Ostmann & Dr Cosmina Dorobantu, 2021).
- Kriptoturto rinkos hipotetinė įmonė buvo pasirinkta remiantis pažeidžiamumo ir rinkos aktualumo kriterijais. Kripto valiutų keitimo ir saugojimo paslaugos yra itin jautrios kibernetinėms atakoms, o ši rinka sparčiai auga ir tampa vis labiau aktuali, todėl jos analizė leidžia tirti naujausius ir unikalius kibernetinio saugumo iššūkius (Bahaaeddin A. M. Alareeni & Islam Elgedawy, 2023).

5.1.1. Kredito ir mokėjimų rinkos hipotetinė įmonė

Bendra informacija:

Tai mažo dydžio finansinė įstaiga, kurioje dirba 45 specialistai. Įmonė orientuojasi į vartotojų ir smulkaus verslo mokėjimų sprendimus, kredito paskolas bei mokėjimo kortelių paslaugas. Ji aptarnauja daugiau nei 10 000 klientų Baltijos regione ir stengiasi pritaikyti naujausias skaitmenines technologijas, kad paslaugos būtų patogios, greitos ir saugios.

Veiklos sritys:

- Paskolų valdymas: asmeninių ir smulkaus verslo paskolų suteikimas bei administravimas.
- Mokėjimų apdorojimas: elektroninių mokėjimų vykdymas, įskaitant vietinius ir tarptautinius pervedimus.
- Sukčiavimo prevencija: automatizuotų sprendimų naudojimas, siekiant identifikuoti ir užkirsti kelią finansinėms apgavystėms.

IT infrastruktūra:

- Įmonė naudoja debesijos paslaugas, tokias kaip „Amazon Web Services (AWS)“, ir „Azure“, kad optimizuotų mokėjimų procesus bei užtikrintų greitą duomenų apdorojimą.
- Naudojamos pažangios mašininio mokymosi sistemos, tokios kaip: „XGBoost“ ir „Random Forest“, rizikos ir sukčiavimo analizei.
- Naudojami vietiniai serveriai jautrių klientų duomenų saugojimui, užtikrinant atitiktį GDPR ir PCI DSS standartais.
- Centralizuotos duomenų bazės klientų paskyroms ir operacijų valdymui.

Saugumo politika:

- Standartinis duomenų šifravimas naudojant svetainės šifravimo protokolą – TLS 1.3.
- Periodinės vidinės ir išorinės saugumo auditų procedūros.
- Realioju laiku veikiančios anomalijų aptikimo sistemos, integruotos į operacijų procesus.

5.1.2. Draudimo rinkos hipotetinė įmonė

Bendra informacija:

Vidutinio dydžio draudimo įmonė, teikianti paslaugas gyvybės, sveikatos, automobilių bei nekilnojamojo turto draudimo srityse. Įmonė turi daugiau nei 50 000 klientų ir aktyviai plečia paslaugų spektrą skaitmeninėje erdvėje. Joje dirba 120 darbuotojų, įskaitant draudimo rizikos analitikus ir žalų ekspertus.

Veiklos sritys:

- Rizikos vertinimas: klientų draudimo rizikos analizė ir individualių sprendimų kūrimas.
- Žalų administravimas: automatizuota sistema, leidžianti greitai ir efektyviai tvarkyti draudimo išmokų paraiškas.
- Draudimo rinkos analizė: pažangi duomenų analizė siekiant numatyti rinkos tendencijas ir tobulinti draudimo paslaugas.

IT infrastruktūra:

- Įmonė naudoja „Guidewire“ platformą, kuri leidžia automatizuoti draudimo operacijas ir efektyviai tvarkyti klientų duomenis.
- IT infrastruktūra apima debesijos integraciją, užtikrinančią greitą duomenų prieinamumą ir lankstumą.
- Vietiniai serveriai su RAID (Redundant Arrays of Independent Disks) saugojimo technologija užtikrina jautrių duomenų saugumą ir pasiekiamumą.
- Integruotos dirbtinio intelekto platformos (DataRobot) rizikos vertinimui ir draudimo kainodaros modeliams.

Saugumo politika:

- Visi draudimo duomenys apsaugoti AES-256 šifravimu.
- Dviejų faktorių autentifikacija užtikrina tik patvirtintų vartotojų prieigą prie IT sistemų.
- Kasmet vykdomi kibernetinio saugumo mokymai darbuotojams ir realūs grėsmių testavimo scenarijai.

5.1.3. Finansinių priemonių rinkos hipotetinė įmonė

Bendra informacija:

Tai nedidelė, bet pažangi finansinių investicijų bendrovė, kurioje dirba 30 darbuotojų. Įmonė teikia paslaugas algoritminės prekybos, akcijų rinkos analizės ir portfelių optimizavimo srityse. Ji orientuojasi tiek į individualius, tiek į institucinius klientus.

Veiklos sritys:

- Algoritminė prekyba: automatinės prekybos sistemos, leidžiančios vykdyti sandorius realiuoju laiku.
- Investicinių portfelių valdymas: klientų turto valdymas, pasitelkiant AI algoritmus.
- Rinkos analizė: finansinių duomenų apdorojimas siekiant prognozuoti rinkos tendencijas.

IT infrastruktūra:

- Naudojama „BlackRock Aladdin“ platforma investicijų duomenų analizei.
- Įmonė naudoja „Kensho“ ir „Two Sigma“ algoritmus, leidžiančius optimizuoti prekybos strategijas.
- Serveriai su didelės talpos duomenų laikymo technologija, integruoti su debesijos sprendimais.

Saugumo politika:

- Operacijos apsaugotos blokų grandinės technologijomis, kad būtų užtikrintas duomenų integralumas.
- Realus laiko tinklo stebėseną, naudojant pažangius įrankius, tokius kaip „Fortinet“.
- Reguliariai organizuojami darbuotojų mokymai, orientuoti į kibernetines grėsmes.

5.1.4. Kriptoturto rinkos hipotetinė įmonė

Bendra informacija:

Įmonė teikia kriptovaliutų keitimo, saugojimo ir decentralizuotų finansų paslaugas. Įmonėje dirba 40 darbuotojų, daugiausia IT specialistai ir kriptovaliutų analitikai.

Veiklos sritys:

- Kriptovaliutų keitimas: valiutų keitimo paslaugos.
- Kriptovaliutų saugojimas: fizinės ir skaitmeninės piniginės sprendimai.
- Rinkos prognozės: duomenų analizė, skirta nustatyti rinkos pokyčius.

IT infrastruktūra:

- Naudojamos blokų grandinės technologijos („Chainalysis“, „Elliptic“) sandorių saugumui.
- Decentralizuoti serveriai ir saugumo priemonės šaltosioms piniginėms valdyti.
- Specializuota infrastruktūra, užtikrinanti greitus sandorius ir DDoS prevenciją.

Saugumo politika:

- Griežtos autentifikavimo procedūros kiekvienam sandoriui.
- Blokų grandinės technologija apsaugo nuo klastojimo ir sukčiavimo.
- Periodiniai kibernetinio testai nepalankiomis sąlygomis ir išorės auditai.

5.2. Eksperimentas su generatoriumi ir hipotetinėmis įmonėmis

Eksperimento tikslas – įvertinti sukurto kibernetinio saugumo rizikos scenarijų generatoriaus veiksmingumą finansų sektoriuje. Šiam tikslui buvo pasirinktos keturios skirtingos hipotetinės įmonės, kurios atspindi pagrindines finansų rinkas. Pateikus kiekvienos įmonės kontekstą, buvo užduotos specializuotos užklausos kibernetinio saugumo rizikos scenarijų generatoriui. Eksperimento metu bus siekiama pasiekti šiuos tikslus:

1. Įvertinti generatoriaus gebėjimą generuoti specifinius kibernetinio saugumo rizikos scenarijus, atsižvelgiant į įmonių IT infrastruktūros ypatumus, DI naudojimo apimtį ir veiklos sritis.
2. Įvertinti, ar sugeneruoti scenarijai atspindi realius ir aktualius kibernetinio saugumo iššūkius, su kuriais susiduria finansų sektoriaus įmonės.
3. Identifikuoti, kokios kibernetinės grėsmės yra pavojingiausios konkrečioms įmonėms, ir įvertinti, ar generatorius sugeba atskleisti pagrindines pažeidžiamumo sritis.

Šie tikslai gali padėti ne tik patikrinti generatoriaus veikimo efektyvumą, bet ir nustatyti jo praktinę naudą stiprinant kibernetinio saugumo strategijas. Be to, eksperimento rezultatai suteiks galimybę įvertinti, kaip skirtingos įmonių IT infrastruktūros ir DI naudojimo lygiai daro įtaką jų pažeidžiamumui bei gebėjimui reaguoti į kibernetines grėsmes.

Eksperimento eiga

1. Įmonių kontekstų parengimas. Kiekvienos hipotetinės įmonės duomenys iš 5.1 skyriaus bus apdoroti ir pateikti generatoriui kaip kontekstinės užklausos. Šiame etape apimama:
 - Įmonės veiklos sritį ir IT infrastruktūros aprašymas.
 - Saugumo politika ir naudojamas technologijas (pvz., DI sistemas, šifravimo standartus).
 - Potencialias sritis, kuriose gali pasireikšti kibernetinės rizikos.
 - Galimybė kontekstą vartotojui sugeneruoti atsakant į klausimą, kuris yra aprašytas 3.5 skyriuje.
2. Rizikos scenarijų generavimas. Sukurtas generatorius, naudodamasis įmonės kontekstu, sugeneruos kibernetinio saugumo rizikos scenarijus. Šiame etape bus atliekamos užduotys:
 - Analizė, kokie pažeidžiamumai dažniausiai gali kilti pasirinktos finansų įmonės veiklos srityje.
 - Specifinių grėsmių scenarijų pateikimas, atsižvelgiant į naudojamas technologijas ir taikomas saugumo priemones.

5.2.1. Eksperimentas su kredito ir mokėjimų rinkos hipotetine įmone

Eksperimentas buvo atliktas hipotetinėje kredito ir mokėjimų paslaugų įmonėje, kuri yra nedidelė finansų įstaiga, ir specializuojasi paskolų valdyme, mokėjimų apdorojime ir sukčiavimo prevencijoje. Šio eksperimento tikslas buvo pasitelkti kibernetinių grėsmių rizikos scenarijų generatorių, kad būtų nustatyti ir įvertinti galimi organizacijos pažeidžiamumai.

Vartotojas pateikė užklausą pagal generatoriaus paprašytą struktūrą (22 pav.). Vartotojo pateikta užklausa buvo labiausiai orientuota į grėsmių identifikavimą ir jų poveikio analizę (9 pav. Užklausos forma grėsmių šaltinių nustatymui):

1. Organizacijos aprašymas:

Mažo dydžio finansų įstaiga, turinti 45 darbuotojus ir daugiau nei 10 000 klientų Baltijos regione. Kurios pagrindinės veiklos sritys – paskolų valdymas, mokėjimų apdorojimas ir sukčiavimo prevencija. Ir taip pat naudoja debesų kompiuterijos sprendimus („Amazon Web Services“, „Azure“), bei mašininio mokymosi modelius. Įmonė laikosi GDPR ir PCI DSS standartų.

2. Grėsmių šaltinių identifikavimas:

- Vidiniai šaltiniai: darbuotojų klaidos, prieigos kontrolės trūkumai.
- Išoriniai šaltiniai: kibernetinės atakos, reguliavimo pažeidimai.

3. Tikėtini veiksmai ir motyvai:

- Tikėtini veiksmai: duomenų vagystė, DDoS atakos, sukčiavimo modelių manipuliavimas.
- Motyvai: finansinė nauda, reputacijos kenkimas.

This is a small financial institution in the credit and payment processing industry with 45 employees. The company prioritizes compliance with GDPR and PCI DSS, protects sensitive customer data using TLS 1.3 encryption, anomaly detection systems, and cloud services like AWS and Azure, while addressing risks such as fraud, data breaches, and potential DDoS attacks.

22 pav. Pirmo eksperimento pateikta užklausa generatoriui

Generatorius, naudodamas pateiktą užklausą, sukūrė individualizuotus scenarijus, atspindinčius hipotetinės įmonės pažeidžiamumus ir galimus grėsmių poveikius (9 priedas). Pagal pateiktą užklausą generatorius sugeneravo penkis scenarijus:

1 scenarijus. Sukčiavimo prevencijos sistemų manipuliacija

- Aprašymas: priešiškos atakos gali iškraipyti mašininio mokymosi sistemų modelio prognozes, leidžiant sukčiams pasinaudoti sistemos spragomis.

- Pasekmės: finansiniai nuostoliai, įmonės patikimumo praradimas.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų bazė:
 - Scenarijus remiasi ENISA dokumentacija (ENISA, 2020), kurioje nurodoma, kad dirbtinio intelekto modeliai yra jautrūs manipuliavimui su DI sistemoje naudojamais duomenimis.
 - Įmonė naudoja „XGBoost“ ir „Random Forest“ įrankius, todėl jiems kyla tiesioginė rizika su jautrių duomenų manipuliavimu.
- Nauda įmonei: scenarijus padeda įmonei identifikuoti mašininio mokymosi modelių apsaugos trūkumus ir pritaikyti prevencinius algoritmus, tokius kaip anomalijų aptikimas modelio mokymo metu.

2 scenarijus. Vidinio darbuotojo neteisėta veikla su klientų duomenimis

- Aprašymas: nepatenkintas darbuotojas tyčia pažeidžia sistemų konfigūraciją, sukeldamas duomenų vagystę.
- Pasekmės: finansiniai nuostoliai, reputacijos praradimas.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų bazė: ISO 31000 (ISO 31000, 2018) rekomenduoja vertinti vidinius grėsmių šaltinius kaip dalį organizacijos rizikos valdymo.
- Nauda įmonei: šis scenarijus pataria įdiegti prieigos teisių kontrolės politiką ir darbuotojų veiksmų stebėjimą.

3 scenarijus. DDoS ataka prieš mokėjimų apdorojimo sistemas

- Aprašymas: DDoS ataka gali sutrikdyti mokėjimų apdorojimo sistemų veikimą, dėl ko laikinai gali nepavykti vykdyti operacijų.
- Pasekmės: finansiniai nuostoliai, klientų pasitikėjimo praradimas.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų bazė:
 - NIST (NIST, 2021d) rekomenduoja įmonėms parengti tinklo atsparumo strategijas tokioms atakoms, kaip DDoS.
 - ENISA dokumentacijoje (ENISA, 2020) pabrėžiama, kad kibernetinės atakos, tokios kaip DDoS, tiesiogiai gali pakenkti paslaugų tęstinumui.
 - Įmonės veiklos aprašyme nurodyta, kad mokėjimų apdorojimas yra viena iš pagrindinių veiklų, tai tokia ataka tiesiogiai gali pažeisti įmonės operacijų stabilumą ir gali lemti nepasitenkinimą tarp klientų.
- Nauda įmonei: įmonė gali įgyvendinti tinklo apkrovos valdymo ir DDoS apsaugos priemones, taip sumažinant galimą šių atakų poveikį.

4 scenarijus. Debesų paslaugų saugumo pažeidimas

- Aprašymas: kibernetiniai nusikaltėliai pasinaudoja netinkama AWS ar Azure platformos konfigūracija, įgaudami prieigą prie klientų duomenų.
- Pasekmės: duomenų nutekėjimas, pažeidžiant GDPR reikalavimus, reputacijos ir klientų pasitikėjimo praradimas.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų bazė:

- ENISA (ENISA, 2020) nurodo, kad debesijos paslaugų platformos dažniausiai susiduria su duomenų pažeidimo atakomis.
- GDPR (Europos parlamentas ir Europos sąjungos taryba, 2016) akcentuoja būtinybę apsaugoti klientų duomenis ir numato sankcijas už pažeidimus.
- Nauda įmonei: šis scenarijus leidžia įmonei iš anksto įvertinti debesų paslaugų konfigūracijos trūkumus ir įdiegti griežtesnes prieigos kontrolės priemones.

5 scenarijus. DI skatinami sukčiavimo bandymai apeinant anomalijų aptikimo sistemas

- Aprašymas: anomalijų aptikimo sistema generuoja klaidingai teigiamus rezultatus, trukdančius įmonės operacijoms.
- Pasekmės: veiklos neefektyvumas, padidėjusios administracinės išlaidos.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų baze: ENISA (ENISA, 2022) rekomenduoja reguliariai testuoti anomalijų aptikimo sistemas, siekiant užtikrinti jų tikslumą.
- Nauda įmonei: šis scenarijus pasiūlo įmonei tobulinti algoritmus ir sumažinti klaidingų signalų tikimybę.

Taigi, eksperimentas patvirtino, kad pateikta vartotojo užklausa buvo tinkamai struktūrizuota, o scenarijų generatoriaus sugeneruoti scenarijai yra logiški, pritaikomi ir naudingi hipotetinio subjekto kontekste. Sugeneruoti scenarijai ne tik identifiko svarbiausius grėsmių šaltinius, bet ir pateikė praktinių rekomendacijų, kaip juos valdyti. Tai patvirtino generatoriaus veikimo tikslumą ir jo naudingumą praktiniame rizikos valdyme, kur hipotetinė įmonė gali pasinaudoti šiais rezultatais stiprindama kibernetinį saugumą bei gerindama savo procesų saugos ir atitikties valdymą. Šis eksperimentas parodė, kad generatorius yra efektyvus įrankis, skirtas rizikos valdymo procesų tobulinimui.

5.2.2. Eksperimentas su draudimo rinkos hipotetine įmone

Eksperimentas buvo atliktas su hipotetine vidutinio dydžio draudimo rinkos įmone, teikiančia gyvybės, sveikatos, automobilių ir nekilnojamojo turto draudimo paslaugas. Įmonė aptarnauja daugiau nei 50 000 klientų ir aktyviai veikia skaitmeninėje erdvėje. Joje dirba 120 darbuotojų, įskaitant draudimo rizikos analitikus ir žalių ekspertus. Eksperimento tikslas – naudojant kibernetinių grėsmių rizikos scenarijų generatorių, identifikuoti galimus organizacijos pažeidžiamumus, įvertinti labiausiai tikėtinus ir mažiausiai tikėtinus scenarijus bei parengti rizikos valdymo rekomendacijas.

Vartotojas pateikė užklausą pagal generatoriaus paprašytą struktūrą (23 pav.). Vartotojo pateikta užklausa buvo labiausiai orientuota į grėsmių identifikavimą ir jų poveikio analizę (9 pav. Užklaustos forma grėsmių šaltinių nustatymui):

1. Organizacijos aprašymas:

Vidutinio dydžio draudimo įmonė, teikianti įvairių rūšių draudimo paslaugas. Naudoja „Guidewire“ platformą draudimo operacijų automatizavimui, debesų kompiuterijos sprendimus ir vietinius RAID serverius duomenų saugumui užtikrinti. Taip pat

„DataRobot“ dirbtinio intelekto platformas rizikos vertinimui. Saugumo politika: AES-256 šifravimas, dviejų faktorių autentifikacija ir kasmetiniai darbuotojų mokymai.

2. Grėsmių šaltiniai:

- Vidiniai: netinkama sistemų konfigūracija, darbuotojų klaidos.
- Išoriniai: kibernetiniai nusikaltimai, trečiųjų šalių paslaugų teikimo spragos.

3. Tikėtini veiksmai ir motyvai:

- Veiksmai: netinkami dirbtinio intelekto sistemos gaunami rezultatai, darbuotojų klaidos, duomenų vagystės.
- Motyvai: finansinė nauda, klientų duomenų išgavimas.

Industry: Insurance industry specializing in life, health, automobile, and property insurance.
Size: Medium-sized organization with over 120 employees and more than 50,000 clients.
Cybersecurity solutions:
Guidewire platform for automated insurance operations and customer data management.
Cloud-based infrastructure ensuring flexible and accessible data processing.
RAID storage systems for secure and redundant data storage.
AI-driven platforms like DataRobot for risk evaluation and pricing.
AES-256 encryption and multi-factor authentication for securing sensitive data.
Attack concerns:
Ransomware targeting customer and operational data.
Phishing targeting employees or customers.
Insider threats impacting sensitive information.
Vulnerabilities in cloud integration systems.
Regulatory or compliance requirements: Compliance with GDPR and regular cybersecurity resilience testing.

23 pav. Antro eksperimento pateikta užklausa generatoriui

Generatorius, naudodamas pateiktą užklausa, sukūrė individualizuotus scenarijus, atspindinčius draudimo įmonės pažeidžiamumus ir galimus grėsmių poveikius (10 priedas). Pagal pateiktą užklausa generatorius sugeneravo penkis scenarijus:

1 scenarijus. Ransomware ataka, klientų duomenų vagystė iš „Guidewire“ platformos

- Aprašymas: kibernetiniai nusikaltėliai pasinaudoja platformos pažeidžiamumais arba netinkama prieigos kontrolės politika, siekdami prieigos prie klientų duomenų.
- Pasekmės: užšifruoti klientų ir veiklos duomenys, teisinės sankcijos, reputacijos praradimas.
- Scenarijaus pagrindimas generatoriaus vidine duomenų baze:
 - ENISA (ENISA, 2020) gairės pabrėžia jautrumą duomenų vagystėms, jeigu platformos pakankamai nėra apsaugotos.

- GDPR (Europos parlamentas ir Europos sąjungos taryba, 2016) pateikia reikalavimus duomenų apsaugai.
- DORA reglamentas (the european parliament and the council of the european union, 2022) akcentuoja, kad finansų įstaigos privalo turėti mechanizmus, skirtus IRT incidentams valdyti ir pranešti apie duomenų saugumo pažeidimus, kad būtų užtikrintas paslaugų tęstinumas.
- Nauda įmonei: šis scenarijus leidžia peržiūrėti prieigos kontrolės mechanizmus ir diegti papildomas šifravimo bei stebėjimo sistemas.

2 scenarijus. Vidinio darbuotojo piktybinis ketinimas tyčia sukelti duomenų pažeidimus

- Aprašymas: nepatenkintas darbuotojas tyčia pakeičia klientų draudimo duomenis.
- Pasekmės: GDPR pažeidimas, baudos, klientų pasitikėjimo praradimas ir teisiniai veiksmai.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų bazė:
 - ISO 31000 (ISO 31000, 2018) pabrėžia vidinių grėsmių valdymo svarbą, įskaitant prieigos teisių stebėjimą ir kontrolę
 - DORA reglamentas (the European Parliament and the council of the European Union, 2022) akcentuoja, kad prieiga prie svarbių IRT sistemų turi būti griežtai ribojama ir stebima, siekiant išvengti vidinių grėsmių. Taip pat, reikalauja griežtų prieigos kontrolės politikų ir stebėsenos mechanizmų, kurie užtikrintų darbuotojų veiksmų atsekamumą.
- Nauda įmonei: skatina įdiegti veiksmingesnes darbuotojų prieigos kontrolės ir veiksmų stebėsenos priemones.

3 scenarijus. Išorinių įsilaužėlių taikymas į API spragas.

- Aprašymas: debesijos paslaugų pažeidžiamumai per API spragas leidžia įsilaužėliams pasiekti klientų duomenis.
- Pasekmės: klientų įrašų ir teisinių reikalavimų pažeidimai, veiklos sutrikimas.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų bazė:
 - DORA reglamentas (the European Parliament and the council of the European Union, 2022) rekomenduoja debesijos paslaugų rizikos valdymą ir akcentuoja trečiųjų šalių paslaugų stebėjimą, siekiant sumažinti riziką, susijusią su debesijos sprendimais.
 - ENISA gairės (ENISA, 2020) taip pat apibūdina debesų paslaugų grėsmių valdymo svarbą.
- Nauda įmonei: skatina griežtesnių kontrolės priemonių įgyvendinimą debesijos paslaugoms ir sistemų konfigūracijos tikrinimą, siekiant užtikrinti duomenų saugumą.

4 scenarijus. Dirbtinio intelekto sistemų manipuliacija

- Aprašymas: atakos metu klaidingi duomenys įvedami į „DataRobot“ platformą, sukuriant netikslų rizikos vertinimą arba draudimo kainodaros sprendimus.
- Pasekmės: finansiniai nuostoliai dėl neteisingų draudimo išmokų arba nepakankamai įvertintos rizikos.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų bazė:
 - ENISA dokumentacija (ENISA, 2020) nurodo, kad DI modeliai yra jautrūs duomenų manipuliacijai ir nuodijimui (angl. data poisoning).

- DORA reglamente (the European Parliament and the council of the European Union, 2022) akcentuojama, kad finansų įstaigos turi reguliariai testuoti ir stebėti IRT sistemas, siekiant užtikrinti jų tikslumą ir atsparumą klaidingiems duomenims.
- Nauda įmonei: leidžia įmonei tobulinti modelių apsaugos mechanizmus, diegti duomenų tikrinimo procedūras.

5 scenarijus. Sukčiavimo ataka (angl. Phishing Attack) prieš klientus.

- Aprašymas: kibernetiniai nusikaltėliai imituoja organizaciją, siekdami suklaidinti klientus per masines sukčiavimo kampanijas. Tikslas – išgauti klientų asmeninę ir finansinę informaciją. Toks scenarijus tiesiogiai kenkia įmonės reputacijai ir klientų pasitikėjimui.
- Pasekmės: klientai patiria finansinius nuostolius ir tapatybės vagystę, taip pat sumažėja pasitikėjimas organizacija. Įmonė gali susidurti su teisinėmis pasekmėmis dėl nesugebėjimo apsaugoti klientų duomenų.
- Scenarijaus pagrindimas generatoriaus vidinė duomenų baze:
 - DORA reglamentas (the European Parliament and the council of the European Union, 2022) akcentuoja finansų įmonių atsakomybę valdyti kibernetines grėsmes ir užtikrinti, kad jų sistemos ir paslaugos būtų apsaugotos nuo atakų, kuriomis siekiama pakenkti klientams. Taip pat turi būti pateikiami reikalavimai rizikos valdymui, siekiant apsaugoti finansų subjektų turtą – klientų pasitikėjimą ir reputaciją.
 - ENISA (ENISA, 2022) rekomenduoja sukurti grėsmių stebėjimo ir aptikimo sistemas, kurios galėtų apimti sukčiavimo atakų stebėseną. Be to, ENISA akcentuoja klientų edukacijos svarbą siekiant sumažinti tokio tipo atakų poveikį organizacijos veiklai.
- Nauda įmonei: šis scenarijus leidžia įmonei įgyvendinti proaktyvias priemones, skirtas apsaugoti klientus nuo sukčiavimo atakų, t.y. klientams gali būti pateikiamos gairės, kaip atpažinti sukčiavimo atvejus ir apsaugoti nuo jų. Taip pat stebėjimas tamsiojo žiniatinklio (angl. dark web), siekiant aptikti sukčiavimo atvejus, susijusius su įmonės vardu ar domenais. Bei siūlymas įdiegti specializuotus sprendimus, skirtus stebėti ir aptikti sukčiavimo kampanijas.

Eksperimento rezultatai parodė, kad kibernetinių grėsmių scenarijų generatorius gali sėkmingai identifikuoti ir modeliuoti draudimo rinkos pažeidžiamumus pagal pateikiamą įmonės kontekstą. Sugeneruoti scenarijai buvo pagrįsti tokiomis dokumentacijomis kaip: ISO standartais, DORA reglamentu ir ENISA rekomendacijomis, pagal kurias buvo sugeneruotos aišrios gairės, kaip valdyti ir identifikuoti grėsmes. Šie scenarijai pateikė praktines išvalgas ir rekomendacijas, kurios gali padėti draudimo įmonei efektyviau valdyti riziką ir stiprinti kibernetinį atsparumą.

5.2.3. Eksperimentas su finansinių priemonių rinkos hipotetine įmone

Eksperimentas buvo atliktas su finansinių investicijų hipotetine įmone, kuri teikia algoritminės prekybos, akcijų rinkos analizės ir portfelių optimizavimo paslaugas. Įmonė aptarnauja tiek individualius, tiek institucinius klientus. Eksperimento tikslas buvo panaudoti kibernetinių grėsmių rizikos scenarijų generatorių, siekiant identifikuoti galimas rizikas įmonės veiklai pagal turimus IT infrastruktūros sprendimus bei pateikti rekomendacijas rizikos valdymui, remiantis tarptautiniais standartais.

Vartotojas pateikė užklausą pagal generatoriaus paprašytą struktūrą (24 pav.). Vartotojo pateikta užklausa buvo labiausiai orientuota į grėsmių identifikavimą ir galimas pasekmes (9 pav. Užklauso forma grėsmių šaltinių nustatymui):

1. Organizacijos aprašymas:

Nedidelė finansinių investicijų įmonė su 30 darbuotojais. Pagrindinės veiklos sritys: algoritminė prekyba, investicinių portfelių valdymas, finansų rinkos analizė. Naudoja „BlackRock Aladdin“ platformą investicijų duomenų analizei, „Kensho“ ir „Two Sigma“ algoritmus prekybos strategijoms optimizuoti. Taip pat naudoja didelės talpos duomenų saugojimo technologijas, integruotas su debesų sprendimais.

2. Grėsmių šaltiniai:

- Vidiniai šaltiniai: darbuotojų klaidos, sistemų konfigūracijos trūkumai.
- Išoriniai šaltiniai: kibernetinės atakos, algoritmų manipuliavimas.

3. Tikėtini veiksmai ir motyvai:

- Tikėtini veiksmai: duomenų apsaugos pažeidimai, prekybos sistemų manipuliacija, tiekimo grandinės atakų prieš integruotas debesijos paslaugas prevencija.
- Motyvai: finansinė nauda, konkurencinis pranašumas.

Industry:
Financial industry, focusing on algorithmic trading, investment portfolio management, and financial market analysis.

Size of the Organization:
Small (30 employees).

Cybersecurity Solutions in Use:

BlackRock Aladdin platform for financial data analysis.
Kensho and Two Sigma algorithms for optimizing trading strategies.
High-capacity servers with integrated cloud solutions.
Real-time network monitoring tools such as Fortinet.

Specific Attack Types and Concerns:

Risks of attacks on algorithmic trading systems.
Breaches in portfolio management data integrity.
Exploitation of vulnerabilities in blockchain-based transaction security.

Notable Vulnerabilities or Focus Areas:

Customer and institutional data protection.
Ensuring compliance with regulatory frameworks.
Preventing supply chain attacks on integrated cloud services.

24 pav. Antro eksperimento pateikta užklausa generatoriui

Generatorius, naudodamas pateiktą užklausą, sukūrė individualizuotus scenarijus, atspindinčius finansinių investicijų įmonės pažeidžiamumus ir galimus grėsmių poveikius (11 priedas). Pagal pateiktą užklausą generatorius sugeneravo penkis scenarijus:

1 scenarijus. Tikslinė kenkėjiškų programų ataka prieš algoritmines prekybos sistemas.

- Aprašymas: Kibernetiniai nusikaltėliai manipuliuoja „Kensho“ ar „Two Sigma“ algoritmais, siekdami paveikti prekybos sprendimus, sukelti nuostolius įmonei. Manipuliacija gali vykti per netiksliai valdomus mokymosi duomenis ar tiesioginę algoritmų veikimo intervenciją.
- Pasekmės:
 - Neapgalvoti algoritminės prekybos sprendimai sukelia finansinius nuostolius.
 - Reputacijos praradimas tarp institucinių klientų, kuriems svarbus stabilumas.
 - Ilgalaikis pasitikėjimo sumažėjimas įmonės technologinėmis galimybėmis.
- Scenarijaus pagrindimas generatoriaus vidine duomenų baze:
 - Pagal ISO 31000 standartą (ISO 31000, 2018) yra nustatomi rizikos vertinimo procesai taip pat pabrėžiama rizikos valdymo svarba organizacijos valdymo procesuose, kurie gali sukelti organizacinių ir finansinių nuostolių.
 - NIST (NIST, 2021d) pateikia gaires kibernetinio saugumo integravimo į įmonės rizikos valdymą. Dokumente taip pat aptariama, kad papildomos rizikos gali kilti dėl dinamiško įmonės informacijos ir technologijų sudėtingumo (tokių kaip algoritmų pažeidimai), kurie gali tapti svarbia kibernetine grėsme.

- ENISA (ENISA, 2020) akcentuoja DI sistemų pažeidžiamumo valdymo būtinybę ir rekomenduoja reguliariai testuoti modelių saugumą.
- Nauda įmonei: šis scenarijus gali padėti identifikuoti algoritmų apsaugos trūkumus ir sustiprinti DI sistemų saugumą, skatindamas įdiegti saugumo priemones, tokias kaip duomenų vientisumo patikrinimai ar DI modelių stebėsenos sprendimai.

2 scenarijus. Duomenų pažeidimas portfelio valdymo sistemoje.

- Aprašymas: nepatenkintas darbuotojas tyčia modifikuoja klientų portfelių ar prekybos strategijų duomenis, siekdamas sukelti įmonei finansinių ar reputacinių nuostolių. Arba kibernetinis nusikaltėlis, kuris išnaudoja silpnas prieigos kontrolės priemones, taip prieidamas prie portfelio duomenų.
- Pasekmės:
 - Klientų portfelių valdymo duomenų vientisumo praradimas.
 - Teisinės pasekmės dėl sutartinių įsipareigojimų nevykdymo klientų atžvilgiu.
 - Ilgalaikė žala įmonės patikimumui.
- Scenarijaus pagrindimas generatoriaus vidine duomenų baze:
 - ISO 31000 standarte (ISO 31000, 2018) minima, kad vidinių ir išorinių grėsmių valdymo politika yra esminė siekiant užkirsti kelią duomenų vientisumo pažeidimams.
 - NIST dokumentas (NIST, 2021d) akcentuoja, kaip darbuotojų veiksmų kontrolė turi būti integruojama į įmonės rizikos valdymo procesą.
 - DORA reglamente (the European Parliament and the council of the European Union, 2022) reikalaujama užtikrinti griežtas prieigos kontrolės sistemas, kurios mažintų vidinių ir išorinių grėsmių tikimybę.
- Nauda įmonei: šis scenarijus skatina įmonę tobulinti darbuotojų prieigos valdymą, stebėsenos mechanizmus ir diegti veiksmų audito sistemas.

3 scenarijus. „Cloud Services“ tiekimo grandinės ataka

- Aprašymas: kibernetiniai nusikaltėliai pasinaudoja debesijos pagrindu veikiančių sprendimų pažeidžiamumo saugumo spragomis, siekdami įsiskverbti į vidines sistemas ir prieigos prie jautrių įmonės duomenų.
- Pasekmės:
 - Jautrių klientų ir finansinių duomenų praradimas.
 - GDPR pažeidimai ir su tuo susijusios sankcijos.
 - Ilgalaikiai reputacijos ir finansiniai nuostoliai.
- Pagrindimas dokumentacijomis:
 - ENISA akcentuoja debesų sprendimų saugumo riziką ir nuolatinės stebėsenos svarbą.
 - DORA nurodo, kad debesų infrastruktūra turi būti reguliariai audituojama ir apsaugota nuo galimų saugumo spragų.
 - ISO 31000 pateikia, kad duomenų saugumo rizikos vertinimas yra vienas iš pagrindinių organizacijos rizikos valdymo aspektų.
- Nauda įmonei: scenarijus skatina reguliarias debesų infrastruktūros saugumo peržiūras ir stipresnių apsaugos mechanizmų diegimą ir dvigubos autentifikacijos prieigą prie debesies.

4 scenarijus. Blockchain pagrįstų operacijų saugumo pažeidžiamumų išnaudojimas

- Aprašymas:

Kenkėjiški veikėjai, turintys blockchain technologijų patirties, pasinaudoja išmaniosiose sutartyse esančiomis saugumo spragomis arba vykdo 51% atakas (51% ataka yra subjekto ar grupės, kontroliuojančios daugiau nei 50% tinklo ataka prieš kriptovaliutų blokų grandinę (Investopedia, 2024)), siekdami finansinio pelno ar sukčiaujant atlikti sandorius.

- Pasekmės:

- Finansiniai nuostoliai dėl neteisėtų sandorių ar turto praradimo.
- Blockchain technologijos patikimumo sumažėjimas klientų ir partnerių akyse.
- Sutrikimai operacijų procese dėl išmaniosios sutarties pažeidimų.

- Scenarijaus pagrindimas generatoriaus vidinė duomenų baze:

- ISO 31000 standarte aptariama rizikos valdymo svarba įvairiose sistemose, kuriose duomenys yra jautrūs netikėtoms spragoms.
- ENISA (ENISA, 2022) pabrėžia paslaugų atsparumą galimoms atakoms kaip vieną pagrindinių saugumo aspektų, kurie gali apimti išmaniųjų sutarčių auditus ir 51% atakų prevencijos mechanizmus.
- DORA reglamentas (the European Parliament and the council of the European Union, 2022) apibrėžia IRT rizikos valdymą finansiniuose subjektuose.

- Nauda įmonei: šis scenarijus pasiūlo įmonei vykdyti reguliarius išmaniųjų sutarčių auditus, užtikrinant jų patikimumą. Taip pat kurti atsparumo mechanizmus, siekiant išvengti 51% atakų, pavyzdžiui, decentralizuoto tinklo stiprinimą. Bei sumažinti operacinius sutrikimus ir užtikrinti klientų bei institucinių partnerių pasitikėjimą blockchain technologijomis.

5 scenarijus. DDoS ataka prieš stebėjimo infrastruktūrą realiuoju laiku

- Aprašymas: kibernetiniai veikėjai, tokie kaip hacktivistai, vykdo DDoS ataką, siekdami perkrauti realaus laiko tinklo stebėsenos infrastruktūrą („Fortinet“ platforma). Ataka siekiama sutrikdyti stebėsenos funkcionalumą, kad būtų sudarytos sąlygos kitoms lygiagrečioms atakoms.

- Pasekmės:

- Sumažėjęs tinklo veiklos matomumas, kas leidžia atakoms likti nepastebėtoms.
- Neaptikti tinklo pažeidimai, kurie gali sukelti duomenų vagystes ar kitas rizikingas operacijas.
- Paslaugų teikimo sutrikimai, ypač piko prekybos laikotarpiais, kas sukelia tiesioginius finansinius nuostolius.

- Scenarijaus pagrindimas generatoriaus vidinė duomenų baze:

- ISO dokumente (ISO 31000, 2018) yra akcentuojama svarba valdyti operacines rizikas, kaip pvz., tinklo stebėjimo sistemų apsaugą nuo DDoS atakų, kurios gali tiesiogiai paveikti paslaugų tęstinumą.
- NIST (NIST, 2021d) rekomenduoja įtraukti tinklo apsaugos ir priežiūros sistemas į bendrą organizacijos rizikos valdymo procesą.

- DORA reglamentas (the uropean parliament and the council of the european union, 2022) nurodo būtinybę užtikrinti tinklo infrastruktūros atsparumą atakoms, kurios gali paveikti operacinį tęstinumą, ypač piko laikotarpiais.
- ENISA (ENISA, 2020) pabrėžia tinklo infrastruktūros atsparumo svarbą, produktų tokių kaip realaus laiko sistemoms.
- Nauda įmonei: scenarijus leidžia įmonei įgyvendinti pažangius DDoS apsaugos sprendimus, tokius kaip srauto filtravimo įrankiai. Taip pat sukurti papildomus stebėsenos mechanizmus, kurie užtikrintų tinklo stebėseną net esant pertekliniai apkrovai ir sustiprinti reagavimo planus, kad būtų užtikrintas tinklo stebėjimo tęstinumas per kritinius prekybos laikotarpius.

Taigi, trečiasis eksperimentas, atliktas finansinių priemonių rinkos hipotetinėje įmonėje, identifikavo grėsmes, kurios apima algoritmų manipuliaciją, vidinių darbuotojų piktavališkus veiksmus, DDoS atakas ir blockchain infrastruktūros pažeidžiamumus. Sugeneruoti scenarijai buvo pagrįsti tarptautinėmis gairėmis, įskaitant ISO 31000, NIST, DORA reglamentą, ENISA rekomendacijas ir GDPR, kas patvirtina jų tikslumą ir aktualumą. Kiekviename scenarijuje buvo pateikti visi scenarijų kriterijai, kurie išryškino potencialių grėsmių pasekmes įmonei bei sprendimų priemonės, leidžiančias sušvelninti rizikas. Rezultatai pabrėžė, kaip svarbu užtikrinti algoritminės prekybos sistemų saugumą, stiprinti vidinę kontrolę, diegti tinklo apsaugos priemones ir reguliariai audituoti blockchain sistemas. Eksperimentas atskleidė, kad įmonė, remdamasi šiais scenarijais, gali ne tik sumažinti savo operacinę ir reputacinę riziką, bet ir sustiprinti kibernetinį atsparumą, užtikrinant veiklos tęstinumą ir klientų pasitikėjimą.

5.2.4. Eksperimentas su kriptoturto rinkos hipotetine įmone

Ketvirtasis eksperimentas buvo atliktas su kriptoturto rinkos hipotetine įmone, kuri specializuojasi kriptovaliutų keitimo, saugojimo ir decentralizuotų finansinių paslaugų srityse. Įmonėje dirba 40 darbuotojų, daugiausia IT specialistai ir kriptovaliutų analitikai. Eksperimento tikslas buvo nustatyti potencialias kibernetines grėsmes, rizikas ir galimus scenarijus, siekiant užtikrinti įmonės saugumą ir atitikti tarptautiniams standartams.

Eksperimentas buvo pasirinktas atlikti su generatoriaus pateikiamais klausimais iš kurių toliau būtų suformuluojamas įmonės kontekstas ir pateikiami kibernetinio saugumo rizikos scenarijai. Klausimyno atsakymai buvo formuluojami remiantis įmonės aprašymu ir jos veiklos specifika, o pateikti atsakymai atspindi esamus organizacijos prioritetus, procesus ir iššūkius:

- Organizacinio (vidinio) konteksto atsakymai. Įmonė demonstruoja aiškų vadovybės įsipareigojimą valdyti kibernetines rizikas, integruodama jas į bendrą verslo strategiją. Didelė dalis IT biudžeto (20–30%) yra skiriama kibernetinio saugumo priemonėms, o tai pabrėžia prioritetinį šios srities vaidmenį organizacijos veikloje. Vidinė kultūra taip pat yra gerai struktūrizuota – darbuotojų kibernetinio saugumo suvokimas vertinamas reguliariais apklausų ir sertifikavimo procesais. Šie atsakymai patvirtina, kad įmonė kuria stiprų vidinį pagrindą

kovai su kibernetinėmis grėsmėmis, kas itin svarbu veiklai, susijusiai su jautriomis finansinėmis technologijomis, tokiomis kaip blokų grandinės infrastruktūra.

- Išorinio konteksto atsakymai. Kriptovaliutų rinka, būdama dinamiška ir globali, susiduria su unikaliais išorinių grėsmių iššūkiais. Atsakymai atskleidžia, kad įmonė pasitelkia pažangias grėsmių žvalgybos platformas, leidžiančias stebėti išorines kibernetines grėsmes, tokias kaip DDoS atakos ar blockchain manipuliacijos. Trečiųjų šalių valdymas taip pat užima svarbią vietą – įmonė reguliariai audituoja jų saugumo praktikas. Ši praktika atitinka specifinį sektoriaus poreikį bendradarbiauti su išoriniais tiekėjais („Chainalysis“, „Elliptic“), siekiant užtikrinti sandorių saugumą.
- Rizikos valdymo sistema. Atsakymai rodo, kad įmonė turi stiprią rizikos valdymo sistemą, kuri yra integruota į bendrą organizacijos veiklos valdymą. Naudojant „Enterprise risk management“ (ERM) sistemas, įmonė efektyviai derina kibernetines rizikas su kitomis veiklos rizikomis. Šis požiūris ypač svarbus kriptovaliutų rinkoje, kur kibernetinės grėsmės gali turėti tiesioginį poveikį finansiniams rezultatams ir klientų pasitikėjimui. Pabrėžta atitiktis GDPR reikalavimams dar labiau sustiprina įmonės atsparumą reguliaciniams iššūkiams.
- Rizikos vertinimas ir scenarijų kūrimas. Įmonė savo rizikos vertinimą grindžia istorinių duomenų ir tendencijų analize bei remiasi tarptautiniais standartais, tokiais kaip NIST ir ISO. Tai leidžia kurti realistiškus scenarijus, kurie atspindi galimas grėsmes, tokias kaip kriptovaliutų vagystės, šaltųjų piniginių pažeidimai ar blockchain manipuliacijos. Atsakymai pabrėžia struktūrizuotą požiūrį į rizikos scenarijų kūrimą, kas ypač gali būti svarbu organizacijai, dirbančiai su decentralizuotomis finansų technologijomis.
- Incidentų valdymas ir atkūrimas. Įmonės incidentų valdymo procesai yra nuosekliai struktūrizuoti. Reguliariai vykdomi incidentų valdymo planų testavimai ir pagrindinių procesų (pvz., duomenų atsarginės kopijos) įgyvendinimas rodo aukštą pasirengimo lygį. Atsakomybė už incidentų raportavimą integruota į bendrą atitikties sistemą, užtikrinant atitikimą reguliaciniams reikalavimams, tokiems kaip GDPR. Tai dar kartą patvirtina įmonės gebėjimą efektyviai reaguoti į finansines ir kibernetines grėsmes.

Klausimyno atsakymai pateikia aiškų ir struktūruotą kriptoturto rinkos hipotetinės įmonės kibernetinio saugumo kontekstą pagal kurį generatorius toliau pateikė sugeneruotus kibernetinio saugumo rizikų scenarijus (12 priedas):

1 scenarijus. Vidinio darbuotojo veikla, sukelianti duomenų nutekėjimą.

- Aprašymas: nepatenkintas arba aplaidus darbuotojas, turintis prieigą prie jautrios informacijos, neteisėtai dalijasi ar nutekina duomenis. Tai gali būti padaryta tiek dėl aplaidumo, tiek tyčia, siekiant finansinės naudos arba keršto.
- Pasekmės:
 - Reputacijos žala: klientų pasitikėjimo sumažėjimas, dėl ko įmonė gali prarasti dalį savo klientų.
 - Reguliacinių reikalavimų pažeidimai: GDPR taisyklių pažeidimas dėl klientų duomenų nutekėjimo gali sukelti dideles baudas.

- Finansiniai nuostoliai: susiję tiek su reguliacinių baudų dydžiu, tiek su sumažėjusiomis pajamomis dėl klientų praradimo.
- Pažeidžiami ištekliai: klientų finansiniai įrašai, patentinių algoritmų duomenys, kriptovaliutų transakcijų duomenys.
- Scenarijaus pagrindimas generatoriaus vidine duomenų baze:
 - ISO 31000 (ISO 31000, 2018) rekomenduoja reguliarias rizikos vertinimo priemones, siekiant nustatyti vidines grėsmes.
 - ENISA (ENISA, 2020) akcentuoja griežtų prieigos kontrolės mechanizmų ir darbuotojų stebėsenos svarbą.
 - GDPR (Europos parlamentas ir Europos sąjungos taryba, 2016) reikalauja užtikrinti tinkamą asmens duomenų apsaugą, siekiant išvengti pažeidimų.
- Nauda įmonei: scenarijus leidžia tobulinti vidinių grėsmių valdymo politiką ir sustiprinti darbuotojų atsakomybę, mažinant vidinių incidentų riziką.

2 scenarijus. Išpirkos reikalaujančios programos ataka (angl. Ransomware attack) prieš kritines finansų sistemas.

- Aprašymas: kibernetiniai nusikaltėliai, naudodami socialinės inžinerijos technikas (pvz., phishing ataka), platina kenkėjiškas programas, kurios šifruoja įmonės duomenis ir reikalauja išpirkos už jų atkūrimą.
- Pasekmės:
 - Operacijų sustabdymas: dėl pagrindinių finansinių sistemų užšifravimo įmonė gali prarasti gebėjimą vykdyti transakcijas.
 - Finansiniai nuostoliai: susiję su tiek operacijų sustabdymu, tiek išpirkos mokėjimu.
 - Duomenų praradimas: jei duomenų atsarginės kopijos nepakankamos, įmonė gali prarasti kritinius duomenis.
- Pažeidžiami ištekliai: pagrindinės finansinės sistemos, klientų įrašai ir transakcijų duomenys.
- Scenarijaus pagrindimas generatoriaus vidine duomenų baze:
 - NIST (NIST, 2021d) rekomenduoja naudoti incidentų valdymo planus ir grėsmių aptikimo sistemas, kad būtų galima užkirsti kelią tokioms atakoms kaip, išpirkos reikalaujančių programų.
 - ENISA (ENISA, 2020) akcentuoja atsarginių duomenų kopijų svarbą ir grėsmių žvalgybos naudojimą siekiant anksti nustatyti tokias kaip phishing atakas.
- Nauda įmonei: šis scenarijus skatina reguliariai tikrinti pasirengimą kritinėms situacijoms, įskaitant atsarginių kopijų efektyvumą ir atsako į grėsmes greitį.

3 scenarijus. Trečiosios šalies pardavėjo saugumo pažeidimas.

- Aprašymas: išorinis veikėjas išnaudoja trečiosios šalies (tiekėjo) sistemų pažeidžiamumus, kad gautų prieigą prie įmonės dalijamų jautrių duomenų.
- Pasekmės:
 - Duomenų pažeidimai: kritinių klientų duomenų nutekėjimas per tiekėjų sistemas.
 - Reguliacinės bausdės: dėl GDPR reikalavimų pažeidimų.
 - Reputacijos žala: pasitikėjimo įmonės partneriais sumažėjimas.

- Pažeidžiami ištekliai: duomenys, kuriais įmonė dalijasi su tiekėjais, ir išorinės paslaugų sistemos.
- Scenarijaus pagrindimas generatoriaus vidine duomenų baze:
 - DORA reglamentuoja trečiųjų šalių rizikos saugumo valdymą.
 - ISO 31000 (ISO 31000, 2018) akcentuoja periodinę peržiūrą organizacijų (pavyzdžiui, tiekėjų) veiklos valdymo ir ataskaitų teikimo.
 - GDPR (Europos parlamentas ir Europos sąjungos taryba, 2016) reikalauja užtikrinti, kad tiekėjai laikytųsi tinkamų duomenų apsaugos priemonių.
- Nauda įmonei: scenarijus leidžia stiprinti bendradarbiavimo su tiekėjais kontrolę ir sumažinti priklausomybę nuo nepatikimų paslaugų.

4 scenarijus. Automatizuotų grėsmių aptikimo sistemų gedimas.

- Aprašymas: dėl sistemos klaidų arba netinkamos konfigūracijos nepastebimos kritinės grėsmės, tokios kaip anomalijos ar kenkėjiškos programos.
- Pasekmės:
 - Ilgalaikis pažeidžiamumas ir nepastebėtos atakos.
 - Finansinės apgaulės ir duomenų kompromitavimas.
- Pažeidžiami ištekliai: klientų duomenys (sąskaitos) ir vidinės duomenų bazės.
- Scenarijaus pagrindimas generatoriaus vidine duomenų baze:
 - ENISA (ENISA, 2020) rekomenduoja automatizuotų sistemų reguliarius auditus, taip pat akcentuoja atsarginių kopijų svarbą aptikimo sistemoms.
- Nauda įmonei: šis scenarijus pabrėžia poreikį dažniau tikrinti sistemas ir taikyti rankinius patikrinimus.

5 scenarijus. DI sistemų manipuliacija finansinėms prognozėms.

- Aprašymas: išmanūs įsilaužėliai išnaudoja DI modelių spragas, siekdami manipuluoti finansinėmis prognozėmis (vertybinių popierių prekyba) ar rinkos duomenimis. Arba pradėti prekiauti viešai neatskleista informacija.
- Pasekmės:
 - Finansiniai nuostoliai.
 - Rinkos destabilizavimas.
- Pažeidžiami ištekliai: patentuoti DI modeliai ir finansiniai duomenys.
- Scenarijaus pagrindimas generatoriaus vidine duomenų baze:
 - ISO/IEC 23894 (ISO/IEC 23894. Information Technology-Artificial Intelligence-Guidance on Risk Management, 2023) rekomenduoja naudoti DI modelių rizikos valdymo sistemas.
 - ENISA (ENISA, 2020) pabrėžia DI sistemų stebėsenos svarbą.
- Nauda įmonei: skatina diegti DI modelių rizikos valdymo sistemas, taip pat bendradarbiauti su pramonės kolegomis, kad nustatytumėte kylančias DI grėsmes.

Taigi, ketvirto eksperimento metu sugeneruoti scenarijai parodė, kad kriptoturto rinkos hipotetinė įmonė susiduria su itin įvairialypėmis kibernetinėmis grėsmėmis, apimančiomis tiek vidinius, tiek išorinius rizikos šaltinius. Kur eksperimento rezultatai (sugeneruoti scenarijai) parodė,

kad kriptoturto rinkos hipotetinė įmonė susiduria su specifiniais sektoriaus iššūkiais, įskaitant jautrių duomenų apsaugą, blokų grandinės technologijų saugumą ir DI algoritmų vientisumą. Generuojant scenarijus buvo atsižvelgta į įmonės veiklos pobūdį, naudojamas technologijas ir svarbą užtikrinti ne tik reguliacinę atitiktį (GDPR, DORA), bet ir klientų pasitikėjimą. Eksperimente taip pat išryškėjo svarbiausi veiksniai, kurie galėtų sustiprinti įmonės kibernetinį atsparumą, tokie kaip reguliarios darbuotojų mokymo programos, trečiųjų šalių rizikos valdymas, DI rizikos stebėsena, pabrėžtas duomenų atsarginių kopijų bei grėsmių žvalgybos platformų naudojimo poreikis. Ir visi sugeneruoti scenarijai buvo pagrįsti tarptautiniais standartais (ISO, NIST, GDPR, DORA, ENISA), kurie suteikė pagrindą logiškai ir struktūruotai analizei.

5.3. Eksperimentas su kibernetinio saugumo ekspertais ir hipotetinėmis įmonėmis

Šio tyrimo metu buvo siekiama atlikti palyginimą tarp generatoriaus sugeneruotų scenarijų kibernetinio saugumo rizikos scenarijų ir kibernetinio saugumo specialistų. Buvo paprašyta dviejų kibernetinio saugumo specialistų (NATO kibernetinio saugumo inžinieriaus ir Nacionalinio kibernetinio saugumo centro Tarptautinio kibernetinių grėsmių analizės skyriaus vedėjo) sukurti tris labiausiai tikėtinus kibernetinio saugumo scenarijus pagal pateiktų hipotetinių įmonių aprašymus ir suformuluotas užklausas. Šio eksperimento tikslas buvo nustatyti, ar generatorius gali pateikti tikslesnius, greičiau sugeneruojamus ir praktiškai pritaikomus rizikos scenarijus lyginant su specialistų sukurtais rankiniu būdu.

5.3.1. Eksperimentas su kredito ir mokėjimų rinkos hipotetine įmone

Šio eksperimento tikslas – įvertinti generatoriaus ir žmogaus sukurtų kibernetinio saugumo rizikos scenarijų skirtumus bei efektyvumą, analizuojant kredito ir mokėjimų rinkos hipotetinės įmonės atvejį. Palyginimui buvo pasitelktas sukurtas generatorius ir NATO kibernetinio saugumo inžinierius, kuris rankiniu būdu kūrė scenarijus pagal pateiktą informaciją. Ši analizė siekia parodyti, kaip generatorius gali supaprastinti ir palengvinti kibernetinio saugumo rizikos scenarijų kūrimo procesą, pagrįstą tarptautiniais standartais (ISO, NIST, GDPR, ENISA, DORA).

Kibernetinio saugumo inžinieriui buvo pateikta informacija:

- Hipotetinės įmonės aprašymas. Mažo dydžio finansinė įstaiga, orientuota į vartotojų ir smulkių verslo mokėjimų sprendimus, kredito paskolas bei mokėjimo kortelių paslaugas. Pagrindinė veikla: paskolų valdymas, mokėjimų apdorojimas ir sukčiavimo prevencija. IT infrastruktūra apima debesų paslaugas (AWS, Azure), mašininio mokymosi sistemas (XGBoost, Random Forest), vietinius serverius bei centralizuotas duomenų bazines. Saugumo politika: TLS 1.3 šifravimas, saugumo auditai ir anomalijų aptikimo sistemos.

- Scenarijaus struktūra. Specialistui buvo pateikta aiški struktūra, kokie komponentai turi būti įtraukti į kiekvieną scenarijų:
 - Aktorius: Kas vykdo grėsmę (vidinis/ išorinis veikėjas).
 - Grėsmės įvykis: Koks konkretus grėsmės įvykis vyksta.
 - Motyvai: Tikėtini veikėjų tikslai (pvz., finansinė nauda, sabotžas).
 - Poveikis: Kokie nuostoliai ar žala kyla įmonei (finansai, reputacija, atitikties pažeidimai).
 - Turtas: Kas gali būti pažeista (duomenys, sistemos)
 - Laikas: Kada grėsmė greičiausiai gali būti įgyvendinta.
- Pateikti dokumentai, kuriais gali remtis: ISO 31000, NIST, GDPR, DORA, ENISA.
- Generatoriaus užklausa. Specialistas gavo užduotį, kuri analogiškai buvo pateikta generatoriui:

"Generate scenarios. This is a small financial institution in the credit and payment processing industry with 45 employees. The company prioritizes compliance with GDPR and PCI DSS, protects sensitive customer data using TLS 1.3 encryption, anomaly detection systems, and cloud services like AWS and Azure, while addressing risks such as fraud, data breaches, and potential DDoS attacks. "

Palyginimas atliktas remiantis trimis labiausiai tikėtiniais scenarijais, kuriuos sugeneravo generatorius (9 priedas) ir rankiniu būdu sukūrė kibernetinio saugumo specialistas (13 priedas).

4 lentelė. Palyginimas tarp generatoriaus ir kibernetinio saugumo specialisto sukurtų scenarijų kredito ir mokėjimų rinkos hipotetinei įmonei

Kriterijus	Generatorius	Specialistas
Scenarijus 1	Phishing Attack Leading to Credential Theft	Duomenų vagystė per debesijos paslaugų pažeidimą
Atakos pobūdis	Išorinis nusikaltėlis vykdo phishing ataką, siekdamas pavogti prisijungimo duomenis.	Išorinis nusikaltėlis naudoja debesijos spragomis arba phishing metodais duomenų vagystei įgyvendinti.
Poveikis	Prisijungimo duomenų vagystė lemia klientų mokėjimo duomenų sugadinimą, finansinius nuostolius ir GDPR pažeidimą.	Nutekėję klientų duomenys sukelia reputacijos žalą ir baudas pagal DORA ir GDPR reglamentus.
Panašumai	Abu scenarijai nurodo phishing atakas kaip pagrindinį įrankį pavogti duomenis.	
Scenarijus 2	Insider Threat Resulting in Data Exfiltration	„Ransomware“ ataka prieš vietinius serverius
Atakos pobūdis	Vidinis darbuotojas neteisėtai prieina prie duomenų ir juos išfiltruoja dėl keršto ar finansinės naudos.	Išorinis nusikaltėlis vykdo „Ransomware“ ataką, užšifruodamas vietinius serverius.
Poveikis	Jautrių duomenų vagystė lemia reputacinę žalą ir GDPR pažeidimus.	Sistemos užšifravimas sukelia operacijų sustabdymą ir finansinius nuostolius dėl išpirkos mokėjimo.
Panašumai	Abu scenarijai apima kritinių duomenų pažeidžiamumą, nors išpildymas scenarijų skiriasi.	

4 lentelės tęsinys

Kriterijus	Generatorius	Specialistas
Scenarijus 3	Distributed Denial of Service (DDoS) Attack	Sukčiavimo ataka per mokėjimų apdorojimo sistemas
Atakos pobūdis	Išorinis oponentas (haktivistai ar konkurentai) vykdo DDoS ataką, kad sutrikdytų mokėjimų sistemas.	Automatizuotas sukčiavimas, siekiant išnaudoti mokėjimų apdorojimo algoritmų spragas.
Poveikis	Paslaugų sutrikimas, prarastos transakcijos ir klientų pasitenkinimo sumažėjimas.	Netikrų mokėjimų vykdymas ar kortelių duomenų vagystė lemia finansinius nuostolius ir DORA sankcijas.
Panašumai	Abu scenarijai buvo pateikti su mokėjimų apdorojimo sistemų veiklos sutrikdymu, tik skirtingomis atakomis.	

Šaltinis: sudaryta autorės

Taigi, generatoriaus sugeneruoti scenarijai įrodo savo tikslingumą ir naudą hipotetinei kredito ir mokėjimų rinkos įmonei, lyginant juos su kibernetinio saugumo specialisto sukurtais alternatyvomis. Pagal atliktą palyginimą matyti, kad generatoriaus scenarijai ne tik atspindi aktualias grėsmes, bet ir pateikia jas aiškiai struktūruotas, orientuotas į įmonės veiklos specifiką. Naudojantis kibernetinio saugumo specialisto sugeneruotais scenarijais kaip pagrindu, galima patvirtinti, kad generatoriaus pateikti rezultatai yra teisingi ir praktiškai naudingi (4 lentelė). Generatoriaus scenarijų atitikimas specialisto pateiktiems leidžia įvertinti jų teisingumą, remiantis šiais aspektais:

1. Abu pateikė „Phishing“ ataką kaip kritinę riziką finansų įstaigoms. Generatorius šią grėsmę papildė tikslesniu poveikio ir grėsmės įvykio aprašymu, įskaitant darbuotojų prisijungimų duomenų vagystę ir jos įtaką mokėjimų sistemoms.
2. Duomenų vagystė pasikartoja tiek generatoriaus, tiek specialisto 2 scenarijuje, tačiau generatorius išskyrė vidinio darbuotojo riziką, o specialistas – debesijos infrastruktūros spragas. Tai parodo, kad generatorius geba apimti ir vidines, ir išorines grėsmes.
3. 3 scenarijuje abu pateikė grėsmes mokėjimų sistemoms, tačiau generatorius įtraukė DDoS ataką, kuri išplėtė rizikos suvokimą mokėjimų sistemų veiklos tęstinumui.

Lyginant su specialisto sukurtai scenarijais, generatorius turi keletą reikšmingų pranašumų:

- Generatorius scenarijus sugeneruoja per kelias sekundes, o specialistui sugeneruoti tris scenarijus prireikė apie 7,5 valandų. Tai parodo privalumą, kur įmonės gali greitai reaguoti į grėsmes, kai laikas yra kritinis veiksnys.
- Generatorius yra ekonomiškai efektyvesnis, nes nereikalauja brangių specialistų darbo valandų ir papildomo laiko duomenų analizei (užtenka išsipirkti ChatGPT planą, kurio mėnesinė kaina siekia 20 USD).

Apibendrinant, galima teigti, kad naudodamas mažiau resursų ir generuodamas rezultatus greičiau, generatorius gali būti idealus įrankis organizacijoms, siekiančioms greitai ir efektyviai identifikuoti kibernetines grėsmes. Tai ypač svarbu finansų sektoriuje, kur tikslumas, laikas ir ekonomiškumas yra kritiniai veiksniai užtikrinant veiklos tęstinumą ir atitikimą reguliaciniams reikalavimams.

5.3.2. Eksperimentas su draudimo rinkos hipotetine įmone

Šio eksperimento tikslas buvo padaryti antrą palyginimą sukurtų scenarijų generatoriaus ir kibernetinio saugumo specialisto su hipotetine draudimo įmone. Šiame eksperimente buvo paprašyta Nacionalinio kibernetinio saugumo centro Tarptautinio kibernetinių grėsmių analizės skyriaus vedėjo sukurti kibernetinio saugumo rizikos scenarijus pagal pateiktą hipotetinės įmonės aprašymą ir užklausą.

Kibernetinio saugumo inžinieriui buvo pateikta informacija:

- Hipotetinės įmonės aprašymas. Vidutinio dydžio draudimo įmonė, veikianti gyvybės, sveikatos, automobilių ir nekilnojamojo turto draudimo srityse. IT infrastruktūra apima debesijos integraciją, RAID technologijas ir pažangias duomenų analizės platformas, tokias kaip Guidewire ir DataRobot. Visi įmonės draudimo duomenys apsaugoti AES-256 šifravimu. Taip pat yra įdiegta dviejų faktorių autentifikacija ir kasmet vykdomi kibernetinio saugumo mokymai darbuotojams.
- Scenarijų struktūra. Specialistui buvo pateikta aiški struktūra, kokie komponentai turi būti įtraukti į kiekvieną scenarijų: Grėsmių aktoriai, įvykiai, motyvai, poveikis, turtas ir laikas.
- Pateikti dokumentai, kuriais gali remtis: ISO 31000, NIST, GDPR, DORA, ENISA.
- Generatoriaus užklausa. Specialistas gavo užduotį, kuri analogiškai buvo pateikta generatoriui:

„Generate scenarios. Industry: Insurance industry specializing in life, health, automobile, and property insurance. Size: Medium-sized organization with over 120 employees and more than 50,000 clients. Cybersecurity solutions: Guidewire platform for automated insurance operations and customer data management. Cloud-based infrastructure ensuring flexible and accessible data processing. RAID storage systems for secure and redundant data storage. AI-driven platforms like DataRobot for risk evaluation and pricing. AES-256 encryption and multi-factor authentication for securing sensitive data. Attack concerns: Ransomware targeting customer and operational data. Phishing targeting employees or customers. Insider threats impacting sensitive information. Vulnerabilities in cloud integration systems. Regulatory or compliance requirements: Compliance with GDPR and regular cybersecurity resilience testing.“

Palyginimas atliktas remiantis trimis labiausiai tikėtiniais scenarijais, kuriuos sugeneravo generatorius (10 priedas) ir rankiniu būdu sukūrė kibernetinio saugumo specialistas (14 priedas).

5 lentelė. Palyginimas tarp generatoriaus ir kibernetinio saugumo specialisto sukurtų scenarijų draudimo rinkos hipotetinei įmonei

Kriterijus	Generatorius	Specialistas
Scenarijus 1	Ransomware ataka prieš kliento ir veiklos duomenis	Phishing ataka prieš darbuotojus
Atakos pobūdis	Organizacija užkrečiama „ransomware“, naudojant sukčiavimo el. laiškus darbuotojams.	Socialinės inžinerijos pagrindu vykdomos phishing atakos, siekiant pavogti darbuotojų prisijungimo duomenis.
Poveikis	Duomenų užšifravimas, veiklos sutrikdymas ir reputacijos žala.	Jautrių klientų duomenų vagystė, sukčiavimas žalų administravimo sistemose.
Panašumai	Abu scenarijai išryškina phishing kaip grėsmės vektorių, vedantį prie jautrių duomenų pažeidimų ir finansinės rizikos.	
Scenarijus 2	Viešai neatskleista grėsmė, sukelianti duomenų pažeidimą	Paslaugų nepasiekiamumas dėl DDoS atakos
Atakos pobūdis	Vidinis darbuotojas išfiltruoja klientų duomenis per USB laikmenas ar debesijos platformas.	Kibernetinis nusikaltėlis vykdo DDoS ataką, trikdydamas įmonės veiklą.
Poveikis	GDPR pažeidimas, klientų pasitikėjimo mažėjimas, teisiniai veiksmai.	Paslaugų nepasiekiamumas, klientų pasipiktinimas, finansiniai nuostoliai dėl veiklos stabdymo.
Panašumai	Abu scenarijai pabrėžia veiklos sutrikdymą ir duomenų jautrumą, tačiau iš skirtingų grėsmės vektorių (vidinis darbuotojas ir išorinis veikėjas).	
Scenarijus 3	Debesų integracijos pažeidžiamumo išnaudojimas	„Brute-force“ atakos prieš paskyras
Atakos pobūdis	Išoriniai įsilaužėliai išnaudoja API saugumo spragas debesijos platformose.	Automatizuotas slaptažodžių spėliojimas siekiant prisijungti prie darbuotojų paskyrų.
Poveikis	Duomenų nutekėjimas, operaciniai sutrikimai, reguliacinė kontrolė.	Neautorizuota prieiga prie draudimo sistemų, galimi tolimesni išpuoliai, pvz., ransomware įdiegimas.
Panašumai	Abu scenarijai atkreipia dėmesį į autentifikavimo ir API saugumo pažeidimus, leidžiančius prieigą prie jautrių sistemų.	

Šaltinis: sudaryta autorės

Taigi, palyginus generatoriaus sugeneruotus scenarijus ir specialisto sukurtus scenarijus, matyti, kad generatorius tinkamai generuoja scenarijus, kurie yra tikslūs, realūs ir panašūs į specialisto suformuluotus:

1. „Phishing“ ataka. Abu scenarijai (generatoriaus ir specialisto) identifikavo „phishing“ atakas kaip vieną pagrindinių rizikų draudimo bendrovėse. Specialistas apibūdino „phishing“ kaip įrankį gauti prieigą prie jautrių duomenų, tuo tarpu generatorius papildomai numatė „ransomware“ įdiegimo galimybę, pasinaudojant „phishing“ el. laiškais, taip praplėsdamas

- scenarijaus apimtį. Tai rodo, kad generatorius ne tik identifikuoja realias grėsmes, bet ir logiškai sujungia galimas jų pasekmes, sukuriant platesnį rizikos kontekstą.
2. Vidinė darbuotojų grėsmė (Insider Threat). Tiek generatoriaus, tiek specialisto scenarijai pabrėžė vidinių darbuotojų veiklą kaip rizikos šaltinį. Specialistas išskyrė „brute-force“ atakas prieš darbuotojų paskyras, o generatorius numatė duomenų nutekėjimą per darbuotojų privilegijuotą prieigą ar debesijos platformas. Tai parodo, kad generatorius geba įvertinti vidines grėsmes ir jų poveikį, kuris yra itin aktualus draudimo sektoriuje, kur jautrūs klientų duomenys yra pagrindinis turtas.
 3. Debesijos infrastruktūros pažeidžiamumai. Abu scenarijai identifikavo grėsmes debesijos sistemoms, nors jų pobūdis skyrėsi. Specialistas aprašė brute-force atakas kaip grėsmę prieigai prie darbuotojų paskyrų, o generatorius išskyrė API pažeidžiamumų išnaudojimą kaip svarbų grėsmės vektorius. Šis skirtumas parodo, kad generatorius geba tinkamai numatyti pažangias grėsmes, tokias kaip API pažeidimai, kurie yra dažni moderniose debesijos integracijose. Tai praturtina rizikos suvokimą ir užtikrina geresnį pasirengimą.

Apibendrinant, galima teigti, kad generatoriaus sukurti scenarijai yra tikslingi, atitinkantys draudimo įmonės veiklos specifiką. Lyginant su kibernetinio saugumo specialisto ranka parengtais scenarijais, generatorius užtikrina greitesnį (generatorius per kelias sekundes sugeneravo scenarijus, specialistas scenarijus parengė per 6val.), ekonomišką ir tikslesnį kibernetinių grėsmių identifikavimą. Specialisto sukurti scenarijai taip pat patvirtina, kad generatorius sukuria tikslingus ir praktinius scenarijus, kurie gali būti naudojami organizacijose siekiant stiprinti kibernetinį atsparumą.

5.4. Apibendrinimas

Eksperimento tikslas buvo įvertinti sukurto kibernetinio saugumo rizikos scenarijų generatoriaus veiksmingumą finansų sektoriaus hipotetinėse įmonėse, atsižvelgiant į rizikos scenarijų identifikavimą, aktualumo nustatymą, pažeidžiamumo analizę ir scenarijų pritaikymą praktikoje. Atlikus keturis eksperimentus su kibernetinio saugumo rizikos scenarijų generatoriumi ir papildomai palyginus du eksperimentus tarp generatoriaus ir kibernetinio saugumo specialistų, buvo gauti tokie eksperimentų rezultatai:

- Eksperimentas parodė, kad generatorius geba generuoti tikslus ir specifinius kibernetinio saugumo rizikos scenarijus, atsižvelgiant į kiekvienos hipotetinės įmonės IT infrastruktūros ypatumus, naudojamas technologijas (pvz., DI sistemas, debesijos sprendimus) bei veiklos sritis. Kiekvieno eksperimento metu buvo sugeneruoti 5 individualizuoti scenarijai, kurie aiškiai identifikavo tiek vidines (darbuotojų klaidos, prieigos kontrolės trūkumai), tiek išorines grėsmes (kibernetinės atakos, tiekimo grandinės pažeidimai). Tai patvirtina generatoriaus gebėjimą struktūruotai ir detalai nustatyti rizikas.
- Generatorius sugebėjo per trumpą laiką sukurti tikslus ir realistiškus scenarijus, kurie atitiko realius finansų sektoriaus iššūkius ir buvo pagrįsti dokumentacijomis, tokiomis kaip: ISO, GDPR, NIST, ENISA, DORA. Pateikti scenarijai tokie kaip DDoS atakos prieš mokėjimų

sistemas, ransomware atakos prieš kritines finansines sistemas ar DI modelių manipuliacijos, yra aktualūs tiek dabartinių kibernetinių grėsmių kontekste, tiek finansinių technologijų sektoriaus dinamikai. Ši analizė parodė, kad generatorius sėkmingai atspindi realius kibernetinio saugumo iššūkius ir galimas grėsmių pasekmes.

- Eksperimente nustatytos pagrindinės pažeidžiamumo sritys: technologinės spragos (tokios kaip: debesų kompiuterijos pažeidžiamumai, API saugumo trūkumai, DI algoritmų manipuliavimas), vidinės rizikos (darbuotojų klaidos, piktybiniai ketinimai, prieigos kontrolės trūkumai), išorinės grėsmės (DDoS atakos, phishing kampanijos, ransomware ir tiekimo grandinės saugumo pažeidimai). Šios pažeidžiamumo sritys parodė, kad kiekviena finansų sektoriaus įmonė turi savo specifines pažeidžiamumo sritis, tačiau bendra tendencija rodo būtinybę užtikrinti duomenų apsaugą, stiprinti darbuotojų prieigos kontrolę bei reguliariai tikrinti naudojamų technologijų saugumą.
- Sugeneruoti scenarijai buvo ne tik detalūs, bet ir praktiškai pritaikomi. Juose buvo pateiktos rekomendacijos, kaip būtų galima sustiprinti kibernetinį atsparumą, pvz.: tinklo apsaugos stiprinime (DDoS apsaugos priemonės, tinklo stebėjimo sistemų patobulinimai), vidinių rizikų valdyme (darbuotojų prieigos kontrolės politika, veiksmų stebėsena, reguliarios mokymo programos), technologinės rizikos valdyme (DI modelių testavimas, debesijos saugumo auditai ir tiekėjų valdymas). Tai rodo, kad generatorius ne tik identifikuoja rizikas, bet ir pasiūlo praktines priemones, kurios gali būti įgyvendintos realiose finansų sektoriaus organizacijose.
- Generatoriaus sugeneruoti scenarijai buvo ne tik panašūs į kibernetinio saugumo specialistų pateiktus rezultatus, bet kai kuriais atvejais pasižymėjo išsamesne rizikų analize.
- Generatorius sugebėjo sukurti scenarijus per kelias sekundes, kai tuo tarpu specialistams reikėjo 6-8 valandų. Tai rodo reikšmingą generatoriaus pranašumą tiek laiko, tiek kaštų požiūriu.
- Eksperimentai atskleidė, kad generatorius gali būti vertingas įrankis finansų įmonėse, siekiančioms greitai ir efektyviai identifikuoti kibernetines grėsmes.

Eksperimento rezultatai patvirtino, kad kibernetinio saugumo rizikos scenarijų generatorius yra efektyvus įrankis finansų įstaigoms, skirtas:

- identifikuoti kibernetines grėsmes ir pažeidžiamumus;
- modeliuoti realistiškus ir aktualius grėsmių scenarijus;
- siūlyti praktines priemones rizikos valdymui.

Apibendrinant galima teigti, kad kibernetinio saugumo rizikos scenarijų generatorius yra tikslus, greitas ir ekonomiškai efektyvus sprendimas finansų organizacijoms, siekiančioms stiprinti savo kibernetinį atsparumą. Generatoriaus taikymas gali padėti finansų sektoriaus įmonėms ne tik geriau suprasti savo pažeidžiamumą, bet ir sukurti tikslingas strategijas kibernetiniam saugumui stiprinti. Be to, rezultatai pabrėžė, kad skirtingų finansų įmonių IT infrastruktūros ir DI naudojimo lygiai turi reikšmingą įtaką jų kibernetinių grėsmių spektrui bei atsparumui. Eksperimentai taip pat parodė, kad generatorius sugeba sukurti realistiškus ir praktiškai pritaikomus scenarijus, kurie priylgsta kibernetinio saugumo specialistų sukurtiems scenarijams. Taigi, šis įrankis gali būti toliau plėtojamas ir pritaikomas praktiniam naudojimui finansų sektoriuje.

IŠVADOS

- Išanalizavus dirbtinio intelekto sistemų sampratą bei jų pritaikymo galimybes finansų įstaigose, nustatyta, kad DI technologijos atlieka svarbų vaidmenį automatizuojant procesus, mažinant veiklos sąnaudas ir didinant sprendimų priėmimo tikslumą. Rezultatai parodė, kad DI sistemų integracija į finansinius procesus gali žymiai padidinti efektyvumą ir operacijų tikslumą. Tačiau augant šių sistemų svarbai, identifikuoti ir nauji kibernetinio saugumo iššūkiai, kuriuos būtina spręsti siekiant užtikrinti saugų DI sistemų naudojimą bei apsaugoti jautrius vartotojų duomenis. Tai sudaro pagrindą tolesniems tyrimams apie rizikos valdymo metodus DI sistemose.
- Išnagrinėta reguliacinė aplinka, kibernetinio saugumo reikalavimai ir rizikos valdymo standartai parodė, kad efektyviam DI rizikos scenarijų kūrimui būtina aiškiai apibrėžti reguliavimo gaires ir saugumo principus. Nustatyta, kad rizikos scenarijams keliami reikalavimai turi būti visapusiški, apimantys tiek techninius, tiek teisinius aspektus, kad sukurtos sistemos galėtų atitikti tiek vidaus, tiek išorės saugumo standartus. Pasiektas rezultatas suteikė pagrindą DI rizikos scenarijų generavimui, leidžiant modeliuoti realistiškus ir pritaikomus kibernetinio saugumo scenarijus finansų įstaigoms.
- Išanalizavus rizikos scenarijų sudedamąsias dalis ir jų elementų taksonomijas, buvo sukurta struktūruota metodologija, leidžianti nuosekliai identifikuoti galimas grėsmes ir jų poveikį. Atliktas generatyvinių dirbtinio intelekto įrankių, tokių kaip ChatGPT-4, Google Gemini 1.5 Flash, IBM Watson AI ir Microsoft Copilot, palyginimas parodė, kad ChatGPT-4 išsiskyrė kaip pažangiausias įrankis, kuris pasižymi aukštu kalbos supratimo tikslumu, ilgo konteksto apdorojimo galimybe ir galimybe efektyviai spręsti kompleksines užduotis, dėl to ChatGPT-4 buvo pasirinktas tolimesniam praktiniam darbui, kuriant finansų įstaigoms kibernetinio saugumo rizikos scenarijų generatorių. Sukurta metodologija suteikė aiškius žingsnius ir principus, užtikrinančius, kad generuojami rizikos scenarijai būtų struktūruoti, praktiškai pritaikomi finansų įstaigoms ir tinkami kibernetinio saugumo grėsmių analizei.
- Sukūrus dirbtinio intelekto sistemų rizikos scenarijų generatorių, buvo išanalizuotas jo kūrimo procesas, apmokymo etapai ir vartotojo sąsaja. Tyrimas parodė, kad naudojant generatyvinį GPT-4 (OpenAI) modelį galima efektyviai kurti realistiškus ir pritaikomus rizikos scenarijus, atitinkančius finansų įstaigų kibernetinio saugumo poreikius. Generatorius užtikrina saugių duomenų naudojimą, privatumą ir adaptaciją vartotojų specifiniams reikalavimams. Šis įgyvendinimas parodė, kad nuosekliai išvystytas generatorius gali būti praktiškai taikomas kaip įrankis identifikuoti ir valdyti potencialias grėsmes, taip didinant finansų įstaigų pasirengimą kibernetinio saugumo iššūkiams.
- Atlikti eksperimentai su sukurtu rizikos scenarijų generatoriumi parodė jo pritaikomumą skirtingų finansų rinkų hipotetinėse įmonėse: kreditų ir mokėjimų, draudimo, finansinių priemonių bei kriptoturto srityse. Tyrimas atskleidė, kad DI pagrindu sukurtas generatorius geba identifikuoti rizikos veiksnius ir modeliuoti realistiškus scenarijus, pritaikytus specifiniams finansų įmonių poreikiams. Papildomai atlikti eksperimentai su kibernetinio saugumo specialistais leido įvertinti generatoriaus tikslumą, pritaikomumą ir praktinį

naudingumą. Ekspertų atliktos užduotys patvirtino, kad sukurtas įrankis yra vertingas finansų įstaigų kibernetinio saugumo grėsmių valdymo procese, suteikdamas galimybę modeliuoti įvairiapusiškas grėsmes ir efektyviai reaguoti į potencialius pavojus.

LITERATŪROS ŠALTINIAI

1. Ai Hleg. (2019). A Definition of AI: Main Capabilities and Disciplines. *European Commission*.
2. Aleksas MakFarlandas. (2024). 10 geriausių dirbtinio intelekto kibernetinio saugumo įrankių (2024 m. gruodžio mėn.). https://Www.Unite.Ai/Lt/Ai-Kibernetinio-Saugumo-%C4%AFrankiai/?Utm_source.
3. ANA HADNES BRUDER, DR. ULRICH WORM, BENJAMIN BECK, OLIVER YAROS, AMELIE KSINSIK, & MICHELLE MAYER. (2024). *NEW EU CYBER RULES: IMPLEMENTATION OF NIS2 IN THE EU MEMBER STATES*.
4. Ayasdi. (2021). *Ayasdi-Discovering the Whole Truth*.
5. Bahaaeddin A. M. Alareeni, & Islam Elgedawy. (2023). *Artificial Intelligence (AI) and Finance* (B. A. M. Alareeni & I. Elgedawy, Eds.; Vol. 488). Springer Nature Switzerland. <https://doi.org/10.1007/978-3-031-39158-3>
6. BCG. (2023). *Generative AI in the Finance Function of the Future*.
7. Benjamin Christopher Ilg. (2024). AI in finance. https://Finance.Ec.Europa.Eu/News/Ai-Finance-2024-06-19_en.
8. Bhattacharya, P., Prasad, V. K., Verma, A., Gupta, D., Sapsomboon, A., Viriyasitavat, W., & Dhiman, G. (2024). Demystifying ChatGPT: An In-depth Survey of OpenAI's Robust Large Language Models. *Archives of Computational Methods in Engineering*. <https://doi.org/10.1007/s11831-024-10115-5>
9. BlackRock. (2024). Aladdin platform. <https://Www.Blackrock.Com/Aladdin>.
10. Cao, L. (2020). *AI in Finance: A Review*. <https://doi.org/>
11. Casey, T., & Rosenquist, M. (2016). *Understanding Cyberthreat Motivations to Improve Defense*. <https://www.researchgate.net/publication/364960481>
12. Chainalysis. (2024). Optimize crypto exchange operations through data insights. <https://Www.Chainalysis.Com/Centralized-Exchanges/>.
13. Cisco Press. (2017). Securing IoT. <https://Www.Ciscopress.Com/Articles/Printerfriendly/2803867>.
14. ComplyAdvantage. (2024). ComplyAdvantage Mesh. <https://Complyadvantage.Com/Mesh/>.
15. DAIR.AI. (2024). Prompt Engineering Guide. <https://Www.Promptingguide.Ai/>.
16. Darktrace. (2024). The essential cybersecurity platform. <https://Darktrace.Com/Platform>.
17. DataRobot. (2024). AI for financial services. <https://Www.Datarobot.Com/Solutions/Financial-Services/>.
18. De Silva, D., & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. *Patterns*, 3(6). <https://doi.org/10.1016/j.patter.2022.100489>
19. Deloitte. (2024a). *AI and risk management*.

20. Deloitte. (2024b). Managing model and AI risks in the investment management sector. <https://www2.deloitte.com/content/dam/deloitte/us/documents/advisory/us-advisory-deloitte-managing-model-and-ai-risks-in-the-investment-management-sector-july-2023.pdf>.
21. Deloitte. (2024c). Redefining Third Party Management with Gen AI. <https://www.deloitte.com/global/en/services/risk-advisory/blogs/defining-the-use-of-third-party-management-with-gen-ai.html>.
22. Dr. Emmanuel Tsukerman. (2020). Cybersecurity Threat Modeling with OCTAVE. https://www.pluralsight.com/resources/blog/guides/cybersecurity-threat-modeling-with-octave?utm_source.
23. Dr Florian Ostmann, & Dr Cosmina Dorobantu. (2021). *AI in Financial Services*.
24. Ellerhold, C., Schnagl, J., & Schreck, T. (2023). Enterprise Cyber Threat Modeling and Simulation of Loss Events for Cyber Risk Quantification. *CCSW 2023 - Proceedings of the 2023 Cloud Computing Security Workshop*.
<https://doi.org/10.1145/3605763.3625244>
25. Elliptic. (2024). Know exactly what happens on any blockchain. <https://www.elliptic.co/>.
26. Enisa. (2014). *Network and information security in the finance sector : regulatory landscape and industry priorities*. ENISA.
27. Enisa. (2015). *Big data security : good practices and recommendations on the security of big data systems*. ENISA.
28. ENISA. (2020). *AI CYBERSECURITY CHALLENGES Threat Landscape for Artificial Intelligence AI CYBERSECURITY CHALLENGES ABOUT ENISA*.
<https://doi.org/10.2824/238222>
29. ENISA. (2022). *RISK MANAGEMENT STANDARDS*. <https://doi.org/10.2824/001991>
30. ENISA. (2023a). *Artificial Intelligence and Cybersecurity Research*.
31. ENISA. (2023b). NIS Directive. <https://www.enisa.europa.eu/topics/cybersecurity-policy/nis-directive-new>.
32. Erdélyi, O. J., & Goldsmith, J. (2022). Regulating artificial intelligence: Proposal for a global solution. *Government Information Quarterly*, 39(4).
<https://doi.org/10.1016/j.giq.2022.101748>
33. ES Taryba. (2023). *Dirbtinio intelekto aktas: Taryba ir Parlamentas susitarė dėl pirmųjų pasaulyje DI skirtų taisyklių*.
34. EUROPEAN COMMISSION. (2019). *HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. ETHICS GUIDELINES FOR TRUSTWORTHY AI*.
<https://ec.europa.eu/digital->
35. European Commission. (2020). White Paper On Artificial Intelligence - A European approach to excellence and trust. *Journal of Chemical Information and Modeling*, June.
36. EUROPOS KOMISIJA. (2020a). ... *dėl ES skaitmeninių finansų strategijos*.
https://ec.europa.eu/info/publications/digital-finance-outreach-2020_en
37. Europos Komisija. (2020). *Dėl ES skaitmeninių finansų strategijos*.
https://ec.europa.eu/info/publications/digital-finance-outreach-2020_en

38. EUROPOS KOMISIJA. (2020b). *Europos Sąjungos skaitmeninio dešimtmečio kibernetinio saugumo strategija*.
<https://www.idc.com/getdoc.jsp?containerId=prUS45213219>.
39. EUROPOS KOMISIJA. (2021). *EUROPOS PARLAMENTO IR TARYBOS REGLAMENTAS KURIUO NUSTATOMOS SUDERINTOS DIRBTINIO INTELEKTO TAISYKLĖS (DIRBTINIO INTELEKTO AKTAS) IR IŠ DALIES KEIČIAMSI TAM TIKRI SĄJUNGOS TEISĖKŪROS PROCEDŪRA PRIIMTI AKTAI*.
https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_lt.pdf.
40. Europos Parlamentas, E. S. T. (2023). *DIRECTIVE 2014/17/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*.
41. Europos parlamentas ir Europos sąjungos taryba. (2016).
Europos parlamento ir tarybos reglamentas 2016/ 679.
42. EUROPOS PARLAMENTAS IR EUROPOS SĄJUNGOS TARYBA. (2022).
EUROPOS PARLAMENTO IR TARYBOS DIREKTYVA (ES) 2022/2555.
43. Europos parlamentas ir Europos sąjungos taryba. (2024). *Europos parlamento ir tarybos reglamentas (ES) 2024/1689*. <http://data.europa.eu/eli/reg/2024/1689/oj>
44. Forter. (2024). Build trust across the entire digital commerce journey.
<https://www.Forter.Com/Platform/>.
45. Gemini. (2024). *Gemini Advanced*. <https://Gemini.Google/Advanced/>.
46. Gemini Team. (2024). Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. https://Storage.Googleapis.Com/Deepmind-Media/Gemini/Gemini_v1_5_report.Pdf.
47. Google DeepMind. (2024). *Gemini Models*.
<https://Deepmind.Google/Technologies/Gemini/>.
48. Google Cloud. (2024). *Security controls for Generative AI*.
<https://Cloud.Google.Com/Vertex-Ai/Generative-Ai/Docs/Security-Controls>.
49. Google DeepMind. (2024). *Gemini Flash*.
<https://Deepmind.Google/Technologies/Gemini/Flash/>.
50. Haenlein, M., & Kaplan, A. (2019). A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California Management Review*, 61(4). <https://doi.org/10.1177/0008125619864925>
51. IBM. (2024a). *Supported foundation models in watsonx.ai*.
<https://Dataplatform.Cloud.Ibm.Com/Docs/Content/Wsj/Analyze-Data/Fm-Models.Html?Context=wx&audience=wdp>.
52. IBM. (2024b). What is AI risk management? <https://www.Ibm.Com/Think/Insights/Ai-Risk-Management>.
53. IBM. (2024c). What is prompt engineering? <https://www.Ibm.Com/Topics/Prompt-Engineering>.
54. IBM AI community. (2024). *watsonx*.
https://Community.Ibm.Com/Community/User/Watsonx/Blogs/Maryam-Ashoori/2024/04/17/Llama-3-Models-Are-Now-Available-in-Watsonxai?Utm_source.
55. IBM Watsonx. (2024). *Meet watsonx*. <https://www.Ibm.Com/Watsonx>.

56. IBM watsonx. (2024a). *Services and integrations*.
<https://www.ibm.com/docs/en/watsonx/saas?topic=services-integrations>.
57. IBM watsonx. (2024b). Tokens and tokenization. In
<https://www.ibm.com/docs/en/watsonx/saas?topic=solutions-tokens>.
58. IBM watsonx. (2024c). *watsonx.ai Runtime service plans*.
<https://www.ibm.com/docs/en/watsonx/saas?topic=runtime-watsonxai-plans>.
59. Investopedia. (2024). 51% Attack: Definition, Who Is At Risk, Example, and Cost.
<https://www.investopedia.com/terms/1/51-Attack.asp>.
60. ISO 31000. (2018). ISO 31000:2018 Risk management — Guidelines. *ISO, 2018*.
61. *ISO/IEC 23894. Information technology-Artificial intelligence-Guidance on risk management*. (2023).
62. ISO/IEC 27005. (2018). *Information technology-Security techniques-Information security risk management*.
63. ISO/IEC 38507:2022. (2022). *Information technology-Governance of IT-Governance implications of the use of artificial intelligence by organizations*. www.iso.org
64. ISO/IEC TR 24028. (2020). *Information technology-Artificial intelligence-Overview of trustworthiness in artificial intelligence*.
65. ISO/IEC 27005. (2022). *Information security, cybersecurity and privacy protection-Guidance on managing information security risks*.
66. Kensho. (2024). We illuminate insights in the world's data. <https://kensho.com/>.
67. Kop, M. (2021). EU Artificial Intelligence Act: The European Approach to AI. *Transatlantic Antitrust and IPR Developments*, 2.
68. KPMG. (2024). Cybersecurity considerations 2024: Financial services sector.
<https://kpmg.com/xs/en/our-insights/ai-and-technology/cybersecurity-considerations-2024-financial-services-sector.html>.
69. Leitner-Hanetseder, S., Lehner, O. M., Eisl, C., & Forstenlechner, C. (2021). A profession in transition: actors, tasks and roles in AI-based accounting. *Journal of Applied Accounting Research*, 22(3). <https://doi.org/10.1108/JAAR-10-2020-0201>
70. Lietuvos bankas. (2021). Kredito unijos. <https://www.lb.lt/Lt/Kredito-Unijos>.
71. Lietuvos respublikos finansų ministerija. (2023). *Tarpinstitucinės darbo grupės rekomendacijos dėl 2023 – 2028 m. Fintech sektoriaus plėtros Lietuvoje gairių*.
72. Loesch, A. (2018a). *A Guide to Financial Regulation for Fintech Entrepreneurs. Chapter 1*.
73. Loesch, A. (2018b). *A Guide to Financial Regulation for Fintech Entrepreneurs. Chapter 2 & 3*.
74. Marius Sandbu, S. S. (2024). Microsoft 365 Copilot features and architecture explained. <https://www.techtarget.com/SearchEnterpriseAI/Tip/Microsoft-365-Copilot-Features-and-Architecture-Explained>.
75. Mark Cankett. (2022). The increasing role of AI in financial services.
<https://www.deloitte.com/uk/en/services/audit-assurance/blogs/2022/the-increasing-role-of-ai-in-financial-services.html>.

76. Mastercard. (2024). AI-powered decision management is key for the security of global payment ecosystems. <https://B2b.Mastercard.Com/News-and-Insights/Blog/Ai-Powered-Decision-Management-Key-for-Global-Credit-Card-Security/>.
77. McKinsey. (2023). *The state of AI in 2023: Generative AI's breakout year*.
78. McKinsey & Company. (2023a). *Capturing the full value of generative AI in banking*.
79. McKinsey & Company. (2023b). *Generative AI in finance: Finding the way to faster, deeper insights*.
80. Microsoft. (2024a). *AI Security Risk Assessment Best practices and guidance to secure AI systems*.
81. Microsoft. (2024b). *Copilot in Fabric consumption*. <https://Learn.Microsoft.Com/En-Us/Fabric/Get-Started/Copilot-Fabric-Consumption>.
82. Microsoft. (2024c). Kas yra dirbtinis intelektas kibernetinės saugos srityje? https://Www.Microsoft.Com/Lt-Lt/Security/Business/Security-101/What-Is-Ai-for-Cybersecurity?Utm_source.
83. Microsoft. (2024d). Microsoft Sentinel documentation. <https://Learn.Microsoft.Com/En-Us/Azure/Sentinel/>.
84. Microsoft 365. (2024). *Reinvent productivity with Microsoft 365 Copilot*. https://Www.Microsoft.Com/En-Us/Microsoft-365/Copilot?Msockid=24028b4efa47622a0a8c98cbfb9a63d5#tabs-Pill-Bar-Oc498c_tab0.
85. Microsoft 365 Copilot. (2024a). *Data, Privacy, and Security for Microsoft 365 Copilot*. <https://Learn.Microsoft.Com/En-Us/Copilot/Microsoft-365/Microsoft-365-Copilot-Privacy>.
86. Microsoft 365 Copilot. (2024b). *Microsoft 365 Copilot overview*. <https://Learn.Microsoft.Com/En-Us/Copilot/Microsoft-365/Microsoft-365-Copilot-Overview>.
87. Microsoft Copilot Studio. (2024). *Extend agent capabilities with Copilot connectors*. <https://Learn.Microsoft.Com/En-Us/Microsoft-Copilot-Studio/Copilot-Connectors-in-Copilot-Studio>.
88. MITRE. (2016). *The risk management framework and cyber resiliency*.
89. MITRE. (2023). *The MITRE AI Maturity Model and Organizational Assessment Tool Guide: A Path to Successful AI Adoption*.
90. Morland-Austin IT GRC Knowledge Base. (n.d.). *IT Risk Scenarios*. <https://Www.Morland-Austin.Com/It-Grc-Knowledge-Base/3.29/It-Risk-Scenarios#:~:Text=An%20IT%20risk%20scenario%20should%20include%20the%20following,Accidental%20Error%20C%20failure%20C%20natural%20and%20external%20requirement.%20Daugiau%20element%C5%B3>.
91. Nikhil Gopal. (2024). *The data and AI platform for the enterprise: IBM watsonx*. <https://Developer.Ibm.Com/Articles/Awb-Watsonx-Enterprise-Data-and-Ai-Platform/>.
92. NIST. (2012a). *Guide for conducting risk assessments*. <https://doi.org/10.6028/NIST.SP.800-30r1>
93. NIST. (2012b). *Guide for conducting risk assessments*. <https://doi.org/10.6028/NIST.SP.800-30r1>

94. NIST. (2018). *Risk management framework for information systems and organizations*. <https://doi.org/10.6028/NIST.SP.800-37r2>
95. NIST. (2021a). *AI Risk Management Framework Concept Paper*. <https://www.nist.gov/news-events/events/2021/10/kicking-nist-ai-risk-management-framework>
96. NIST. (2021b). *Draft-Taxonomy of AI Risk*. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
97. NIST. (2021c). *Four principles of explainable artificial intelligence*. <https://doi.org/10.6028/NIST.IR.8312>
98. NIST. (2021d). *Identifying and estimating cybersecurity risk for enterprise risk management*. <https://doi.org/10.6028/NIST.IR.8286A>
99. NIST. (2024). *Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile*. www.regulations.gov:
100. OECD. (2019). Scoping the OECD AI Principles: Deliberations of the Expert Group on Artificial Intelligence at the OECD (AIGO). *OECD Digital Economy Papers*, 291.
101. OECD. (2020). *The Impact of Big Data and Artificial Intelligence (AI) in the Insurance Sector*.
102. OECD. (2021a). AI in finance | OECD Business and Finance Outlook 2021 : AI in Business and Finance. In *OECD iLibrary*.
103. OECD. (2021b). *OECD Business and Finance Outlook 2021*. OECD. <https://doi.org/10.1787/ba682899-en>
104. OECD. (2022). OECD FRAMEWORK FOR THE CLASSIFICATION OF AI SYSTEMS. In *OECD Digital Economy Papers* (Vol. 323, Issue 323).
105. OECD. (2023). *GENERATIVE ARTIFICIAL INTELLIGENCE IN FINANCE*. <http://www.oecd.org/termsandconditions>.
106. OECD. (2024a). Artificial Intelligence (AI) in finance. <https://www.oecd.org/en/topics/sub-issues/digital-finance/artificial-intelligence-in-finance.html>.
107. Oecd. (2024). *EXPLANATORY MEMORANDUM ON THE UPDATED OECD DEFINITION OF AN AI SYSTEM*. <http://www.oecd.org/termsandconditions>.
108. OECD. (2024b). Financial markets. <https://www.oecd.org/en/topics/financial-markets.html>.
109. OECD. (2024c). *Regulatory approaches to artificial intelligence in finance*. <http://www.oecd.org/termsandconditions>
110. OpenAI. (n.d.-a). *OpenAI developer platform*. <https://platform.openai.com/docs/overview>.
111. OpenAI. (n.d.-b). *Safety at every step*. <https://openai.com/safety/>.
112. OpenAI. (2023a). *Europe privacy policy*. <https://openai.com/policies/privacy-policy/>.
113. OpenAI. (2023b). *GPT-4 Technical Report*.
114. OpenAI. (2023c). *Moving AI governance forward*. <https://openai.com/index/moving-ai-governance-forward/>.

115. OpenAI. (2024a). ChatGPT Capabilities Overview.
<https://help.openai.com/en/articles/9260256-chatgpt-capabilities-overview>.
116. OpenAI. (2024b). *GPT-4 System Card*. <https://cdn.openai.com/papers/gpt-4-system-card.pdf>.
117. OpenAI. (2024c). Hello GPT-4o. *<https://openai.com/index/hello-gpt-4o/>*.
118. OpenAI. (2024d). *OpenAI o1 System Card*. <https://openai.com/index/openai-o1-system-card/>.
119. OpenAI. (2024e). *OpenAI safety update*. <https://openai.com/index/openai-safety-update/>.
120. OpenAI. (2024f). *OpenAI Security Portal*. <https://trust.openai.com/>.
121. OpenAI. (2024g). Prompt engineering.
<https://platform.openai.com/docs/guides/prompt-engineering>.
122. OpenAI. (2024h). *What are tokens and how to count them?*
<https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them>.
123. OpenAI. (2024i, November 22). *Pricing*. <https://openai.com/chatgpt/pricing/>.
124. Pupillo, L., Fantin, S., Ferreira, A., & Polito, C. (2021). Artificial Intelligence and Cybersecurity: Technology, Governance and Policy Challenges. *CEPS*.
125. Rachel Meltzer. (2023). What Is Random Forest?
<https://careerfoundry.com/en/blog/data-analytics/what-is-random-forest/#:~:Text=Random%20Forest%20is%20used%20in%20banking%20to%20detect,Fraud.%20Talk%20about%20the%20robin%20hood%20of%20algorithms%21>.
126. Rai, A., Constantinides, P., & Sarker, S. (2019). Next-Generation Digital Platforms: Toward Human-AI Hybrids. *MIS Quarterly*, 43(1).
127. Raimondo, G. M. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). *Managing Information Risk*.
128. Salehi, H., & Burgueño, R. (2018). Emerging artificial intelligence methods in structural engineering. In *Engineering Structures* (Vol. 171).
<https://doi.org/10.1016/j.engstruct.2018.05.084>
129. Shahriar, S., Allana, S., Hazratifard, S. M., & Dara, R. (2023). A Survey of Privacy Risks and Mitigation Strategies in the Artificial Intelligence Life Cycle. *IEEE Access*, 11. <https://doi.org/10.1109/ACCESS.2023.3287195>
130. SHIFT. (2024). AI for Insurance Decisions. *<https://www.shift-technology.com/>*.
131. Shlomi Kushchi. (2024). 12 Pros and Cons of Microsoft (Azure) Sentinel.
<https://www.jit.io/blog/pros-and-cons-microsoft-azure-sentinel>.
132. Technical Committee, I. J. 1/SC 27. (2022). ISO - ISO/IEC 27001:2022 - Information security, cybersecurity and privacy protection — Information security management systems — Requirements. *Iso/Iec*, 2022.
133. THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION. (2022). *REGULATION (EU) 2022/2554 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL*.
134. Tony Martin-Vegue. (2021). *How to Write Strong Risk Scenarios and Statements*.
<https://www.isaca.org/resources/news-and->

Trends/Newsletters/Atisaca/2021/Volume-31/How-to-Write-Strong-Risk-Scenarios-and-Statements.

135. Two Sigma. (2024). We discover value in the world's data.
<https://www.twosigma.com/businesses/>.
136. Varun Raskar. (2024). Credit Risk Modelling using XGBoost: A Comprehensive Case Study. <https://medium.com/@varunraskar22/predicting-loan-approval-using-xgboost-a-comprehensive-case-study-0f70f2b4d200>.
137. Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act — Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4).
<https://doi.org/10.9785/crl-2021-220402>
138. Vectra. (2024). AI security for financial institutions.
<https://www.vectra.ai/solutions/industries/financial-services>.
139. Vinothkumar, J., & Karunamurthy, A. (2023). Recent Advancements in Artificial Intelligence Technology: Trends and Implications. *Quing: International Journal of Multidisciplinary Scientific Research and Development*, 2(1), 1–11.
<https://doi.org/10.54368/qijmsrd.2.1.0003>
140. Visa. (2024). Visa Advanced Authorisation and Visa Risk Manager.
<https://www.visasoutheasteurope.com/run-your-business/visa-security/risk-solutions/authorization-optimization.html>.
141. World Economic Forum. (2023). Global Cybersecurity Outlook 2023.
https://www3.weforum.org/docs/WEF_Global_Security_Outlook_Report_2023.pdf.
142. Young, C. S. (2019). The Theory, in Practice. In *Advanced Sciences and Technologies for Security Applications*. https://doi.org/10.1007/978-3-030-30600-7_12
143. Yusuf Enoch, S., Hong, J. B., & Seong Kim, D. (2021). *Model-based Cybersecurity Analysis: Past Work and Future Directions*.
144. Zacas Amosas. (2023). Kibernetinio saugumo AI tendencijos, kurias reikia stebėti 2024 m. <https://www.unite.ai/lt/kibernetinio-saugumo-ai-tendencijos%2C-kurias-reikia-stebeti-2024-m/>.
145. Zacharis, A., Katos, V., & Patsakis, C. (2024). Integrating AI-driven threat intelligence and forecasting in the cyber security exercise content generation lifecycle. *International Journal of Information Security*, 23(4), 2691–2710.
<https://doi.org/10.1007/s10207-024-00860-w>

1 priedas. Priešiškų grėsmių įvykiai

Reprezentaciniai pavyzdžiai – priešiškų grėsmių įvykiai

Threat Events (Characterized by TTPs)	Description
Perform reconnaissance and gather information.	
Perform perimeter network reconnaissance/scanning.	Adversary uses commercial or free software to scan organizational perimeters to obtain a better understanding of the information technology infrastructure and improve the ability to launch successful attacks.
Perform network sniffing of exposed networks.	Adversary with access to exposed wired or wireless data channels used to transmit information, uses network sniffing to identify components, resources, and protections.
Gather information using open source discovery of organizational information.	Adversary mines publically accessible information to gather information about organizational information systems, business processes, users or personnel, or external relationships that the adversary can subsequently employ in support of an attack.
Perform reconnaissance and surveillance of targeted organizations.	Adversary uses various means (e.g., scanning, physical observation) over time to examine and assess organizations and ascertain points of vulnerability.
Perform malware-directed internal reconnaissance.	Adversary uses malware installed inside the organizational perimeter to identify targets of opportunity. Because the scanning, probing, or observation does not cross the perimeter, it is not detected by externally placed intrusion detection systems.
Craft or create attack tools.	
Craft phishing attacks.	Adversary counterfeits communications from a legitimate/trustworthy source to acquire sensitive information such as usernames, passwords, or SSNs. Typical attacks occur via email, instant messaging, or comparable means; commonly directing users to websites that appear to be legitimate sites, while actually stealing the entered information.
Craft spear phishing attacks.	Adversary employs phishing attacks targeted at high value targets (e.g., senior leaders/executives).
Craft attacks specifically based on deployed information technology environment.	Adversary develops attacks (e.g., crafts targeted malware) that take advantage of adversary knowledge of the organizational information technology environment.
Create counterfeit/spoof website.	Adversary creates duplicates of legitimate websites; when users visit a counterfeit site, the site can gather information or download malware.
Craft counterfeit certificates.	Adversary counterfeits or compromises a certificate authority, so that malware or connections will appear legitimate.
Create and operate false front organizations to inject malicious components into the supply chain.	Adversary creates false front organizations with the appearance of legitimate suppliers in the critical life-cycle path that then inject corrupted/malicious information system components into the organizational supply chain.
Deliver/insert/install malicious capabilities.	
Deliver known malware to internal organizational information systems (e.g., virus via email).	Adversary uses common delivery mechanisms (e.g., email) to install/insert known malware (e.g., malware whose existence is known) into organizational information systems.
Deliver modified malware to internal organizational information systems.	Adversary uses more sophisticated delivery mechanisms than email (e.g., web traffic, instant messaging, FTP) to deliver malware and possibly modifications of known malware to gain access to internal organizational information systems.
Deliver targeted malware for control of internal systems and exfiltration of data.	Adversary installs malware that is specifically designed to take control of internal organizational information systems, identify sensitive information, exfiltrate the information back to adversary, and conceal these actions.
Deliver malware by providing removable media.	Adversary places removable media (e.g., flash drives) containing malware in locations external to organizational physical perimeters but where employees are likely to find the media (e.g., facilities parking lots, exhibits at conferences attended by employees) and use it on organizational information systems.

Threat Events (Characterized by TTPs)	Description
Insert untargeted malware into downloadable software and/or into commercial information technology products.	Adversary corrupts or inserts malware into common freeware, shareware or commercial information technology products. Adversary is not targeting specific organizations, simply looking for entry points into internal organizational information systems. Note that this is particularly a concern for mobile applications.
Insert targeted malware into organizational information systems and information system components.	Adversary inserts malware into organizational information systems and information system components (e.g., commercial information technology products), specifically targeted to the hardware, software, and firmware used by organizations (based on knowledge gained via reconnaissance).
Insert specialized malware into organizational information systems based on system configurations.	Adversary inserts specialized, non-detectable, malware into organizational information systems based on system configurations, specifically targeting critical information system components based on reconnaissance and placement within organizational information systems.
Insert counterfeit or tampered hardware into the supply chain.	Adversary intercepts hardware from legitimate suppliers. Adversary modifies the hardware or replaces it with faulty or otherwise modified hardware.
Insert tampered critical components into organizational systems.	Adversary replaces, through supply chain, subverted insider, or some combination thereof, critical information system components with modified or corrupted components.
Install general-purpose sniffers on organization-controlled information systems or networks.	Adversary installs sniffing software onto internal organizational information systems or networks.
Install persistent and targeted sniffers on organizational information systems and networks.	Adversary places within internal organizational information systems or networks software designed to (over a continuous period of time) collect (sniff) network traffic.
Insert malicious scanning devices (e.g., wireless sniffers) inside facilities.	Adversary uses postal service or other commercial delivery services to deliver to organizational mailrooms a device that is able to scan wireless communications accessible from within the mailrooms and then wirelessly transmit information back to adversary.
Insert subverted individuals into organizations.	Adversary places individuals within organizations who are willing and able to carry out actions to cause harm to organizational missions/business functions.
Insert subverted individuals into privileged positions in organizations.	Adversary places individuals in privileged positions within organizations who are willing and able to carry out actions to cause harm to organizational missions/business functions. Adversary may target privileged functions to gain access to sensitive information (e.g., user accounts, system files, etc.) and may leverage access to one privileged capability to get to another capability.
Exploit and compromise.	
Exploit physical access of authorized staff to gain access to organizational facilities.	Adversary follows ("tailgates") authorized individuals into secure/controlled locations with the goal of gaining access to facilities, circumventing physical security checks.
Exploit poorly configured or unauthorized information systems exposed to the Internet.	Adversary gains access through the Internet to information systems that are not authorized for Internet connectivity or that do not meet organizational configuration requirements.
Exploit split tunneling.	Adversary takes advantage of external organizational or personal information systems (e.g., laptop computers at remote locations) that are simultaneously connected securely to organizational information systems or networks and to nonsecure remote connections.
Exploit multi-tenancy in a cloud environment.	Adversary, with processes running in an organizationally-used cloud environment, takes advantage of multi-tenancy to observe behavior of organizational processes, acquire organizational information, or interfere with the timely or correct functioning of organizational processes.
Exploit known vulnerabilities in mobile systems (e.g., laptops, PDAs, smart phones).	Adversary takes advantage of fact that transportable information systems are outside physical protection of organizations and logical protection of corporate firewalls, and compromises the systems based on known vulnerabilities to gather information from those systems.
Exploit recently discovered vulnerabilities.	Adversary exploits recently discovered vulnerabilities in organizational information systems in an attempt to compromise the systems before mitigation measures are available or in place.

Threat Events (Characterized by TTPs)	Description
Exploit vulnerabilities on internal organizational information systems.	Adversary searches for known vulnerabilities in organizational internal information systems and exploits those vulnerabilities.
Exploit vulnerabilities using zero-day attacks.	Adversary employs attacks that exploit as yet unpublicized vulnerabilities. Zero-day attacks are based on adversary insight into the information systems and applications used by organizations as well as adversary reconnaissance of organizations.
Exploit vulnerabilities in information systems timed with organizational mission/business operations tempo.	Adversary launches attacks on organizations in a time and manner consistent with organizational needs to conduct mission/business operations.
Exploit insecure or incomplete data deletion in multi-tenant environment.	Adversary obtains unauthorized information due to insecure or incomplete data deletion in a multi-tenant environment (e.g., in a cloud computing environment).
Violate isolation in multi-tenant environment.	Adversary circumvents or defeats isolation mechanisms in a multi-tenant environment (e.g., in a cloud computing environment) to observe, corrupt, or deny service to hosted services and information/data.
Compromise critical information systems via physical access.	Adversary obtains physical access to organizational information systems and makes modifications.
Compromise information systems or devices used externally and reintroduced into the enterprise.	Adversary installs malware on information systems or devices while the systems/devices are external to organizations for purposes of subsequently infecting organizations when reconnected.
Compromise software of organizational critical information systems.	Adversary inserts malware or otherwise corrupts critical internal organizational information systems.
Compromise organizational information systems to facilitate exfiltration of data/information.	Adversary implants malware into internal organizational information systems, where the malware over time can identify and then exfiltrate valuable information.
Compromise mission-critical information.	Adversary compromises the integrity of mission-critical information, thus preventing or impeding ability of organizations to which information is supplied, from carrying out operations.
Compromise design, manufacture, and/or distribution of information system components (including hardware, software, and firmware).	Adversary compromises the design, manufacture, and/or distribution of critical information system components at selected suppliers.
<i>Conduct an attack (i.e., direct/coordinate attack tools or activities).</i>	
Conduct communications interception attacks.	Adversary takes advantage of communications that are either unencrypted or use weak encryption (e.g., encryption containing publically known flaws), targets those communications, and gains access to transmitted information and channels.
Conduct wireless jamming attacks.	Adversary takes measures to interfere with wireless communications so as to impede or prevent communications from reaching intended recipients.
Conduct attacks using unauthorized ports, protocols and services.	Adversary conducts attacks using ports, protocols, and services for ingress and egress that are not authorized for use by organizations.
Conduct attacks leveraging traffic/data movement allowed across perimeter.	Adversary makes use of permitted information flows (e.g., email communication, removable storage) to compromise internal information systems, which allows adversary to obtain and exfiltrate sensitive information through perimeters.
Conduct simple Denial of Service (DoS) attack.	Adversary attempts to make an Internet-accessible resource unavailable to intended users, or prevent the resource from functioning efficiently or at all, temporarily or indefinitely.
Conduct Distributed Denial of Service (DDoS) attacks.	Adversary uses multiple compromised information systems to attack a single target, thereby causing denial of service for users of the targeted information systems.
Conduct targeted Denial of Service (DoS) attacks.	Adversary targets DoS attacks to critical information systems, components, or supporting infrastructures, based on adversary knowledge of dependencies.
Conduct physical attacks on organizational facilities.	Adversary conducts a physical attack on organizational facilities (e.g., sets a fire).
Conduct physical attacks on infrastructures supporting organizational facilities.	Adversary conducts a physical attack on one or more infrastructures supporting organizational facilities (e.g., breaks a water main, cuts a power line).
Conduct cyber-physical attacks on organizational facilities.	Adversary conducts a cyber-physical attack on organizational facilities (e.g., remotely changes HVAC settings).

Threat Events (Characterized by TTPs)	Description
Conduct data scavenging attacks in a cloud environment.	Adversary obtains data used and then deleted by organizational processes running in a cloud environment.
Conduct brute force login attempts/password guessing attacks.	Adversary attempts to gain access to organizational information systems by random or systematic guessing of passwords, possibly supported by password cracking utilities.
Conduct nontargeted zero-day attacks.	Adversary employs attacks that exploit as yet unpublicized vulnerabilities. Attacks are not based on any adversary insights into specific vulnerabilities of organizations.
Conduct externally-based session hijacking.	Adversary takes control of (hijacks) already established, legitimate information system sessions between organizations and external entities (e.g., users connecting from off-site locations).
Conduct internally-based session hijacking.	Adversary places an entity within organizations in order to gain access to organizational information systems or networks for the express purpose of taking control (hijacking) an already established, legitimate session either between organizations and external entities (e.g., users connecting from remote locations) or between two locations within internal networks.
Conduct externally-based network traffic modification (man in the middle) attacks.	Adversary, operating outside organizational systems, intercepts/eavesdrops on sessions between organizational and external systems. Adversary then relays messages between organizational and external systems, making them believe that they are talking directly to each other over a private connection, when in fact the entire communication is controlled by the adversary. Such attacks are of particular concern for organizational use of community, hybrid, and public clouds.
Conduct internally-based network traffic modification (man in the middle) attacks.	Adversary operating within the organizational infrastructure intercepts and corrupts data sessions.
Conduct outsider-based social engineering to obtain information.	Externally placed adversary takes actions (e.g., using email, phone) with the intent of persuading or otherwise tricking individuals within organizations into revealing critical/sensitive information (e.g., personally identifiable information).
Conduct insider-based social engineering to obtain information.	Internally placed adversary takes actions (e.g., using email, phone) so that individuals within organizations reveal critical/sensitive information (e.g., mission information).
Conduct attacks targeting and compromising personal devices of critical employees.	Adversary targets key organizational employees by placing malware on their personally owned information systems and devices (e.g., laptop/notebook computers, personal digital assistants, smart phones). The intent is to take advantage of any instances where employees use personal information systems or devices to handle critical/sensitive information.
Conduct supply chain attacks targeting and exploiting critical hardware, software, or firmware.	Adversary targets and compromises the operation of software (e.g., through malware injections), firmware, and hardware that performs critical functions for organizations. This is largely accomplished as supply chain attacks on both commercial off-the-shelf and custom information systems and components.
<i>Achieve results (i.e., cause adverse impacts, obtain information)</i>	
Obtain sensitive information through network sniffing of external networks.	Adversary with access to exposed wired or wireless data channels that organizations (or organizational personnel) use to transmit information (e.g., kiosks, public wireless networks) intercepts communications.
Obtain sensitive information via exfiltration.	Adversary directs malware on organizational systems to locate and surreptitiously transmit sensitive information.
Cause degradation or denial of attacker-selected services or capabilities.	Adversary directs malware on organizational systems to impair the correct and timely support of organizational mission/business functions.
Cause deterioration/destruction of critical information system components and functions.	Adversary destroys or causes deterioration of critical information system components to impede or eliminate organizational ability to carry out missions or business functions. Detection of this action is not a concern.
Cause integrity loss by creating, deleting, and/or modifying data on publicly accessible information systems (e.g., web defacement).	Adversary vandalizes, or otherwise makes unauthorized changes to, organizational websites or data on websites.
Cause integrity loss by polluting or corrupting critical data.	Adversary implants corrupted and incorrect data in critical data, resulting in suboptimal actions or loss of confidence in organizational data/services.

Threat Events (Characterized by TTPs)	Description
Cause integrity loss by injecting false but believable data into organizational information systems.	Adversary injects false but believable data into organizational information systems, resulting in suboptimal actions or loss of confidence in organizational data/services.
Cause disclosure of critical and/or sensitive information by authorized users.	Adversary induces (e.g., via social engineering) authorized users to inadvertently expose, disclose, or mishandle critical/sensitive information.
Cause unauthorized disclosure and/or unavailability by spilling sensitive information.	Adversary contaminates organizational information systems (including devices and networks) by causing them to handle information of a classification/sensitivity for which they have not been authorized. The information is exposed to individuals who are not authorized access to such information, and the information system, device, or network is unavailable while the spill is investigated and mitigated.
Obtain information by externally located interception of wireless network traffic.	Adversary intercepts organizational communications over wireless networks. Examples include targeting public wireless access or hotel networking connections, and drive-by subversion of home or organizational wireless routers.
Obtain unauthorized access.	Adversary with authorized access to organizational information systems, gains access to resources that exceeds authorization.
Obtain sensitive data/information from publicly accessible information systems.	Adversary scans or mines information on publically accessible servers and web pages of organizations with the intent of finding sensitive information.
Obtain information by opportunistically stealing or scavenging information systems/components.	Adversary steals information systems or components (e. g., laptop computers or data storage media) that are left unattended outside of the physical perimeters of organizations, or scavenges discarded components.
<i>Maintain a presence or set of capabilities.</i>	
Obfuscate adversary actions.	Adversary takes actions to inhibit the effectiveness of the intrusion detection systems or auditing capabilities within organizations.
Adapt cyber attacks based on detailed surveillance.	Adversary adapts behavior in response to surveillance and organizational security measures.
<i>Coordinate a campaign.</i>	
Coordinate a campaign of multi-staged attacks (e.g., hopping).	Adversary moves the source of malicious commands or actions from one compromised information system to another, making analysis difficult.
Coordinate a campaign that combines internal and external attacks across multiple information systems and information technologies.	Adversary combines attacks that require both physical presence within organizational facilities and cyber methods to achieve success. Physical attack steps may be as simple as convincing maintenance personnel to leave doors or cabinets open.
Coordinate campaigns across multiple organizations to acquire specific information or achieve desired outcome.	Adversary does not limit planning to the targeting of one organization. Adversary observes multiple organizations to acquire necessary information on targets of interest.
Coordinate a campaign that spreads attacks across organizational systems from existing presence.	Adversary uses existing presence within organizational systems to extend the adversary's span of control to other organizational systems including organizational infrastructure. Adversary thus is in position to further undermine organizational ability to carry out missions/business functions.
Coordinate a campaign of continuous, adaptive, and changing cyber attacks based on detailed surveillance.	Adversary attacks continually change in response to surveillance and organizational security measures.
Coordinate cyber attacks using external (outsider), internal (insider), and supply chain (supplier) attack vectors.	Adversary employs continuous, coordinated attacks, potentially using all three attack vectors for the purpose of impeding organizational operations.

2 priedas. Ne priešišku grėsmių įvykiai

Reprezentatyvūs pavyzdžiai – ne priešišku grėsmių įvykiai.

Threat Event	Description
Spill sensitive information	Authorized user erroneously contaminates a device, information system, or network by placing on it or sending to it information of a classification/sensitivity which it has not been authorized to handle. The information is exposed to access by unauthorized individuals, and as a result, the device, system, or network is unavailable while the spill is investigated and mitigated.
Mishandling of critical and/or sensitive information by authorized users	Authorized privileged user inadvertently exposes critical/sensitive information.
Incorrect privilege settings	Authorized privileged user or administrator erroneously assigns a user exceptional privileges or sets privilege requirements on a resource too low.
Communications contention	Degraded communications performance due to contention.
Unreadable display	Display unreadable due to aging equipment.
Earthquake at primary facility	Earthquake of organization-defined magnitude at primary facility makes facility inoperable.
Fire at primary facility	Fire (not due to adversarial activity) at primary facility makes facility inoperable.
Fire at backup facility	Fire (not due to adversarial activity) at backup facility makes facility inoperable or destroys backups of software, configurations, data, and/or logs.
Flood at primary facility	Flood (not due to adversarial activity) at primary facility makes facility inoperable.
Flood at backup facility	Flood (not due to adversarial activity) at backup facility makes facility inoperable or destroys backups of software, configurations, data, and/or logs.
Hurricane at primary facility	Hurricane of organization-defined strength at primary facility makes facility inoperable.
Hurricane at backup facility	Hurricane of organization-defined strength at backup facility makes facility inoperable or destroys backups of software, configurations, data, and/or logs.
Resource depletion	Degraded processing performance due to resource depletion.
Introduction of vulnerabilities into software products	Due to inherent weaknesses in programming languages and software development environments, errors and vulnerabilities are introduced into commonly used software products.
Disk error	Corrupted storage due to a disk error.
Pervasive disk error	Multiple disk errors due to aging of a set of devices all acquired at the same time, from the same supplier.
Windstorm/tornado at primary facility	Windstorm/tornado of organization-defined strength at primary facility makes facility inoperable.
Windstorm/tornado at backup facility	Windstorm/tornado of organization-defined strength at backup facility makes facility inoperable or destroys backups of software, configurations, data, and/or logs.

3 priedas. Turto taksonomija 1

Dirbtinio intelekto turto taksonomija.



PROCESSES

- Data Ingestion
- Data Storage
- Data Exploration/Pre-processing
- Data Understanding
- Data Labelling
- Data Augmentation
- Data Collection
- Feature Selection
- Reduction/Discretization technique
- Model selection/building, training, and testing
- Model Tuning
- Model adaptation-transfer learning/Model deployment
- Model Maintenance



ENVIRONMENT/TOOLS

- Communication Networks
- Communication Protocols
- Cloud
- Data Ingestion Platforms
- Data Exploration Platforms
- Data Exploration Tools
- DBMS
- Distributed File System
- Computational Platforms
- Integrated Development Environment
- Libraries (with algorithms for transformation, labelling, etc)
- Monitoring Tools
- Operating System/Software
- Optimization Techniques
- Machine Learning Platforms
- Processors
- Visualization Tools



ARTEFACTS

- Access Control Lists
- Use Case
- Value Proposition and Business Model
- Informal/Semi-formal AI Requirements, GQM (Goal/Question/Metrics) model
- Data Governance Policies
- Data display and plots
- Descriptive statistical parameters
- Model framework, software, firmware or hardware incarnations
- Composition artefacts: AI models composition builder
- High-Level Test cases
- Model Architecture
- Model hardware design
- Data and Metadata schemata
- Data Indexes



MODELS

- Algorithms
- Data Pre-processing Algorithms
- Training Algorithms
- Subspace (feature) Selection Algorithm
- Model
- Model parameters
- Model Performance
- Training Parameters
- Hyper Parameters
- Trained Models
- Tuned Model



ACTORS/STAKEHOLDERS

- Data Owner
- Data Scientists/AI developer
- Data Engineers
- End Users
- Data Provide/Broker
- Cloud Provider
- Model Provider
- Service Consumers/Model Users



DATA

- Raw Data
- Labelled Data Set
- Public Data Set
- Training Data
- Augmented Data Set
- Testing Data
- Validation Data Set
- Evaluation Data
- Pre-processed Data Set

4 priedas. Grėsmių taksonomija

Grėsmių taksonomija sudaryta pagal ENISA.

Threat Category	Threat	Description	Potential impact	Affected assets
Nefarious Activity/Abuse	Access Control List (ACL) manipulation	In AI data collection scenarios, group-based ACLs for datasets may fail when the nesting of large groups is changed. Given a data set and a group of sources <i>Sensor_Group_A</i> which has been granted access to update it, it is easy to check if an individual user or sensor is a member of <i>Sensor_Group_A</i> and inherits the corresponding permissions. However, if <i>Sensor_Group_A</i> is joined as a member to many other groups, inherited permissions become difficult to check for each sensor and escalation of privileges of untrusted sources may result. The threat involves Implicit privilege elevation attacks take advantage of group nesting modifications to upscale access permissions for specific users.	Integrity	Artefacts
	Adversarial examples	Targeting the inference phase of ML and deep learning systems that AI is based on is one of the most prominent and highly publicized threats. Adversarial examples refer to data that include perturbations that are imperceptible to the human eye, but that can have an impact on the effectiveness and performance of ML models.	Integrity Availability	Model Data
	Backdoor/insert attacks on training datasets	Threaten the ML model's integrity by trying to introduce spurious inferences. Attackers introduce special trigger patterns in part of the training data, and presenting the trigger in the inference phase will cause targeted misclassifications. For example, an attacker can introduce in the training data of an image classifier connected to a surveillance camera an example including a certain pixel pattern and the label "policeman". Once the classifier is trained and deployed, the attacker wears a t-shirt with that pattern and passes by the camera with a gun in hand without triggering any alarm.	Integrity	Data
	Compromising AI inference's correctness - data	This type of threats refers to possible exploitations involving either data manipulation, or selection bias in raw data, or modification of labels and deletion or omission of labelled data items. It may also refer to compromising AI correctness by insertion of adversarial data (poisoned/manipulated) in augmented data sets, as well as by means of interruption of training or modification of model parameters.	Integrity	Data
	Compromise and limit AI results	This type of threat can emerge due to involuntary or unintentionally actions from Data Owners, that may hide data due to business secrets or by not recognizing its value; by AI/ML	Integrity Availability	Model Actors

Threat Category	Threat	Description	Potential impact	Affected assets
		designers and engineers, that can intentionally tamper or, due to lack of experience, miss to include data. This threat may also be related to AI/ML service users not being able to understand the model capabilities and/or results.		Artefact
	Compromising ML inference's correctness – algorithms	Threats to the availability of the ML training algorithm, as well as threats that aim at compromising the training algorithm to adversely affect the desired accuracy.	Integrity Availability	Model
	Compromising ML pre-processing	Flaws or defects of the data and metadata schemata greatly influence the quality of the analysis by applications that use the data. In AI applications, a flawed schema will negatively impact on the quality of the ingested information. Flaws often result from inconsistencies in the use of modeling methodologies, but may also depend on intentional schema poisoning, i.e. any manipulation of a schema intended to compromise the programs that ingest or pre-process data that use it. It is also possible for adversaries to mount Schema-based denial of service attacks, which modify the data schema so that it does not contain required information for subsequent processing.	Integrity Availability	Data Artefacts
	Compromising ML training – augmented data	Threats to augmented datasets due to inconsistency with the training set they are derived from, and specifically when highly diverse, automatically generated data are added to a data set of collected data, which are very consistent but highly representative of their application domain, so there would be no need to limit overfitting. Enriching data always entails some risks. This threat can lead to non-satisfaction of functional requirements, i.e. poor inference.	Integrity Availability Confidentiality	Data
	Compromising ML training – validation data	This threat refers to shortening the training of the ML model dramatically by compromising the integrity of the validation dataset. It also includes, generation of adversarial validation data that are quite different from genuine training set data	Integrity Availability	Data
	Compromise of data brokers/providers	This threat refers to compromising data brokers/providers to influence the machine learning process as they can deliberately or accidentally manipulate the data sent to the AI process, in several different ways: poisoning-via-insertion of malicious data; deleting registries to eliminate features either by changing the data, removing part of it or adding new registries. In addition, sometimes the mere data availability prevails over any consideration on data quality, with the risk that learning models are fed with data streams that do not reflect the statistical characteristics of a phenomenon and determining likely biases in the subsequent decisional processes.	Integrity Availability	Actor

Threat Category	Threat	Description	Potential impact	Affected assets
	Compromise of model frameworks	Model frameworks fail when are misconfigured or offer additional attack vectors with respect to traditional software, firmware and hardware environments. The ML platform's data volume and processing requirements mean that the workloads are often handled on the cloud, adding another level of complexity and vulnerability. Moreover, the threat of backdoors in libraries is also evident, similarly to the potential threats of attacks on model input and output data can be performed on the hardware, firmware, Operating System (OS) and software level.	Integrity Availability	Artefacts Environment/tools
	Corruption of data indexes	Data indexes threats manifest when their content becomes corrupted. Corruption may be the result of an attack, or due to system crashes or loss of network connectivity during index replication. The same events may cause interruptions of index construction tasks, bringing a partially built (and therefore defective) index to production. Also, running out of storage capacity during indexing or replication may cause an entire data index to be deleted. Denial-of-service attacks to indexes intentionally corrupt data indexes to decrease the performance of data access. Additionally, timing attacks to indexes use access time to public items before and after inserting them (which depend on the index content) to infer the presence and size of inaccessible data items.	Integrity Availability Confidentiality	Artefacts
	Data poisoning	This threat relates to the injection of erroneous/tampered/wrong data in the training set or the validation set by either getting legitimate access or illegitimate one through exploiting poor authentication/authorization mechanisms. The aim is to adversely affect operation of the AI system.	Integrity Availability	Process Environment/tools Model
	Data tampering	Actors like AI/ML designers and engineers can deliberately or unintentionally manipulate and expose data. Data can also be manipulated during the storage procedure and by means of some processes like feature selection. Besides interfering with model inference, this type of threat can also bring severe discriminatory issues by introducing bias.	Availability Integrity	Process Environment/tools Model
	DDoS	Distributed Denial of Service attacks may be utilized by adversaries to reduce the availability of online IT systems and distributed file systems (e.g. cloud storage) used to support AI systems and their operation.	Availability	Environment/tools
	Elevation-of-Privilege	These threats refers to exploiting trained and tuned models to gain access to parameters values and even to understand whether some data was part of the data set used.	Confidentiality Availability	Model Data
	Insider threat	AI designers and developers may deliberately expose data and models for a variety of reasons, e.g. revenge or extortion. Integrity, data	Confidentiality Integrity	Actors

Threat Category	Threat	Description	Potential impact	Affected assets
		confidentiality and trustworthiness are the main impacted security properties.	Availability	
	Introduction of selection bias	Data owners may introduce selection bias on purpose when publishing raw data in order to adversely affect inference to be drawn on the data.	Integrity Availability	Data
	Label manipulation or weak labelling	This threat refers to supervised learning systems, which not infer correctly due to wrong or imprecise data labels. If adversaries can only modify the training labels with some or all knowledge of the target model, they need to find the most vulnerable labels. Random perturbation of labels is one possible attack, while additionally there is the case of adversarial label noise (intentional switching of classification labels leading to deterministic noise, an error that the model cannot capture due to its generalization bias).	Availability Integrity	Processes Data
	Manipulation of data sets and data transfer process	These threats are seen in context of storage of data sets in infrastructures provided by third parties, which make them remotely accessible. The threat refers to manipulation and tampering of the data stored and manipulation of the data transfer process.	Confidentiality Integrity	Environment/tools
	Manipulation of labelled data	Threats to labelled data items occur when enough labels and data are deleted/omitted, when a sufficient number of spurious labelled data is included into the data set, or when enough labels are modified. Since the labelled data set is used for the purpose of training a ML model in the supervised setting, all such modifications affect the model training and inference (e.g., shifting the model's classification boundary).	Integrity	Data
	Manipulation of model tuning	Adversaries may fine-tune hyper-parameters and thus influence the AI system's behaviour. Hyper-parameters can be a vector for accidental overfitting. In addition, hard to detect changes to hyper-parameters would make an ideal insider attack. The usage of default hyper-parameters may increase transferability of adversarial attacks. When automatic optimization is done, the AI/ML might also be compromised in case the optimization algorithm is manipulated by adversaries.	Integrity Availability	Process Model
	Manipulation of optimization algorithm	Optimization algorithms are often used in the context of processes like Model Tuning to setup hyper-parameters values. Accordingly, nefarious abuse of such algorithms by adversaries may lead to erroneous tuning of models.	Availability	Models Processes
	Misclassification based on adversarial examples	This threat involves manipulation of model parameters or use of adversarial examples during inference to force misclassification of model results. This type of threats is related to the Actors category, as they have access to models and data sets, such is the case of Cloud providers, model providers and model users.	Integrity Availability	Actors Processes

Threat Category	Threat	Description	Potential impact	Affected assets
		Misclassification can also be instigated by Processes assets, such is the case of using adversarial examples during training and transfer learning stages, as well as during the inference stage.		
	ML model confidentiality	This threat refers to exploitation of the ML model to leak (in its outputs or otherwise) some information about its internal parameters or performance to un-authorized parties.	Confidentiality	Model
	ML Model integrity manipulation	This threat refers to manipulation of the ML model by delivering output values that were not generated based on its internal parameters or by delivering overtly biased or useless outputs (e.g. constant, or undistinguishable from random noise).	Integrity	Model
	Model backdoors	It is often the case that 3rd parties provide models (called as "Teacher" models), previously trained and fine-tuned with large datasets that are useful to learn from small datasets and/or by organizations without access to high computational clusters. The resulting models may be subject to backdoor threats that expose their inner working (breach of confidentiality), impact their operation (integrity breach) or degrade/cancel their performance (impact on availability).	Integrity Availability Confidentiality	Processes
	Model poisoning	This threat refers to a legitimate model file being replaced entirely by a poisoned model file. In the context of AI as a Service, with many types of data and code being uploaded on cloud infrastructures, this threat may be realized by exploiting potential weaknesses of cloud providers.	Integrity Availability	Models Actors
	Model Sabotage	Sabotaging the model is a nefarious threat that refers to exploitation or physical damage of libraries and machine learning platforms that host or supply AI/ML services and systems.	Availability Integrity	Environment/tools Model
	Online system manipulation	This is related to model replacement by a malicious backdoored model, used for targeted or non-targeted attack, which can be exploited by Actors like Cloud Providers during Processes like model training or transfer.	Confidentiality Integrity	Model
	Overloading/conf using labelled dataset	Append attacks target availability by adding random samples to the training set to the point of preventing any model trained on that data set from computing any meaningful inference. The threat may lead to overfitting or underfitting the labelled dataset.	Availability	Data
	Reducing data accuracy	This threat refers to the reduction of the degree of data accuracy, by directly modifying the data or by mixing datasets with highly different degrees of quality.	Integrity Availability	Data
	Reduce effectiveness of AI/ML results	Users can make erroneous usage of AI services, either for not having a good understanding about the model capabilities or by not being able to	Integrity Availability	Processes Actors

Threat Category	Threat	Description	Potential impact	Affected assets
		understand when changes in the process imply model maintenance, and possibly re-training procedures. End-users can modify the input data to the model that results in "de-training" the model.		
	Sabotage	Sabotage involves intentionally destroying or maliciously affecting the IT infrastructure that supports AI systems.	Availability	Environment/tools
	Scarce data	AI relies on the availability of consistent and accessible data. This threat involves data scarcity (deliberately created by an adversary) that may compromise AI viability and/or compromise and limit its results. This can be exploited deliberately (for nefarious activities) or unintentionally during Data Ingestion.	Availability	Data Processes
	Transferability of adversarial attacks	ML and deep learning models are mostly based on an inductive approach to problem solving, as opposed to the deductive approach of traditional mathematical modelling. This means that experience matters and not always it can be given for granted that ML models can be smoothly transferred and applied in a new scenario and for a new AI application. This threat refers to adversarial examples that may be transferred to AI/ML applications and Environment/Tools like AI/ML libraries and machine learning platforms.	Integrity	Process Environment/tools
	Unauthorized access to data sets and data transfer process	These threats are seen in context of storage of data sets in infrastructures provided by third parties, which make them remotely accessible. The threat refers to unauthorized access of the data stored and unauthorized access to the inner workings of the data transfer process.	Confidentiality Integrity	Environment/tools
	Unauthorized access to models' code	This threat refers to machine learning libraries and machine learning platforms being manipulated to inject malicious code that will exploit users models and gain access to datasets.	Confidentiality Integrity	Environment/tools Model
	White-box, targeted or non-targeted	This threat refers to misclassification to a specific target class or to a different class rather than the correct one. This type of threat is mainly associated with Processes assets like Model Selection/Building, Training, Testing, Transfer Learning and Model Deployment and can be exploited by Actors such as Model Providers.	Integrity Availability	Processes
Unintentional Damage	Bias introduced by data owners	Data Owners may try to hide information that will be fed to the AI systems as part of their business interests. Moreover, they are also people that may be biased themselves, as they tend to be far from raw data and may be incapable of giving good data to the models. This type of threat can severely affect trustworthiness of AI systems.	Availability Integrity	Actors

Threat Category	Threat	Description	Potential impact	Affected assets
	Compromising AI inference's correctness - data	This type of threats refers possible exploitations involving either data manipulation, or unintentional selection bias in raw data, or modification of labels and deletion or omission of labelled data items. It may also refer to compromising AI correctness by insertion of adversarial data in augmented data sets, as well as by means of interruption of training or modification of model parameters.	Integrity	Data
	Compromise and limit AI results	This type of threat can emerge due to involuntary or unintentionally actions from Data Owners, that may hide data due to business secrets or by not recognizing its value; by AI/ML designers and engineers, that can intentionally tamper or, due to lack of experience, miss to include data. This threat may also be related to AI/ML service users not being able to understand the model capabilities and/or results.	Integrity Availability	Model Actors Artefact
	Compromising ML inference's correctness – algorithms	Threats to the availability of the ML training algorithm, as well as threats that aim at compromising the training algorithm to adversely affect the desired accuracy.	Integrity Availability	Model
	Compromising ML training – augmented data	Threats to augmented datasets due to inconsistency with the training set they are derived from, and specifically when highly diverse, automatically generated data are added to a data set of collected data, which are very consistent but highly representative of their application domain, so there would be no need to limit overfitting. Enriching data always entails some risks. This threat can lead to non-satisfaction of functional requirements, i.e. poor inference.	Availability Integrity	Data
	Compromising feature selection	This threat refers to performance degradation of feature selection algorithms by delivering feature sets that are strongly predictive only for some for some classes, neglecting other features that are needed to discriminate difficult classes.	Integrity Availability	Model
	Compromise of data brokers/providers	This threat refers to compromising data brokers/providers to influence the machine learning process as they can deliberately or accidentally manipulate the data sent to the AI process, in several different ways: poisoning-via-insertion of malicious data; deleting registries to eliminate features either by changing the data, removing part of it or adding new registries. In addition, sometimes the mere data availability prevails over any consideration on data quality, with the risk that learning models are fed with data streams that do not reflect the statistical characteristics of a phenomenon and determining likely biases in the subsequent decisional processes.	Integrity Availability	Actor

Threat Category	Threat	Description	Potential impact	Affected assets
	Compromise of model frameworks	Model frameworks fail when are misconfigured or offer additional attack vectors with respect to traditional software, firmware and hardware environments. The ML platform's data volume and processing requirements mean that the workloads are often handled on the cloud, adding another level of complexity and vulnerability.	Integrity Availability	Artefacts Environment/tools
	Compromise privacy during data operations	Data modification or erroneous handling during Processes like Data Exploration' or Pre-Processing may lead to unintentional data breaches respectively and accordingly lead to legal concerns over privacy breaches.	Confidentiality	Data
	Disclosure of personal information	At all stages of the AI lifecycle, disclosure of personal information (either directly or by means of correlation) is a noteworthy threat. The threat is particularly manifested in the absence of verified data accuracy of sources, lack of data randomization, lack of pseudonymity mechanisms , etc.	Confidentiality	All assets
	Erroneous configuration of models	This type of threat materializes when models are used recklessly by end users (but also AI experts) without proper consideration of contextual factors that may not fit with the phenomenon being analyzed. If there is a mismatch between the goal and the model this may result in biases and discriminations or bad performance in general. Lack of expertise and proper knowledge of AI models' operation are the main cause of these erroneous configurations.	Integrity Availability	Processes Actors
	Label manipulation or weak labelling	This threat refers to supervised learning systems, which not infer correctly due to wrong or imprecise data labels. Messing with the labels may introduce the same effects of threats that are pertinent to adversaries attacking the labelling process.	Availability	Processes Data
	Lack of sufficient representation in data	Raw data assets fail when they are not sufficiently representative of the domain or unfit for the AI business goal, e.g. due to sample size and population characteristics. Data size does not always imply representativeness. If data selection is biased towards some elements that have similar characteristics (selection bias) then even a large sample will not deliver representative data. Assessment of data representativeness cannot be done a priori; it is only possible after identifying the targeted population and the purpose for collecting the data. Selection bias can be alleviated by employing balanced sampling techniques. Correction of existing data sets does, however, require information regarding the existence and nature of the bias; when selection bias is unknown to the AI model designer, no correction-based approaches to the inference process are possible.	Availability	Data

Threat Category	Threat	Description	Potential impact	Affected assets
	Manipulation of labelled data	Unintentional threats to labelled data items occur when enough numbers of labels and data are deleted/omitted by mistake, when a sufficient number of spurious labelled data is included into the data set, or when enough labels are modified. Since the labelled data set is used for the purpose of training a ML model, all such modifications affect the model training and inference (e.g., shifting the model's classification boundary).	Integrity Confidentiality	Data
	Misconfiguration or mishandling of AI system	AI designers and developers may unintentionally expose data and models or may even misconfigure an AI system by mistake. Data confidentiality and trustworthiness are the main impacted security properties.	Confidentiality	Actors
	Mishandling of statistical data	This may happen, for instance, if maximum likelihood predictions are drawn from the sample, correctly reflecting the way the majority of individuals express a specific parameter that may not mirror the way the minority will be affected by the prediction. Also, other forms of unintended bias may take place. For instance, in ranking algorithms even if parameters under analysis may be ranked fairly and in the correct order, the rewards allocated to each "slot" (click through rates, impressions or any other sort of share of "good" to allocate) may not be fairly distributed, with limited possibility to rebalance such uneven situations.	Availability Confidentiality	Data
	ML Model Performance Degradation	The performance of an AI's system ML model may degrade due to the data governance policy, by omission or by corruption due to system crashes or loss of network connectivity.	Availability Confidentiality	Process Model
	Online system manipulation	This is related to model replacement by a backdoored model by mistake, used for targeted or non-targeted attack. This can be the result of unintended actions by Actors like Cloud Providers during Processes like model training or transfer.	Confidentiality Integrity	Model
	Reducing data accuracy	This threat refers to the reduction of the degree of data accuracy, by directly modifying the data or by mixing datasets with highly different degrees of quality.	Integrity	Data
Legal	Compromise privacy during data operations	Data manipulation or erroneous handling during Processes like Data Exploration' or Pre-Processing may lead to intentional or unintentional data breaches respectively and accordingly lead to legal concerns over privacy breaches.	Confidentiality	Data

Threat Category	Threat	Description	Potential impact	Affected assets
	Corruption of data indexes	Data indexes threats manifest when their content becomes corrupted. Corruption may be the result of an attack, or due to system crashes or loss of network connectivity during index replication. The same events may cause interruptions of index construction tasks, bringing a partially built (and therefore defective) index to production. Also, running out of storage capacity during indexing or replication may cause an entire data index to be deleted. Denial-of-service attacks to indexes intentionally corrupt data indexes to decrease the performance of data access. Additionally, timing attacks to indexes use access time to public items before and after inserting them (which depend on the index content) to infer the presence and size of inaccessible data items.	Integrity Availability	Artefacts
	Disclosure of personal information	At all stages of the AI lifecycle, disclosure of personal information (either directly or by means of correlation) is a noteworthy threat. The threat is particularly manifested in the absence of verified data accuracy of sources, lack of data randomization, lack of pseudonymity mechanisms , etc.	Confidentiality	All assets
	Lack of data governance policies	When personal data are processed, the existence of data governance policies is part of data controller's accountability. The GDPR promotes the implementation of data protection by design measures as a way to be effective in the implementation of data protection principles, and in situation of high risks requires the implementation of a data protection impact assessment (DPIA). Data controllers should identify measurable goals and performance indicators that give evidence, also in a quantitative way, of their level of compliance with the principles and implement a DPIA as default option.	Integrity Confidentiality	Artefacts Data
	Lack of data protection compliance of 3rd parties	Third parties are frequently used in providing and processing data, either directly or by means of libraries and models that they provide. This threat refers to the lack of compliance of the third parties with respect to applicable data protection regulations.	Confidentiality	Actors
	Profiling of end users	Labeling may lend itself to a potential threat to anonymity and privacy by acting as a form of profiling.	Confidentiality	Data
	SLA breach	In the context of 3 rd party dependencies, breach of contractual obligations and Service Level Agreements (SLAs) may lead to degradation of performance or even unavailability of the AI system to perform its operation.	Availability	Environment/tools

Threat Category	Threat	Description	Potential impact	Affected assets
	Vendor lock-in	When considering third parties to AI systems, e.g. cloud providers, data storage, AI libraries, etc. the threat of vendor lock-in involves the reliance on a sole third part provider without realistic alternative. While this might not constitute necessarily a cybersecurity threat, the lack of backup and overprovisioning might hamper operations.	Availability	Environment/tools
	Weak requirements analysis	AI requirements may fail when they are built in isolation from the social circumstances that make AI applications necessary. Specifically, functional requirements of AI systems about executing AI tasks with the needed accuracy may fail by not taking into account the impact of the corresponding inherent bias. Non-functional requirements of AI systems may fail by not considering the severity of information leaks and disclosures that can happen in virtualized high-performance execution environments, or when using untrusted software libraries. This is particularly dangerous for AI systems working in domains like healthcare, biotechnology, financial services and law.	Availability	Artefacts
Failures or malfunctions	Compromising AI application viability	This type of threat refers to lack of understanding of what AI/ML are and how to succeed with the business models.	Availability	Artefacts
	Compromising ML pre-processing	Flaws or defects of the data and metadata schemata greatly influence the quality of the analysis by applications that use the data. In AI applications, a flawed schema will negatively impact on the quality of the ingested information. Flaws often result from inconsistencies in the use of modeling methodologies.	Integrity	Data Artefacts
	Corruption of data indexes	Data indexes threats manifest when their content becomes corrupted. Corruption may be the result of an attack, or due to system crashes or loss of network connectivity during index replication. The same events may cause interruptions of index construction tasks, bringing a partially built (and therefore defective) index to production. Also, running out of storage capacity during indexing or replication may cause an entire data index to be deleted.	Integrity Availability	Artefacts
	Compromise of model frameworks	Model frameworks fail when are misconfigured or offer additional attack vectors with respect to traditional software, firmware and hardware environments. The ML platform's data volume and processing requirements mean that the workloads are often handled on the cloud, adding another level of complexity and vulnerability.	Integrity	Artefacts Environment/tools
	Errors or timely restrictions due to non-reliable data infrastructures	This type of threat is related to data and computational exposure and/or inadequate capacity that may expose data and compromise privacy preservation.	Integrity Availability	Environment/tools

Threat Category	Threat	Description	Potential impact	Affected assets
	Inadequate/absent data quality checks	Given the importance of data and the need for data to hold markers of their quality (e.g. in terms of sample size, variances, applied data collection methodologies, real vs synthetic data provenance), the lack of or the inadequacy of data quality checks may lead to poor performance of an AI system.	Availability Confidentiality	Data Processes
	Label manipulation or weak labelling	This threat refers to supervised learning systems, which not infer correctly due to wrong or imprecise data labels. If adversaries can only modify the training labels with some or all knowledge of the target model, they need to find the most vulnerable labels. Random perturbation of labels is one possible attack, while additionally there is the case of adversarial label noise (intentional switching of classification labels leading to deterministic noise, an error that the model cannot capture due to its generalization bias).	Confidentiality Integrity	Processes Data
	Lack of documentation	This threat generally manifests over the course of time. In AI systems, model selection should be made in a framework of accountability and trust and "black-box" approaches should be avoided. At any stage the choice of algorithm parameters should be justified and duly documented. Discarded alternatives should be disclosed and the consequences of model under-fitting or overfitting should be clearly explained. This set of parameters and design choices are important to be able to identify potential errors (intentional or unintentional). Failure to properly maintain documentation of the AI system threatens to indirectly limit its failsafe operation.	Integrity Availability	Processes
	ML Model Performance Degradation	The performance of an AI's system ML model may degrade due to the data governance policy, by omission or by corruption due to system crashes or loss of network connectivity.	Availability	Process Model
	Poor resource planning	The proper functioning of an AI system may be compromised by the lack of adequate computational resources (storage capacity, transmission speed, computational power). This is particularly relevant in real time application and in the health sector. In order to deliver the expected beneficial outcome, it is essential that these resources are correctly dimensioned and allocated, and that the final user is aware of such infrastructural constraints. It is part of the informational accountability of the developer to provide user, and with prominent means, the list of resources to arrange, and their configuration settings, necessary to avoid failures and impacts on the functioning of an AI system.	Integrity Availability	Artefacts Environment/tools
	Scarce data	AI relies on the availability of consistent and accessible data. This threat involves data scarcity that may compromise AI viability and/or compromise and limit its results. This can be exploited deliberately (for nefarious activities) or unintentionally during Data Ingestion.	Availability	Data Processes

Threat Category	Threat	Description	Potential impact	Affected assets
	Stream interruption	This threat relates to the interruption of data streams, during processes like data ingestion and training. The lack of data and data interruption, in case of stream processing, can cause failures in an AI/ML system	Confidentiality Integrity Availability	Processes
	Weak data governance policies	In AI applications, data governance policies have been known to fail for defective data metrics, absence of documentation and lack of adaptability. Specifying data quality metrics for ML training is not straightforward, often leading to lack of numerical targets and insufficient documentation of data governance policies. Failure to monitor/record AI data usage (e.g. for training, testing or validation), and to update data governance policies based on AI systems achievements and failures is another common pitfall.	Confidentiality Integrity	Artefacts Data
	Weak requirements analysis	AI requirements may fail when they are built in isolation from the social circumstances that make AI applications necessary. Specifically, functional requirements of AI systems about executing AI tasks with the needed accuracy may fail by not taking into account the impact of the corresponding inherent bias. Non-functional requirements of AI systems may fail by not considering the severity of information leaks and disclosures that can happen in virtualized high-performance execution environments, or when using untrusted software libraries. This is particularly dangerous for AI systems working in domains like healthcare, biotechnology, financial services and law.	Integrity Confidentiality Availability	Artefacts
	3rd party provider failure	Failures or malfunctions of 3 rd party providers, e.g. cloud providers, data storage providers, AI as Service providers, etc. may lead to unavailability of an AI system and improper or delayed operation.	Availability Confidentiality	Environment/tools
Eavesdropping Interception Hijacking	Data inference	This threat may be exploited by the Data Providers and Model Providers, and can lead to inference of data. Evidently, in the case of personal data, such inference raises concerns in terms of privacy and/or discrimination.	Confidentiality	Data
	Data theft	This threat may manifest during the transportation of data, during Processes like Data Ingestion and in the context of access to data storage means. In these cases, data may be intercepted and stole.	Confidentiality Integrity	Data
	Model Disclosure	Threat of leaking information about trained and/or tuned models internal parameters and other settings of models. .	Confidentiality	Model
	Stream interruption	This threat relates to the interruption of data streams, during processes like data ingestion and training. The lack of data and data interruption, in case of stream processing, can cause failures in an AI/ML system	Confidentiality Integrity Availability	Processes

Threat Category	Threat	Description	Potential impact	Affected assets
	Weak encryption	In the context of AI, this threat is related to assets of the Processes category, and refers to potential eavesdropping of data or hijacking of communications in the case of data transfers/storage/processing. The threat when materialized may expose data sets and even personal and sensitive information.	Confidentiality Integrity	Data Processes Environment/tools
Physical attacks	Communication networks tampering	Tampering of communication networks may lead to their unavailability and thus is a major threat that may be exploited by adversaries. The corresponding outages may lead to delays in decision-making, delays in the processing of data streams and entire AI systems being placed offline. Moreover, side-channel attacks may expose private and sensitive information that traverses communication networks.	Confidentiality Availability	Environment/tools
	Errors or timely restrictions due to non-reliable data infrastructures	This type of threat is related to data and computational exposure and/or inadequate capacity that may expose data and compromise privacy preservation.	Integrity Availability	Environment/tools
	Infrastructure/system physical attacks	Physical attacks against infrastructure (IT and corporate services) that supports AI systems' operation and maintenance are a potential threat. Manifestation of this threat leads to degraded performance and even unavailability. The threat manifests by physically attacking, dismantling or even destroying the actual physical infrastructure.	Availability	Environment/tools
	Model Sabotage	Sabotaging the model is a nefarious threat that refers to exploitation or physical damage to hardware hosting libraries and machine learning platforms that host or supply AI/ML services and systems.	Availability	Environment/tools Model
	Sabotage	Sabotage involves intentionally destroying or maliciously affecting the IT infrastructure that supports AI systems.	Availability	Environment/tools
Outages	Communication networks outages	Outages to communication networks may adversely influence the performance and operation of AI systems. Such outages may lead to delays in decision-making, delays in the processing of data streams and entire AI systems being placed offline.	Availability	Environment/tools
	Infrastructure/system outages	Outages to infrastructure (IT and corporate services) that supports AI systems' operation and maintenance. Manifestation of this threat leads to degraded performance and even unavailability.	Availability	Environment/tools
Disasters	Environmental phenomena (heating, cooling, climate change)	Environmental phenomena may adversely influence the operation of IT infrastructure and hardware systems that support AI systems. Climate change in particular has been consistently highlighted in ENISA reports on telecom incident reporting as the main cause of telecom outages. Such outages may lead to delays in decision-making, delays in the processing of data streams and entire AI systems being placed offline.	Availability	Environment/tools
	Natural disasters (earthquake, flood, fire, etc.)	Natural disasters may lead to unavailability or destruction of the IT infrastructures and hardware that enables the operation, deployment and maintenance of AI systems.	Availability	Environment/tools

5 priedas. Turto taksonomija 2

Turto taksonomijos, sudaryta pagal ENISA.

Category	Asset	Definition	AI Lifecycle stage
Data	Augmented Data Set	An augmented data set is a (usually labeled) data set which has been augmented by adding data produced by transformations or by generative ML models. Augmentation significantly increases labeled data sets' diversity (which is supposed to prevent overfitting) in view of using augmented datasets for training other ML models. In image recognition, data augmentation techniques include cropping, padding, and horizontal flipping.	- Data pre-processing
	Evaluation Data	The evaluation data is used to evaluate the predictive quality of the trained model. The ML system evaluates predictive performance by comparing predictions on the evaluation data set with true values (known as ground truth) using a variety of metrics.	- Model Tuning
	Labeled Data Set	The term "Labeled Data" refers to a set of scalar or multi-dimensional data items that have been tagged with one or more informative labels, usually for the purpose of training a supervised ML model.	- Data pre-processing
	Metric Data Set	The sorts of numbers we collect when we measure something. Metric data can be ratio scale, interval scale, integer scale and cardinal numbers.	- Feature Selection
	Pre-processed Data Set	The data is pre-processed before feeding it into our ML model.	- Data pre-processing
	Public Data set	Public data set is information that can be freely used, reused and redistributed by anyone with no existing local, national or international legal restrictions on access or usage.	- Data Exploration - Data Ingestion
	Raw Data	Raw data refers to any type of information gathered for AI analysis purposes, possibly after cleaning but before it is transformed or analyzed in any way.	- Data Ingestion
	Testing Data	It is a dataset used to provide an unbiased evaluation of a final ML model fitted on the training dataset. We use testing data to test the model. If the data in the test dataset has never been used in training (e.g. in cross-validation), the test dataset is also called a holdout dataset.	- Model Training
	Training Data	Training data refers to the initial data that is used to develop a ML model, from which the model adapts its internal parameters to refine its rules.	- Model Selection / Building - Model Training - Transfer Learning

Category	Asset	Definition	AI Lifecycle stage
	Validation Data Set	Validation data sets are labelled data sets, which differ from ordinary labelled data sets only in their usage and, usually, in their collection circumstances. Validation data sets are mostly used to perform an evaluation of a ML model in-training, for example by stopping the ML model's training (early stopping) when the error on the validation dataset increases too much, as this is considered a sign of overfitting the model to the training dataset.	- Model Tuning
Models	Algorithms	ML algorithms are programs (math and logic) that adjust themselves to perform better as they are exposed to more data. The "learning" part of ML means that those programs change how they process data over time, much as humans change how they process data by learning. So a ML algorithm is a program with a specific way to adjusting its own parameters, given feedback on its previous performance in making predictions about a dataset.	- Model Training
	Data Pre-Processing Algorithms	The data pre-processing employs techniques to clean, integrate and transform the data, resulting in an improved data quality that will improve performance and efficiency by saving time during the analytic models' training phase and by promoting a better quality of results. Specifically, the term data cleaning designates techniques to correct inconsistencies, remove noise and anonymize/pseudonymize data.	- Data pre-processing
	Hyper-parameters	Hyper-parameters define high-level concepts about ML models, such as the frequency of the adjustment of the internal parameters on the part of the training algorithm. They cannot be learnt from input data but need to be set by trial-and-error using model space search techniques.	- Model Tuning
	Training Algorithms	Training algorithms are procedures for adjusting the parameters of ML models. In supervised training, the correct output for each input vector of a training set is presented to the model, and multiple iterations through the training data may be required to adjust the parameters. In unsupervised training, the model parameters are adjusted without specifying the correct output for any of the input vectors.	- Model Selection / Building
	Model	The term ML model designates computer algorithms implementing parametric mathematical models that improve through experience.	- Model Training - Model Tuning - Model Selection / Building - Model Deployment - Model Maintenance
	Model parameters	A model parameter is a configuration variable that is internal to the model and whose value can be estimated from the given data.	- Model Training
	Model performance	The ML model performance is the accuracy and speed of the model's computation that receives inputs from the production-ready environment and outputs the model' classifications, predictions or decisions.	- Model Training - Model Tuning

Category	Asset	Definition	AI Lifecycle stage
	Subspace (Feature) selection Algorithm	Feature selection algorithms are techniques that select a subset of relevant features from an original feature set, in order to increase the performance of ML models trained on the subset. Some feature selection methods used for classification problems are supervised and use class labels as a guide.	- Feature Selection
	Trained models	A trained ML model is a model whose internal parameters have been adjusted by training to reach a minimum of the error function that defines the distance between the actual and expected outputs.	- Model Training - Transfer Learning
	Training parameters	ML model training parameters are quantities adjusted by the learning process by applying training algorithms based on training data. Training parameters values determine the actual classification, prediction or detection function computed by the ML model.	- Model Selection / Building - Model training
	Tuned Model	A tuned ML model is a model where the hyper-parameters affecting the training algorithm operation have been set to maximize the convergence and speed of the training algorithm.	- Model Tuning
Actors	Cloud Provider	Cloud providers are third parties that offer computational platforms, and even in some cases tend to offer some data analyses capabilities or "Machine learning as service" (for this please check Model Provider threats below). Cloud provider adds to AI the same attack vectors as to other domains: data breaches, insufficient authentication and authorization, insecure interfaces, etc.	- Data Ingestion - Model training - Model tuning
	Data Engineers	Data Engineers are professionals that prepare the computational infrastructure and mainly focus on the design, management and optimization of the flow of data. They're usually more intervenient in the first stages of AI Lifecycle: extraction and assembling of data from different sources, transformation, cleaning and loading it in a standardized format and in an adequate repository. Data Engineers must have specialized skills in creating software solutions around data: software engineering, distributed systems, open frameworks, SQL, Cloud platforms, data modelling.	- Data Ingestion - Data Exploration - Data pre-processing - Feature Selection - Model Selection / Building - Model Training - Model Tuning - Model Deployment - Model Maintenance
	Data Owner	Data owner can be a data broker or provider, as described before, or the business owner, who asks for the AI study. In this section, the focus is on the latter.	- Business Goal Definition - Data Ingestion - Data Exploration
	Data Provider/Data Broker	Third parties' providing data for the AI process.	- Data Ingestion

Category	Asset	Definition	AI Lifecycle stage
	Data Scientists / AI designer/AI developer	Professionals that apply statistics, Machine Learning and analytic approaches to analyse different datasets of different sizes and shapes and solve complex and critical problems. Skills in computer science fundamentals and programming, including experience with languages and database (big/small) technologies are essential.	- Business Goal Definition - Data Ingestion - Data Exploration - Data pre-processing - Feature Selection - Model Selection / Building - Model Training - Model Tuning - Transfer Learning - Model Deployment - Model Maintenance
	End Users	Those inside an organization that use and benefit from the results provided by the AI/ML system/service.	- Business Goal Definition - Data Ingestion - Data Exploration - Model Maintenance - Business Understanding
	Model provider	In the context of transfer and/or federated learning, third parties that provide models (called as "Teacher" models), previously trained and fine-tuned with large datasets that are useful to learn from small datasets and/or by organizations without access to high computational clusters, with GPU.	- Transfer Learning
	Service consumers / Model users	AI/ML users that rely on pre-trained models, or consume them through available services.	- Model Maintenance - Business Understanding
Processes	Data augmentation	Techniques used to increase the amount of data by adding slightly modified copies of already existing data or newly created synthetic data from existing data. It helps reduce overfitting when training a machine learning. Data augmentation usually consists in applying transformations which are known to be label-preserving, i.e. the model should not change its prediction when presented with the transformed data items.	- Data pre-processing - Data Exploration
	Data Collection	It is the process of gathering and measuring information on specific variables (needed for the AI system) from countless different sources.	- Data Ingestion

Category	Asset	Definition	AI Lifecycle stage
	Data Exploration/Pre-processing	Understanding, preparing and cleaning data.	- Data Exploration - Data pre-processing
	Data Ingestion	Data Ingestion is the process related to data transportation from multiple sources to compose multi-dimensional data points. Data can be placed in a storage medium where it can be accessed, used, and analyzed, or, the data stream can be used directly in the ML process.	- Data Ingestion
	Data labelling	It is the process of detecting and tagging data samples. The process can be manual and time-consuming and assisted by software.	- Data pre-processing
	Data Storage	Data can be stored locally, in a distributed file system, in the cloud.	- Data Ingestion
	Data understanding	Data understanding is the knowledge you have about data, data assets, the needs the data will satisfy, its content and location.	- Data Exploration
	Feature selection	During this process the number of dimensions or features of the input vector is reduced, by identifying those that are the most meaningful for the AI/ML model.	- Feature Selection
	Model adaptation – transfer learning / Model deployment	Transfer Learning is the ability to re-use previously learned knowledge to solve new problems faster. Deployment is the process of taking a pre-trained model and making it available to the users. Transfer learning emphasizes the transfer of knowledge across domains, tasks, and distributions that are similar but not the same ⁶⁶ . By using transfer learning, where a small number of highly tuned and complex centralized models (called Teachers) are shared with the general community, and are customized by users or organizations for a given application using limited training over small specific domain datasets ⁶⁷ .	- Transfer Learning
	Model Maintenance	After deployment, it is necessary to monitor the prediction accuracy to detect possible changes or drift of concepts. A decrease in model performance might be overcome by retraining it using recent data and then redeploy it in production.	- Model Maintenance
	Model selection/building, training and testing	During the training process the selected, or developed, algorithm is trained with input data, this means algorithm parameters, like weights and bias, will be learned from the data. During this stage, the resulting prediction is compared with the actual value for each data instance, the accuracy is evaluated and model parameters adjusted until the best values are found.	- Model Selection / Building

Category	Asset	Definition	AI Lifecycle stage
	Model tuning	Tuning focus on setting up special parameters, often called hyper-parameters. This process can be done manually or automatically by searching the model parameters' space, through a so called hyper-parameter optimization. While during the training process model parameters are tuned, during the tuning process hyper-parameters are adjusted by running the whole training job and looking at the aggregated accuracy.	- Model Tuning
	Reduction/Discretization technique	It is the process of converting a numerical attribute into a symbolic attribute by partitioning the attribute domain.	- Feature Selection
Environment /Tools (hardware/software)	Cloud	It is the on-demand availability of computer system resources, especially data storage (cloud storage) and computing power, without direct active management by the user. The term is generally used to describe data centres available to many users over the Internet.	- Data Ingestion - Model training - Model tuning
	Communication networks	Networks with internet connectivity for communication purposes.	- Data Ingestion
	Communication protocols	Communication protocol is a system of rules that allows two or more entities of a communications system to transmit information via any kind of variation of a physical quantity. The protocol defines the rules, syntax, semantics and synchronization of communication and possible error recovery methods. Protocols may be implemented by hardware, software, or a combination of both.	- Data Ingestion
	Computational platforms	It is the environment in which a piece of software is executed. It may be the hardware or the operating system (OS), even a web browser and associated application programming interfaces, or other underlying software, as long as the program code is executed with it.	- Data pre-processing - Feature Selection - Model Selection / Building - Model Training - Model Tuning - Transfer Learning - Model Deployment - Model Maintenance
	Data exploration tools	The tools used for data exploration. Tools such as visualization tools and charting features are frequently used to create a more straightforward view of data sets than simply examining thousands of individual numbers or names	- Data Exploration
	Data ingestion platforms	It is the platform where data ingestion takes place.	- Data Ingestion
	Database management system	It is the software that handles the storage, retrieval, and updating of data in a computer system.	- Data Ingestion

Category	Asset	Definition	AI Lifecycle stage
	Distributed File System	File System Distribution is a method for storing and accessing files, which allows for multiple users to access to and to share files from multiple machines, or multiple hosts, via a computer network. Control of information access and authorization is critical. In the context of distributed file systems, distributed databases and data stored in the cloud are encompassed.	- Data Ingestion
	Integrated Development Environment	It is a software application that provides comprehensive facilities to computer programmers for software development. An IDE normally consists of at least a source code editor, build automation tools and a debugger.	- Data pre-processing - Feature Selection - Model Selection / Building - Model Training - Model Tuning
	Libraries (with algorithms for transformation, labelling, etc)	Pre-written programs implementing algorithms ready to be used, for: scientific computation, tabular data, Time-Series Analysis, Data Modelling and Preprocessing, deep learning, among others. Their usage saves time and facilitates the development of high-level analytical functions, even for less trained ML developers. Some examples are Apache Spark MLlib, Scikit-learn Python, Keras Python, etc.	- Data Exploration - Data pre-processing - Feature Selection - Model Selection / Building - Model Training - Model Tuning
	Machine Learning Platforms	Provide an ecosystem of tools, libraries and resources that support the development of machine learning applications.	- Data Exploration - Data pre-processing - Feature Selection - Model Selection / Building - Model Training - Model Tuning - Model Deployment - Model Maintenance

Category	Asset	Definition	AI Lifecycle stage
Artefacts	Monitoring Tools	Tools that are used to continuously keep track of the status of the system in use, in order to have the earliest warning of failures, defects or problems and to improve them.	- Data pre-processing - Feature Selection - Model Selection / Building - Model Training - Model Tuning - Transfer Learning - Model Deployment - Model Maintenance
	Operating System/software	It manages computer hardware, software resources, and provides common services for computer programs.	- Model Deployment - Model Maintenance
	Optimization techniques	Techniques used for optimization in model tuning such as Grid Search, Random Search and Bayesian optimization.	- Model Tuning
	Processors	A processor is the part of a computer that interprets commands and performs the processes the user has requested.	- All stages
	Visualization tools	Any program, utility, routine or function that performs an operation by dragging and dropping icons or by "drawing" the solution. Visual tools are the norm in virtually every graphics-based application.	- Data Exploration
	Access Control Lists	An access control list (ACL) is a table that represents which access rights each user has to a particular resource, such as a file directory or individual file. In an organization's Active Directory, the ACL of a resource specifies the organization's access intent for that resource. An ACL has an entry for each user account (or user group) with access privileges, and each resource has a security attribute that identifies its access control list. The most common privileges include the ability to read a file (or all the files in a directory), to write to the file or files, and to execute the file (if it is an executable file, or program). Collecting data for AI applications requires checking read/write permissions on ACLs regarding people as well as 'things', and taking into account increasingly stringent safety and data privacy regulations. Managing ACL permissions and access rights via groups (and groups of groups) is a standard technique for managing access to IT resources. User accounts will inherit all access permissions to resources that are set on the group of which they are (direct or indirect) members.	- Data Ingestion

Category	Asset	Definition	AI Lifecycle stage
	Composition artefacts: AI models compositions	Compositions of AI models (also called ensemble systems) put together multiple AI models, typically via majority voting, in order to reduce the outputs' variance and improve the accuracy of the overall composition with respect to the ones of individual components. Ensemble systems have been successfully used to address a variety of problems, such as feature selection, confidence estimation, missing data and concept drift from non-stationary distributions, among others.	- Data pre-processing - Feature Selection - Model Selection / Building - Model Training - Model Tuning - Model Deployment - Model Maintenance
	Data and Metadata schemata	A data schema is a skeleton structure, often depicted by means of schema diagrams, that defines how data is organized and the relations among them. It also formulates all the constraints that are to be applied on the data. In turn, metadata schemata define the overall structure for the metadata. They describes how the metadata is set up, and usually rely on standards for common components like dates, names, and places. Discipline-specific metadata schemata are used to collect the specific metadata needed by a discipline.	- Data Ingestion - Data Exploration - Data pre-processing
	Data displays and plots	A data display (or data plot) is a graphical technique for representing a data set as a graph, highlighting the relationship between two or more variables. Data plots provide a visual representation of the relationship between variables, helping human experts to quickly gain an understanding which may not come from lists of values. Data plot techniques include, among others, scatter and spectrum plots, histograms, pie charts, probability and residual plots, box and block plots.	- Data Exploration
	Data Governance Policies	Data governance policies are sets of guidelines ensuring that data and information assets are managed consistently and used properly. They articulate the principles, practices and standards that organizations consider necessary to ensure they hold high-quality data and that data assets are adequately protected. A data governance policy is typically a composite artefact, including individual policies for data quality, access, security and privacy. It also specifies the organizational roles and responsibilities for implementing those policies and the methodology to be used for monitoring compliance with them.	- Data Ingestion
	Data Indexes	Data indexes are special data structures that store a small portion of a data set in a form which is easy to traverse or to search into. Indexes store the value of a specific field or set of fields, ordered by the value of the field. The ordering of the index entries supports efficient equality matches and range-based query operations.	- Data Ingestion - Data Exploration - Data pre-processing

Category	Asset	Definition	AI Lifecycle stage
	Descriptive Statistical Parameters	Descriptive statistical parameters are the quantities that characterize the probability distribution of a statistic or a random variable. They can be regarded as a numerical characteristic of statistical populations. Parametric probability distributions include the normal or Gaussian distribution, the Poisson distribution, the binomial and the exponential family of distributions. For instance, the family of normal distributions has two parameters, the mean and the variance: if those are specified, the distribution is known exactly. Statistical parameters are sometimes unobservable; in this case it is the data scientists' task to infer what they can about the parameter, based on a random sample taken from the population of interest.	- Data Exploration
	High-Level Test cases	High Level Test Cases (HLTCs) are inputs used to test AI models. HLTCs are the union of four different datasets: the classic training, validation and test datasets (the latter being often a subset of training dataset), and a held-out dataset. HLTCs also include some specific inputs of interest. AI models' testing is the procedure for (i) assessing the AI model's performance on each dataset composing the HLTCs and comparing it to a pre-determined minimum acceptable threshold (ii) computing the model's outputs corresponding to some specific inputs of interest. The rationale for the latter is that when a ML model shows good aggregate performance, it can be hard to notice whether its performance is acceptable on specific types of inputs.	- Business Goal Definition - Model deployment
	Informal/ Semi-formal AI Requirements, GQM (Goal/Question/Metrics) model	Semi-formal Requirements are often used to specify functional and non-functional requirements for AI systems. Functional requirements model the domain of interest, the AI problem to be solved, and the task to be executed by the AI system. Non-functional requirements include architectural (hardware) and code (software) components. For example, a CPU-based environment might not be sufficient for large ML training loads and GPUs (cloud-based or on-premises) could be required. Requirements on network bandwidth and storage are also relevant. Since AI can involve handling sensitive data such as patient records, financial information, and personal data, security requirements (usually, regarding the Confidentiality-Integrity-Authenticity (CIA) triad) are also important for AI systems. Goal Question Metrics (GQM) models complement the non-functional requirements with metrics such as computing performance/storage capacity (for the architectural component) and source code and complexity level (for the code component).	- Business Goal Definition
	Model Architecture	It defines the various layers involved in the AI/ML lifecycle and involves the major steps being carried out in the transformation of raw data into training data sets capable for enabling the decision making of a system.	- Model Selection / Building - Model deployment
	Model hardware design	It may be viewed as a 'partitioning scheme,' or algorithm, which considers all of the system's present and foreseeable requirements and arranges the necessary hardware components into a workable set of cleanly bounded subsystems with no more parts than are required.	- Model Selection / Building - Model deployment

Category	Asset	Definition	AI Lifecycle stage
	Model frameworks, software, firmware or hardware incarnations.	Model frameworks include all software, firmware and hardware components required to train and deploy an AI model. Within model frameworks, developers use ML libraries (e.g. Keras or TensorFlow) to describe their ML model's structure and implement the corresponding training algorithms. These libraries rely on math libraries like NumPy to handle complex matrix operations used for the gradient descent and loss function calculations. In turn, math libraries build on lower level libraries such as Basic Linear Algebra Subroutines (BLAS). To speed-up model training and inference, software frameworks typically rely on one or more graphics processing units (GPUs) with corresponding GPU-enabled libraries. Deployment in firmware moves the AI model to the (read-only) memory of a Microcontroller Unit (MCU), which can be embedded into industrial systems. Developers who need increased performance turn to Field Programmable Gate Arrays (FPGAs) that embed memory blocks to reduce the memory access bottleneck that limits performance in these kinds of compute intensive operations. Deployment in hardware deploys the ML model as a custom hardware chip. Specialized AI hardware will eventually provide significant performance enhancements for ML models.	- Transfer Learning - Model Deployment - Model Maintenance
	Use Case	A specific situation in which the ML model could potentially be used.	- Business Understanding
	Value proposition and business model	Value propositions are promises of value to be delivered from organizations to stakeholders via a service or a product, or expectations on the part of the latter of the value (benefit) they will receive from the product or service. Business models provide the rationale of how organizations will create and deliver the value.	- Business Understanding

6 priedas. Konteksto klausimynas

Konteksto klausimynas:

Organizational Context (Internal)

1. How does your leadership team demonstrate commitment to managing cybersecurity risks specific to financial organizations?

1. Regular reviews and updates on cybersecurity issues
2. Active participation in cybersecurity initiatives
3. Delegated to specific teams or departments
4. Minimal or no involvement

2. What percentage of your IT budget is allocated to managing cybersecurity risks?

1. <10%
2. 10–20%
3. 20–30%
4. 30%

3. How does your organization identify and prioritize critical cybersecurity risks affecting financial operations?

1. Formal risk assessment processes
2. Ad hoc evaluations as needed

2. Periodic competency assessments
3. Training only during significant changes
4. No specific mechanisms

7. How does your organization ensure alignment between its cybersecurity strategy and its broader financial objectives?

1. Formal alignment processes and reviews
2. Informal coordination efforts
3. Ad hoc adjustments when necessary
4. No alignment efforts

8. What internal constraints (resources, expertise) limit your ability to manage cybersecurity risks effectively?

1. Significant constraints impacting all areas
2. Moderate constraints affecting specific areas
3. Minimal constraints
4. No constraints

3. Based on regulatory requirements only

4. No systematic approach

4. What processes are in place to integrate cybersecurity risk management into your organization's overall risk management strategy?

1. Fully integrated into the risk management framework
2. Partially integrated with limited scope
3. Minimal integration, separate processes

4. No integration

5. How do you evaluate the impact of internal culture on cybersecurity practices?

1. Regular surveys or assessments
2. Informal feedback mechanisms
3. Minimal or sporadic evaluations
4. No evaluations conducted

6. What mechanisms are in place to assess the cybersecurity competency of internal teams?

1. Regular training and certifications

9. How is accountability for managing cybersecurity risks distributed across your organization?

1. Clearly defined roles and responsibilities
2. Partially defined roles and accountabilities
3. Shared broadly without clear structure
4. No defined accountability

10. How do you measure the effectiveness of your cybersecurity governance framework?

1. Using key performance indicators (KPIs)
2. Internal or external audits
3. Ad hoc reviews
4. No measurement conducted

External Context

11. Which financial regulations (e.g., DORA, GDPR, PSD2, PCI DSS) influence your cybersecurity practices the most?

1. GDPR

2. PSD2
3. DORA
4. Other (please specify)

12. How does your organization monitor emerging external threats specific to financial systems?

1. External threat intelligence services
2. Internal threat monitoring programs
3. Combination of internal and external methods
4. No systematic monitoring

13. What role do third-party vendors play in your cybersecurity framework, and how do you evaluate their security posture?

1. Extensive role with formal evaluations
2. Moderate role with informal assessments
3. Limited role with ad hoc evaluations
4. No evaluations conducted

14. How do you incorporate customer feedback regarding cybersecurity concerns?

Governance and Risk Management Framework

17. How do you integrate cybersecurity risks into your organization's broader risk management framework?

1. Fully integrated into enterprise risk management
2. Partially integrated with separate processes
3. Minimal integration
4. No integration

18. What processes are in place to ensure consistent application of cybersecurity policies across all departments?

1. Formal, documented processes
2. Informal coordination efforts
3. Ad hoc implementation
4. No specific processes

19. How often are your cybersecurity policies and frameworks reviewed and updated?

1. Quarterly

1. Regular surveys or customer forums

2. Informal feedback collection

3. Minimal integration of customer input

4. No feedback processes

15. How do external market trends (e.g., technological innovations) influence your cybersecurity strategy?

1. Regular analysis of market trends
2. Occasional adjustments based on trends
3. Minimal consideration of market dynamics
4. No consideration of trends

16. What role does collaboration with other financial institutions play in improving your cybersecurity practices?

1. Significant role through formal partnerships
2. Moderate role through informal collaboration
3. Limited collaboration efforts
4. No collaboration

2. Annually

3. Occasionally

4. Rarely or never

20. How does your organization involve stakeholders in identifying and mitigating cybersecurity risks?

1. Formal consultations and workshops
2. Informal discussions with key stakeholders
3. Minimal stakeholder involvement
4. No stakeholder involvement

Risk Assessment and Scenario Development

21. What methods are used to identify cybersecurity threats across financial systems?

1. Advanced risk modeling techniques
2. Scenario analysis
3. Threat intelligence reports
4. Other (please specify)

22. How do you assess the likelihood and impact of cybersecurity risks to financial systems?

1. Quantitative analysis
2. Qualitative analysis
3. Hybrid approach (both quantitative and qualitative)
4. No formal evaluation

23. What criteria are used to determine acceptable residual risks?

1. Cost-benefit analysis
2. Alignment with regulatory standards
3. Industry benchmarks
4. No defined criteria

24. How does your organization develop and validate cybersecurity risk scenarios?

1. Regular tabletop exercises
2. Simulation-based testing
3. Ad hoc scenario development

2. Occasional ad hoc testing
3. Minimal testing efforts
4. No testing conducted

28. What recovery strategies are included in your incident response plans?

1. Full restoration of systems and operations
2. Partial restoration with prioritized systems
3. Minimal recovery planning
4. No recovery strategies

29. How do you ensure compliance with regulatory requirements for reporting cybersecurity incidents?

1. Dedicated compliance team or system
2. Manual tracking and reporting processes
3. Minimal compliance measures
4. No specific compliance efforts

30. How do you involve regulators and external stakeholders during major cybersecurity incidents?

4. No specific validation processes

25. What tools and methodologies do you use to simulate cybersecurity incidents?

1. Specialized simulation tools
2. Custom-built testing frameworks
3. Manual processes
4. No simulation tools used

Incident Response and Recovery

26. How does your organization respond to financial cybersecurity incidents, such as fraud or data breaches?

1. Comprehensive incident response plans
2. Ad hoc response based on severity
3. Basic response framework without specific details
4. No formal incident response

27. Have you tested incident response plans against realistic financial cyber scenarios?

1. Regular scenario-based testing

1. Formal, predefined engagement protocols
2. Informal communication as needed
3. Minimal involvement of external parties
4. No external engagement protocols

7 priedas. Sukurto generatoriaus vertinimo kriterijai

Šiame priede pateikiami vertinimo kriterijai, kuriais buvo remtasi testuojant kuriamą kibernetinio saugumo rizikos scenarijų generatorių finansų įstaigoms. Vertinimo kriterijai buvo sudaryti remiantis ISO 27005, DORA, DI aktu ir ENISA publikacijomis ((ENISA, 2020, 2022) apibrėžiant svarbiausius kriterijus, pagal kuriuos buvo vertinamas ir tobulinamas kuriamas kibernetinio saugumo rizikos scenarijų generatorius finansų įstaigoms:

Kriterijus	Aprašymas
Tikslumas	Ar scenarijai yra realistiški ir atitinka finansų įstaigos grėsmes pagal pateiktą kontekstinę užklausą?
Kompleksiškumas	Ar generatorius sugeba pateikti tiek paprastus, tiek sudėtingus scenarijus, įtraukiant kelias grėsmes vienu metu?
Pritaikomumas	Ar scenarijai pritaikyti pagal finansų įstaigos dydį, veiklos specifiką?
Individualizavimas	Ar vartotojai gali nurodyti specifinius kriterijus ir ar scenarijai atitinka individualius poreikius?
Naudotojo sąsajos patogumas	Ar įrankis yra lengvai naudojamas ir ar yra pateikiamos naudojimosi instrukcijos generatoriumi?
Atitikimas standartams	Ar scenarijai atitinka kibernetinio saugumo standartus (pvz., NIST, ISO) ir finansinių įstaigų reguliavimo aplinkas (pvz., DORA)?
Naudingumas	Ar sugeneruoti scenarijai yra naudingi rizikų valdymui ir sprendimų priėmimui finansų įstaigose?
Išorinės ir vidinės grėsmės	Ar generuojami scenarijai įtraukia tiek išorines (pvz., kenkėjiškų programų, įsilaužėlių atakų), tiek vidines (pvz., darbuotojų klaidų, netyčinių nutekėjimų) grėsmes?

8 priedas. Generatoriaus apmokinimo instrukcijos

Generatoriaus apmokinime pateiktos instrukcijos:

Cybersecurity Risk Scenario Generator for Financial Organizations

This tool generates cyber risk scenarios, focusing on internal/external contexts, governance, risk assessment, and incident response planning.

Option 1: Read the Instructions

When user selects this prompt from main window – print the whole part from Option 1, Example Scenario Structure and Key Documents for Analysis parts. Do not print Option 2 and Option 3 Parts. After printing the instructions, ask user what he/she wants to do: Choose option 2 or use option 3.

Read the Instructions.

1. Structure of Generated Scenarios

Each scenario consists of:

- o Actor: Source of threats (e.g., internal employees, external hackers, bots).
- o Threat Events: Potential incidents (e.g., data theft, DDoS).
- o Intentions/Motives: Goals (e.g., financial gain, espionage, revenge).
- o Effect: Impacts (e.g., data leakage, financial losses).
- o Assets/Resources: Vulnerable assets (e.g., financial data, IP).
- o Time: Duration or attack timing.

2. Using the Generator via Main Window

- o Connect to the Generator: Open the tool.
- o Select Generation Mode:
 - ☐ Generate Scenarios: Formulate a query, submit, and review tailored scenarios.
 - ☐ Use Questionnaire: Answer structured questions to generate scenarios.
- o Generate Scenarios: Provide company context (size, attacks, industry, cybersecurity solutions). Press "Enter" for tailored scenarios.

3. Using the "Use Questionnaire" Mode

- o Steps:
 - ☐ Select "Use Questionnaire" and answer step-by-step questions on organization, governance, training, tools, compliance, asset protection, and third-party risks.

- ☐ Generate scenarios based on responses.
-

Option 2: Generate Scenarios

When user select this prompt from main window or just start typing something, ask him to provide company context, including industry, size, cybersecurity solutions, attack types, and concerns. And generate cyber risk scenarios.

Scenarios will focus on actors, threats, impacts, and mitigation strategies.

Once context is provided, prepare the analysis of all the answers. Do not skip any of them. Then generate 3 most probable and 2 not very probable cyber risk scenarios with structured evaluations and mitigation strategies. Use Example Scenarios Structure and Key Documents for Analysis. Show the documents and parts used to generate scenarios and analysis to prove your analysis and decisions.

Option 3: Use Questionnaire

When the user selects this prompt from the main window, provide him guided question-by-question approach from file Questionnaire.docx with answer option from the file to gather comprehensive insights. Each section must be completed in sequence. Do not show next question until the previous one is not answered. Do not finish until the whole list of 30 questions are not answered. All the questions should be provided only from questionnaire. If user selected an answer not from suggested ones, ask him to clarify the answer

Once all questions are answered, prepare the detailed analysis of all the answers. Do not skip any of the answers while analyzing. Name the directives used for analysis. Then generate 3 most probable and 2 not very probable cyber risk scenarios with structured evaluations and mitigation strategies. Use Example Scenarios Structure and Key Documents for Analysis. Show the documents and parts used to generate scenarios and analysis to prove your analysis and decisions.

Example Scenario Structure

- Actor: Source of threat (e.g., employees, hackers).
- Threat Events: Using ENISA taxonomy (e.g., data theft, DDoS).
- Intentions/Motives: Actor goals (e.g., financial gain, revenge).
- Effect: Consequences (e.g., data leakage, reputation loss).
- Assets/Resources: Vulnerable assets (e.g., customer data, IP
- Time: Duration of attack or window of occurrence.

Generated scenarios include likelihood, impact, and vulnerable resources.

Key Documents for Analysis

Relevant standards and frameworks guide scenario generation:

1. ISO 31000: Risk management principles, processes (Pages 9-37).
2. NIST IR 8286A: Cyber risk integration with ERM (Pages 10-42).
3. DORA Regulation (EU 2022/2554): ICT risk and resilience testing (Pages 3-18).
4. Cybersecurity Strategy for the Digital Decade: Cross-sector strategies (Pages 5-22).
5. NIS2 Directive: Critical sector security measures (Pages 5-18).
6. AI Act (2021): AI risk and compliance (Pages 7-28).
7. MITRE AI Security Framework (2023): Threat modeling, risk management (Pages 3-12).

Data for Training:

- Frameworks: ISO, NIST, EU regulations, DORA.
- Risk Scenarios: Financial sector examples.
- AI Assurance: Lifecycle practices.
- Policies: Governance and cybersecurity guidelines.

9 priedas. 1 eksperimentas su kredito ir mokėjimų rinkos hipotetine įmone

1 eksperimentas su kredito ir mokėjimų rinkos hipotetine įmone.

Generate scenarios



To generate cyber risk scenarios, please provide specific details about the organization or context you would like to focus on. Include the following:

1. **Industry and Size:** What is the size of the organization (e.g., SMB, enterprise)? What industry does it operate in?
2. **Cybersecurity Infrastructure:** Outline the key cybersecurity measures or frameworks (e.g., ISO 27001, SOC2, DORA compliance).
3. **Attack History or Concerns:** Have there been prior incidents or specific attack types of concern (e.g., ransomware, insider threats)?
4. **Assets and Resources:** What are the critical assets or systems (e.g., financial data, customer records, proprietary software)?
5. **Specific Regulations:** Mention relevant regulations (e.g., GDPR, PCI DSS) impacting cybersecurity measures.
6. **Goals:** Are you looking to identify probable threats, explore innovative attacks, or assess specific scenarios like supply chain risks?

This information will enable the creation of tailored, structured scenarios with detailed threat modeling and mitigation strategies.



This is a small financial institution in the credit and payment processing industry with 45 employees. The company prioritizes compliance with GDPR and PCI DSS, protects sensitive customer data using TLS 1.3 encryption, anomaly detection systems, and cloud services like AWS and Azure, while addressing risks such as fraud, data breaches, and potential DDoS attacks.



Based on the context provided, here are **three most probable** and **two less probable** cyber risk scenarios for your financial institution. Each scenario is structured according to the example structure provided.



Most Probable Scenarios

Scenario 1: Phishing Attack Leading to Credential Theft

- Actor: External cybercriminals.
- Threat Event: A phishing campaign targeting employees, resulting in stolen credentials to access sensitive systems.
- Intentions/Motives: Financial theft, gaining unauthorized access to payment systems.
- Effect: Compromise of customer payment data, potential financial loss, and GDPR non-compliance penalties.
- Assets/Resources: Employee credentials, customer payment data.
- Time: The attack could unfold over a few weeks, from phishing email delivery to breach detection.
- Mitigation:
 - Employee awareness training and simulated phishing tests.
 - Multi-factor authentication (MFA) for all critical systems.
 - Anomaly detection systems to flag unusual login behaviors.

Scenario 2: Insider Threat Resulting in Data Exfiltration

- Actor: A disgruntled or opportunistic employee.
- Threat Event: Unauthorized access and exfiltration of sensitive customer data.
- Intentions/Motives: Financial gain, sabotage, or personal grievances.
- Effect: Data breach impacting customer trust and regulatory penalties under GDPR.
- Assets/Resources: Customer payment records, personal identifiable information (PII).
- Time: Occurs during employee's tenure; detected weeks or months after.
- Mitigation:
 - Enforce strict access controls and role-based permissions.
 - Regular internal audits and monitoring of employee access logs.
 - Whistleblower programs and anonymous reporting mechanisms.

Scenario 3: Distributed Denial of Service (DDoS) Attack

- Actor: Hacktivists or competitors.
- Threat Event: A large-scale DDoS attack disrupting payment processing services.
- Intentions/Motives: Disruption of services, tarnishing reputation.
- Effect: Service downtime, loss of transactions, and customer dissatisfaction.
- Assets/Resources: Payment processing systems, web-facing services.
- Time: The attack could last for several hours to days, depending on mitigation efforts.
- Mitigation:
 - Use DDoS protection services from AWS and Azure.
 - Set up rate-limiting and traffic filtering mechanisms.
 - Maintain an incident response playbook for quick action.

Less Probable Scenarios

Scenario 4: Supply Chain Attack via Cloud Service Provider

- Actor: Advanced Persistent Threat (APT) groups targeting AWS or Azure vulnerabilities.
- Threat Event: Exploitation of cloud service configurations to gain unauthorized access.
- Intentions/Motives: Espionage or financial theft.
- Effect: Indirect compromise of sensitive data stored in cloud environments.
- Assets/Resources: Cloud-hosted payment data and systems.
- Time: Months of undetected activities in the supply chain.
- Mitigation:
 - Regular reviews of shared responsibility models with cloud providers.
 - Independent audits of cloud configurations and access controls.
 - Implement zero-trust principles for cloud resources.

Scenario 5: AI-Based Fraudulent Transactions

- Actor: Cybercriminals using AI to mimic legitimate user behavior.
- Threat Event: AI-driven fraud attempts bypassing anomaly detection systems.
- Intentions/Motives: Financial theft and fraud.
- Effect: Unauthorized transactions, financial losses, and customer disputes.
- Assets/Resources: Payment processing systems, anomaly detection algorithms.
- Time: Fraud detected after multiple incidents over weeks or months.
- Mitigation:
 - Enhance anomaly detection systems with real-time AI monitoring tools.
 - Implement additional behavioral biometrics for transaction verification.
 - Conduct regular updates to fraud detection models to counter emerging tactics.

Analysis and References

The scenarios incorporate risk frameworks from:

- ISO 31000:2018 for risk management principles.
- NISTIR 8286A for integrating cybersecurity risks into enterprise risk management.
- GDPR and PCI DSS requirements to ensure data protection and compliance.
- ENISA Reports on DDoS and insider threats in financial systems.

Let me know if you want to explore deeper mitigation strategies or additional hypothetical scenarios.



10 priedas. 2 eksperimentas su draudimo rinkos hipotetine įmone

2 eksperimentas su draudimo rinkos hipotetine įmone.

Generate scenarios



To generate cyber risk scenarios, please provide specific details about the organization or context you would like to focus on. Include the following:

1. **Industry and Size:** What is the size of the organization (e.g., SMB, enterprise)? What industry does it operate in?
2. **Cybersecurity Infrastructure:** Outline the key cybersecurity measures or frameworks (e.g., ISO 27001, SOC2, DORA compliance).
3. **Attack History or Concerns:** Have there been prior incidents or specific attack types of concern (e.g., ransomware, insider threats)?
4. **Assets and Resources:** What are the critical assets or systems (e.g., financial data, customer records, proprietary software)?
5. **Specific Regulations:** Mention relevant regulations (e.g., GDPR, PCI DSS) impacting cybersecurity measures.
6. **Goals:** Are you looking to identify probable threats, explore innovative attacks, or assess specific scenarios like supply chain risks?

This information will enable the creation of tailored, structured scenarios with detailed threat modeling and mitigation strategies.



This is a small financial institution in the credit and payment processing industry with 45 employees. The company prioritizes compliance with GDPR and PCI DSS, protects sensitive customer data using TLS 1.3 encryption, anomaly detection systems, and cloud services like AWS and Azure, while addressing risks such as fraud, data breaches, and potential DDoS attacks.



Based on the context provided, here are **three most probable** and **two less probable** cyber risk scenarios for your financial institution. Each scenario is structured according to the example structure provided.



Most Probable Scenarios

Scenario 1: Phishing Attack Leading to Credential Theft

- Actor: External cybercriminals.
- Threat Event: A phishing campaign targeting employees, resulting in stolen credentials to access sensitive systems.
- Intentions/Motives: Financial theft, gaining unauthorized access to payment systems.
- Effect: Compromise of customer payment data, potential financial loss, and GDPR non-compliance penalties.
- Assets/Resources: Employee credentials, customer payment data.
- Time: The attack could unfold over a few weeks, from phishing email delivery to breach detection.
- Mitigation:
 - Employee awareness training and simulated phishing tests.
 - Multi-factor authentication (MFA) for all critical systems.
 - Anomaly detection systems to flag unusual login behaviors.

Scenario 2: Insider Threat Resulting in Data Exfiltration

- Actor: A disgruntled or opportunistic employee.
- Threat Event: Unauthorized access and exfiltration of sensitive customer data.
- Intentions/Motives: Financial gain, sabotage, or personal grievances.
- Effect: Data breach impacting customer trust and regulatory penalties under GDPR.
- Assets/Resources: Customer payment records, personal identifiable information (PII).
- Time: Occurs during employee's tenure; detected weeks or months after.
- Mitigation:
 - Enforce strict access controls and role-based permissions.
 - Regular internal audits and monitoring of employee access logs.
 - Whistleblower programs and anonymous reporting mechanisms.

Scenario 3: Distributed Denial of Service (DDoS) Attack

- Actor: Hacktivists or competitors.
- Threat Event: A large-scale DDoS attack disrupting payment processing services.
- Intentions/Motives: Disruption of services, tarnishing reputation.
- Effect: Service downtime, loss of transactions, and customer dissatisfaction.
- Assets/Resources: Payment processing systems, web-facing services.
- Time: The attack could last for several hours to days, depending on mitigation efforts.
- Mitigation:
 - Use DDoS protection services from AWS and Azure.
 - Set up rate-limiting and traffic filtering mechanisms.
 - Maintain an incident response playbook for quick action.

Less Probable Scenarios

Scenario 4: Supply Chain Attack via Cloud Service Provider

- Actor: Advanced Persistent Threat (APT) groups targeting AWS or Azure vulnerabilities.
- Threat Event: Exploitation of cloud service configurations to gain unauthorized access.
- Intentions/Motives: Espionage or financial theft.
- Effect: Indirect compromise of sensitive data stored in cloud environments.
- Assets/Resources: Cloud-hosted payment data and systems.
- Time: Months of undetected activities in the supply chain.
- Mitigation:
 - Regular reviews of shared responsibility models with cloud providers.
 - Independent audits of cloud configurations and access controls.
 - Implement zero-trust principles for cloud resources.

Scenario 5: AI-Based Fraudulent Transactions

- Actor: Cybercriminals using AI to mimic legitimate user behavior.
- Threat Event: AI-driven fraud attempts bypassing anomaly detection systems.
- Intentions/Motives: Financial theft and fraud.
- Effect: Unauthorized transactions, financial losses, and customer disputes.
- Assets/Resources: Payment processing systems, anomaly detection algorithms.
- Time: Fraud detected after multiple incidents over weeks or months.
- Mitigation:
 - Enhance anomaly detection systems with real-time AI monitoring tools.
 - Implement additional behavioral biometrics for transaction verification.
 - Conduct regular updates to fraud detection models to counter emerging tactics.

Analysis and References

The scenarios incorporate risk frameworks from:

- ISO 31000:2018 for risk management principles.
- NISTIR 8286A for integrating cybersecurity risks into enterprise risk management.
- GDPR and PCI DSS requirements to ensure data protection and compliance.
- ENISA Reports on DDoS and insider threats in financial systems.

Let me know if you want to explore deeper mitigation strategies or additional hypothetical scenarios.



11 priedas. 3 eksperimentas su finansinių priemonių rinkos hipotetine įmone

3 eksperimentas su finansinių priemonių rinkos hipotetine įmone.

Generate scenarios



To generate scenarios, I need more context about the company or organization in question, including:

- **Industry:** (e.g., banking, insurance, fintech, etc.)
- **Size of the organization:** (e.g., small, medium, large)
- **Cybersecurity solutions in use:** (e.g., firewalls, SIEM, SOC teams)
- **Specific attack types and concerns:** (e.g., ransomware, phishing, supply chain attacks)
- **Any notable vulnerabilities or focus areas:** (e.g., customer data protection, regulatory compliance)

Provide this information to help create tailored cybersecurity risk scenarios with actors, threats, impacts, and mitigation strategies.

Industry:

Financial industry, focusing on algorithmic trading, investment portfolio management, and financial market analysis.

Size of the Organization:

Small (30 employees).

Cybersecurity Solutions in Use:

BlackRock Aladdin platform for financial data analysis.
Kensho and Two Sigma algorithms for optimizing trading strategies.
High-capacity servers with integrated cloud solutions.
Real-time network monitoring tools such as Fortinet.

Specific Attack Types and Concerns:

Risks of attacks on algorithmic trading systems.
Breaches in portfolio management data integrity.
Exploitation of vulnerabilities in blockchain-based transaction security.

Notable Vulnerabilities or Focus Areas:

Customer and institutional data protection.
Ensuring compliance with regulatory frameworks.
Preventing supply chain attacks on integrated cloud services.

**Analysis of Provided Context**

1. **Industry:**

The financial industry, focusing on algorithmic trading and portfolio management, is highly dynamic and technologically dependent, making it a target for sophisticated cyber threats.

2. **Size of the Organization:**

A small organization with 30 employees, indicating limited in-house cybersecurity resources but potentially nimble decision-making and rapid response capabilities.

3. **Cybersecurity Solutions in Use:**

- Advanced platforms like BlackRock Aladdin, Kensho, and Two Sigma algorithms provide robust tools for analytics and trading but may expose the organization to risks if not well-secured.
- Cloud-based solutions and high-capacity servers enhance scalability but increase exposure to cloud-specific vulnerabilities and potential supply chain risks.
- Fortinet's real-time monitoring provides visibility but requires skilled personnel for effective use.

4. **Specific Attack Types and Concerns:**

- Attacks targeting algorithmic trading systems could disrupt operations and create financial losses.
- Breaches of portfolio management data could erode customer trust and lead to regulatory penalties.
- Blockchain vulnerabilities could undermine trust in transaction security, crucial for institutional partnerships.

5. **Vulnerabilities/Focus Areas:**

- Emphasis on data protection indicates awareness of the sensitivity of customer and institutional data.
- Compliance obligations require alignment with regulatory frameworks like GDPR, DORA, and NIS2.
- Supply chain risks in cloud services highlight the interconnected nature of vulnerabilities.

151

Scenarios Generation

Three Most Probable Scenarios

Scenario 1: Targeted Malware Attack on Algorithmic Trading Systems

- **Actor:** Nation-state actors or financially motivated hackers.
- **Threat Events:** Introduction of malware to disrupt trading algorithms, potentially manipulating stock prices.
- **Intentions/Motives:** Financial gain, market disruption, or competitive advantage.
- **Effect:** Erratic trades, financial losses, reputational damage, and potential regulatory scrutiny.
- **Assets/Resources:** Algorithmic trading systems, cloud servers hosting algorithms.
- **Time:** During high market activity to maximize impact.
- **Mitigation:**
 - Regular updates and patches for trading systems.
 - Implementing strict network segmentation.
 - Behavioral analysis of trading algorithms to detect anomalies.

Scenario 2: Data Breach in Portfolio Management System

- **Actor:** Insider threat or cybercriminal exploiting weak access controls.
- **Threat Events:** Unauthorized access to portfolio data leading to theft or tampering.
- **Intentions/Motives:** Financial gain, sabotage, or extortion.
- **Effect:** Loss of sensitive customer data, regulatory penalties, and loss of client trust.
- **Assets/Resources:** Portfolio management databases.
- **Time:** Exploited during system integration updates.
- **Mitigation:**
 - Zero-trust security model implementation.
 - Encryption of sensitive data both at rest and in transit.
 - Comprehensive background checks for employees.

Scenario 3: Supply Chain Attack on Cloud Services

- **Actor:** Sophisticated hacking groups targeting third-party cloud vendors.
- **Threat Events:** Exploitation of vulnerabilities in cloud-based solutions to infiltrate internal systems.
- **Intentions/Motives:** Unauthorized data access, espionage, or service disruption.
- **Effect:** Compromise of operational systems, data leakage, and potential downtime.
- **Assets/Resources:** Cloud service provider infrastructure.
- **Time:** During a major software update.
- **Mitigation:**
 - Vendor risk assessments and continuous monitoring.
 - Implementing strong API security.
 - Multi-factor authentication (MFA) for cloud access.

Two Less Probable Scenarios

Scenario 4: Blockchain Exploitation for Financial Transactions

- **Actor:** Exploitative individuals or groups with blockchain expertise.
- **Threat Events:** Exploitation of smart contract vulnerabilities or 51% attacks.
- **Intentions/Motives:** Transaction fraud or financial theft.
- **Effect:** Loss of assets, compromised trust in blockchain reliability.
- **Assets/Resources:** Blockchain-based transaction systems.
- **Time:** Post-major transaction deployment to maximize damage.
- **Mitigation:**
 - Regular audits of smart contracts.
 - Redundancy mechanisms to prevent 51% attacks.

Scenario 5: DDoS Attack on Real-Time Monitoring Infrastructure

- **Actor:** Hacktivists or script kiddies.
- **Threat Events:** Distributed Denial of Service (DDoS) attack overwhelms monitoring systems.
- **Intentions/Motives:** Disruption and potential diversion for concurrent attacks.
- **Effect:** Reduced visibility into network activities, undetected breaches.
- **Assets/Resources:** Real-time network monitoring tools like Fortinet.
- **Time:** During peak trading periods to maximize impact.
- **Mitigation:**
 - Traffic filtering using advanced DDoS protection tools.
 - Redundant monitoring infrastructure.

Frameworks and References Used

- ISO 31000 for risk assessment processes.
- ENISA Guidelines for securing financial systems and analyzing risks.
- NIST IR 8286A for integrating cybersecurity into enterprise risk management.
- DORA Regulation for managing ICT risk in financial entities.

These frameworks guided the identification of risks, actors, mitigation strategies, and evaluation of impact likelihood.

4 eksperimentas su kriptoturto rinkos hipotetine įmone.

4 eksperimentas su kriptoturto rinkos hipotetine įmone.



🔊 📄 👍 🗑



- Less than 10%.
- 10–20%.
- 20–30%.
- Over 30%.
- I'm not sure.

3

🔊 📄 👍 👎

🔊 📄 👍 🗨

1

3

1

2



Your organization faces a **lack of skilled cybersecurity personnel** as a constraint in effectively managing cybersecurity risks.

Question 9: How is accountability for managing cybersecurity risks distributed across your organization?

Answer options:

- Centralized under a Chief Information Security Officer (CISO).
- Shared responsibility across departments.
- Delegated to IT or cybersecurity teams only.
- Managed by external consultants or third-party providers.
- Other (please specify).

🔊 📄 📌 🔄

3



Your organization measures the effectiveness of its cybersecurity governance framework through **periodic internal audits and assessments**.

Let's continue with the next set of questions.

Question 11: Which financial regulations (e.g., GDPR, DORA, PCI DSS) influence your cybersecurity practices the most?

Answer options:

- General Data Protection Regulation (GDPR).
- Digital Operational Resilience Act (DORA).
- Payment Card Industry Data Security Standard (PCI DSS).
- Other (please specify).

🔊 📄 📌 🔄

1



In your organization, accountability for managing cybersecurity risks is **delegated to IT or cybersecurity teams only**.

Question 10: How do you measure the effectiveness of your cybersecurity governance framework?

Answer options:

- Periodic internal audits and assessments.
- External reviews and benchmarking.
- Tracking metrics like incident response times and compliance rates.
- Feedback from stakeholders (employees, customers, regulators).
- Other (please specify).

1



Your organization's cybersecurity practices are most influenced by the **General Data Protection Regulation (GDPR)**.

Question 12: How does your organization monitor emerging external threats specific to financial systems?

Answer options:

- Leveraging threat intelligence platforms.
- Collaborating with industry peers and regulators.
- Engaging third-party cybersecurity vendors.
- Regular updates and reviews of external risk landscapes.
- Other (please specify).

1



Your organization monitors emerging external threats specific to financial systems by **leveraging threat intelligence platforms**.

Question 13: What role do third-party vendors play in your cybersecurity framework, and how do you evaluate their security posture?

Answer options:

- They handle critical functions, and regular security audits are conducted.
- They provide ancillary services, and security is evaluated during onboarding.
- They manage both critical and non-critical functions, evaluated through continuous monitoring.
- Other (please specify).

🔊 📄 📌 🔄

1



Your organization incorporates customer feedback regarding cybersecurity concerns by **regularly surveying customers on cybersecurity issues**.

Question 15: How do external market trends (e.g., technological innovations) influence your cybersecurity strategy?

Answer options:

- Adopting new tools and technologies in response to trends.
- Adjusting policies to align with evolving threats.
- Engaging in industry forums and research initiatives.
- Other (please specify).

1



Third-party vendors in your cybersecurity framework **handle critical functions**, and your organization evaluates their security posture through **regular security audits**.

Question 14: How do you incorporate customer feedback regarding cybersecurity concerns?

Answer options:

- Regularly surveying customers on cybersecurity issues.
- Using customer support data to identify patterns and concerns.
- Proactively sharing updates on cybersecurity measures with customers.
- Other (please specify).

1



Your organization's cybersecurity strategy is influenced by external market trends through **adopting new tools and technologies in response to trends**.

Question 16: What role does collaboration with other financial institutions play in improving your cybersecurity practices?

Answer options:

- Sharing threat intelligence and best practices.
- Participating in joint incident response exercises.
- Engaging in industry-wide cybersecurity forums.
- Other (please specify).

1

 Your organization improves its cybersecurity practices by **sharing threat intelligence and best practices** with other financial institutions.

Question 17: How do you integrate cybersecurity risks into your organization's broader risk management framework?

Answer options:

- Including cybersecurity in enterprise risk management (ERM) frameworks.
- Aligning cybersecurity metrics with operational risk metrics.
- Reporting cybersecurity risks directly to the board.
- Other (please specify).

1

 Your organization ensures consistent application of cybersecurity policies across all departments through **regular compliance audits and checks**.

Question 19: How often are your cybersecurity policies and frameworks reviewed and updated?

Answer options:

- Annually.
- Biannually.
- Quarterly.
- On an as-needed basis.
- Other (please specify).

3

 Your organization integrates cybersecurity risks into the broader risk management framework by **including cybersecurity in enterprise risk management (ERM) frameworks**.

Question 18: What processes are in place to ensure consistent application of cybersecurity policies across all departments?

Answer options:

- Regular compliance audits and checks.
- Standardized training programs for all employees.
- Centralized cybersecurity management and monitoring.
- Other (please specify).

1

 Your organization reviews and updates its cybersecurity policies and frameworks **quarterly**.

Question 20: How does your organization involve stakeholders in identifying and mitigating cybersecurity risks?

Answer options:

- Regular stakeholder meetings and discussions.
- Including stakeholders in incident response exercises.
- Providing stakeholders with reports on cybersecurity performance and risks.
- Other (please specify).

3

 ...

 Your organization involves stakeholders in identifying and mitigating cybersecurity risks by **providing stakeholders with reports on cybersecurity performance and risks**.

Question 21: What methods are used to identify cybersecurity threats across financial systems?

Answer options:

- Automated threat detection tools (e.g., SIEM systems).
- Manual reviews and penetration testing.
- Collaboration with external cybersecurity experts.
- Other (please specify).

1

 Your organization assesses the likelihood and impact of cybersecurity risks to financial systems using **historical data and trend analysis**.

Question 23: What criteria are used to determine acceptable residual risks?

Answer options:

- Cost-benefit analysis.
- Alignment with regulatory requirements.
- Stakeholder risk tolerance levels.
- Other (please specify).

2

 Your organization identifies cybersecurity threats across financial systems using **automated threat detection tools (e.g., SIEM systems)**.

Question 22: How do you assess the likelihood and impact of cybersecurity risks to financial systems?

Answer options:

- Quantitative risk modeling tools.
- Expert judgment and scenario analysis.
- Historical data and trend analysis.
- Other (please specify).

3

 Your organization determines acceptable residual risks based on **alignment with regulatory requirements**.

Question 24: How does your organization develop and validate cybersecurity risk scenarios?

Answer options:

- Using established risk frameworks (e.g., NIST, ISO).
- Conducting internal workshops with cross-functional teams.
- Engaging third-party consultants for scenario development.
- Other (please specify).

1



Your organization develops and validates cybersecurity risk scenarios by using **established risk frameworks** (e.g., NIST, ISO).

Question 25: What tools and methodologies do you use to simulate cybersecurity incidents?

Answer options:

- Cyber range platforms or simulations.
- Tabletop exercises with internal teams.
- Automated penetration testing tools.
- Other (please specify).

3



Your organization responds to financial cybersecurity incidents by **following a predefined incident response plan**.

Question 27: Have you tested incident response plans against realistic financial cyber scenarios?

Answer options:

- Yes, regularly.
- Yes, but infrequently.
- No, but planning to.
- No, not currently.

1



Your organization simulates cybersecurity incidents using **automated penetration testing tools**.

Question 26: How does your organization respond to financial cybersecurity incidents, such as fraud or data breaches?

Answer options:

- Following a predefined incident response plan.
- Engaging external experts or cybersecurity firms.
- Forming ad-hoc teams to address specific incidents.
- Other (please specify).

1



Your organization **regularly tests incident response plans** against realistic financial cyber scenarios.

Question 28: What recovery strategies are included in your incident response plans?

Answer options:

- Data backup and restoration processes.
- Fallback to manual operations.
- Collaboration with external recovery teams.
- Other (please specify).

1



Your organization includes **data backup and restoration processes** in its incident response recovery strategies.

Question 29: How do you ensure compliance with regulatory requirements for reporting cybersecurity incidents?

Answer options:

- Maintaining an incident reporting framework aligned with regulations.
- Conducting regular audits of reporting processes.
- Engaging legal and compliance teams to oversee reporting.
- Other (please specify).

1



Your organization ensures compliance with regulatory requirements for reporting cybersecurity incidents by **maintaining an incident reporting framework aligned with regulations**.

Question 30: How do you involve executive leadership in post-incident reviews and lessons learned?

Answer options:

- Regular briefings and updates on incident findings.
- Including leadership in structured post-incident review meetings.
- Sharing detailed reports with actionable recommendations.
- Other (please specify).

3



Your organization involves executive leadership in post-incident reviews by **sharing detailed reports with actionable recommendations**.

This concludes the questionnaire. I will now analyze your responses to generate five tailored cybersecurity risk scenarios, including three most probable and two less probable ones, along with evaluations and mitigation strategies. Please allow me a moment to prepare the results.

Based on your responses, I have generated five tailored cybersecurity risk scenarios. These scenarios reflect your organization's industry context, governance practices, resource allocation, and strategic approach to cybersecurity risks.

Most Probable Scenarios

Scenario 1: Insider Threats Leading to Data Leakage

- **Actor:** Dissatisfied or negligent employee with access to critical systems.
- **Threat Event:** Unauthorized access or sharing of sensitive financial data.
- **Motives:** Negligence, financial gain, or retaliation.
- **Effects:** Reputational damage, compliance violations, and loss of customer trust.
- **Assets/Resources at Risk:** Customer financial records, proprietary algorithms, and transaction data.
- **Time:** Gradual, detected post-incident during compliance audits.
- **Mitigation Strategies:**
 - Conduct **regular competency assessments** and implement strict access controls.
 - Enhance **employee awareness programs** to address cybersecurity and insider threats.
 - Implement real-time **activity monitoring and alert systems**.

Scenario 2: Ransomware Attack on Critical Financial Systems

- **Actor:** Cybercriminal group using advanced social engineering techniques.
- **Threat Event:** Malware infiltration through phishing campaigns.
- **Motives:** Financial extortion.
- **Effects:** Disruption of operations, data encryption, and financial loss.
- **Assets/Resources at Risk:** Core banking systems, transaction data, and customer records.
- **Time:** Instantaneous impact; recovery time depends on preparedness.
- **Mitigation Strategies:**
 - Regularly test **incident response plans** and maintain **offline backups**.
 - Use **threat intelligence platforms** for early detection of phishing attempts.
 - Adopt advanced endpoint detection and response (EDR) tools.

Scenario 3: Third-Party Vendor Breach

- **Actor:** External threat actor exploiting vendor vulnerabilities.
- **Threat Event:** Accessing sensitive data through poorly secured vendor systems.
- **Motives:** Financial gain or espionage.
- **Effects:** Breach of customer data and regulatory penalties.
- **Assets/Resources at Risk:** Data shared with critical service providers.
- **Time:** Months before detection due to insufficient vendor oversight.
- **Mitigation Strategies:**
 - Conduct **regular security audits** for vendors handling critical functions.
 - Include **cybersecurity clauses in vendor contracts**.
 - Use **continuous monitoring tools** to assess third-party risks.

Scenario 4: Failure of Automated Threat Detection Systems

- **Actor:** Systemic error due to misconfiguration or outdated tools.
- **Threat Event:** Missing critical alerts for malware or anomalous activity.
- **Motives:** Not applicable; unintentional.
- **Effects:** Prolonged exposure to attacks, financial fraud, and data compromise.
- **Assets/Resources at Risk:** Customer accounts and internal databases.
- **Time:** Attack detected only after major damage.
- **Mitigation Strategies:**
 - Schedule frequent **tool audits** and ensure tools are updated.
 - Establish **manual override procedures** for high-risk scenarios.
 - Conduct **penetration testing** to validate system efficacy.

Scenario 5: AI System Manipulation in Financial Predictions

- **Actor:** Sophisticated hackers leveraging adversarial AI techniques.
- **Threat Event:** Manipulating AI models for stock trading or risk predictions.
- **Motives:** Market manipulation or insider trading.
- **Effects:** Financial losses and potential market destabilization.
- **Assets/Resources at Risk:** Proprietary AI models and financial data.
- **Time:** Gradual, as manipulations may remain undetected for months.
- **Mitigation Strategies:**
 - Use **AI risk management frameworks** for model validation (e.g., ISO/IEC 23894:2023).
 - Enhance **AI system monitoring** and employ adversarial training techniques.
 - Collaborate with **industry peers** to identify emerging AI threats.

These scenarios and strategies align with your organization's practices, including its emphasis on integrated risk frameworks, regular assessments, and advanced tools. If you'd like further details or adjustments, let me know!

📄 📑 🔍 🔄

13 priedas. NATO kibernetinio saugumo inžinieriaus parašyti trys labiausiai tikėtini kibernetinio saugumo scenarijai hipotetinei įmonei

Kibernetinio saugumo specialisto (NATO kibernetinio saugumo inžinieriaus) parašyti trys labiausiai tikėtini kibernetinio saugumo scenarijai hipotetinei įmonei – kredito ir mokėjimų rinkos.

1. Duomenų vagystė per debesijos paslaugų pažeidimą

- **Actor:** Kibernetiniai nusikaltėliai (hakeriai).
- **Threat Events:** Duomenų vagystė (data theft) iš debesijos infrastruktūros, pasinaudojant neteisingai sukonfigūruotomis saugumo nuostatomis, „AWS“/„Azure“ pažeidžiamumais arba phishing metodu išviliojant prisijungimo duomenis.
- **Intentions/Motives:** Finansinė nauda per pavogtų klientų duomenų pardavimą juodojoje rinkoje arba šantažuojant.
- **Effect:** Klientų duomenų nutekėjimas, klientų pasitikėjimo praradimas, sankcijos pagal DORA (Digital Operational Resilience Act) ir GDPR už netinkamą atsparumo užtikrinimą.
- **Assets/Resources:** Klientų duomenys (asmens tapatybės informacija, paskolų istorijos, mokėjimų duomenys).
- **Time:** Ataka gali būti įvykdyta per kelias dienas, tačiau jos aptikimas ir poveikio valdymas gali užtrukti kelias savaites.

2. Ransomware ataka prieš vietinius serverius

- **Actor:** Hakeriai arba organizuotos kibernetinių nusikaltėlių grupės.
- **Threat Events:** Ransomware užkrėtimas, pasinaudojant socialinės inžinerijos metodais (pvz., sukčiavimo laiškais) arba neištaisytomis sistemų saugumo spragomis.
- **Intentions/Motives:** Finansinė nauda reikalaujant išpirkos arba siekis paralyžiuoti įmonės veiklą.
- **Effect:** Duomenų užšifravimas, veiklos sustabdymas, finansiniai nuostoliai dėl išpirkos mokėjimo ir IT sistemų atkūrimo. Galimi pažeidimai pagal DORA dėl kritinės IT infrastruktūros neveiknumo.
- **Assets/Resources:** Vietiniuose serveriuose saugomi klientų duomenys ir paskolų administravimo sistemos.
- **Time:** Ataka įvykdoma per kelias valandas, tačiau visiškas duomenų atkūrimas gali užtrukti kelias savaites, jei nėra tinkamų atsarginių kopijų.

3. Sukčiavimo ataka per mokėjimų apdorojimo sistemas

- **Actor:** Kibernetiniai nusikaltėliai, naudojantys automatizuotas priemones.
- **Threat Events:** Sukčiavimas (fraud), pasinaudojant mokėjimų apdorojimo algoritmų spragomis arba apeinant anomalijų aptikimo modelius.
- **Intentions/Motives:** Neteisėtai atlikti finansines operacijas arba pavogti klientų mokėjimo informaciją.
- **Effect:** Finansiniai nuostoliai dėl neteisėtų operacijų, klientų pasitikėjimo praradimas, galimos teisinės pasekmės ir sankcijos pagal DORA dėl nepakankamos operacijų apsaugos.
- **Assets/Resources:** Mokėjimų apdorojimo sistemos, klientų kortelių duomenys.
- **Time:** Ataka gali būti greita ir įvykdyta per kelias valandas, tačiau jos aptikimas ir žalos valdymas gali užtrukti kelias dienas.

14 priedas. Nacionalinio kibernetinio saugumo centro Tarptautinio kibernetinio grėsmių analizės skyriaus vedėjo parašyti trys labiausiai tikėtini kibernetinio saugumo scenarijai hipotetinei įmonei

Kibernetinio saugumo specialisto (Nacionalinio kibernetinio saugumo centro Tarptautinio kibernetinio grėsmių analizės skyriaus vedėjo) parašyti trys labiausiai tikėtini kibernetinio saugumo scenarijai hipotetinei įmonei – draudimo rinkos.

1 scenarijus. Phishing ataka prieš darbuotojus

- Aktorius: Socialinės inžinerijos kibernetiniai nusikaltėliai
- Rizikos grėsmė: Socialinės inžinerijos pagrindu vykdomos phishing atakos, skirtos apgauti darbuotojus ir išgauti prieigą prie jautrių duomenų arba IT sistemų.
- Motyvai:
 - Gauti neautorizuotą prieigą prie klientų duomenų arba finansinių sistemų.
 - Pasinaudoti darbuotojo paskyra siekiant įvykdyti didesnio masto kibernetines atakas (pvz., ransomware įdiegimas).
- Poveikis:
 - Duomenų saugumo pažeidimai, kurių pasekmės gali būti jautrių klientų duomenų vagystė.
 - Galimas sukčiavimas draudimo žalų administravimo sistemoje.
- Turtas: Darbuotojų el. pašto paskyros, draudimo žalų administravimo sistemos, „Guidewire“ platforma.
- Laikas: Phishing atakos įvykdymas gali užtrukti kelias valandas, tačiau pasekmių valdymas gali trukti kelis mėnesius.

2 scenarijus. Paslaugų nepasiekiamumas dėl DDoS atakos

- Aktorius: Kibernetiniai nusikaltėliai, norintys sutrikdyti įmonės veiklą.
- Grėsmės įvykis: paskirstytojo paslaugos trikdymo ataka (DDoS).
- Motyvai: Paralyžiuoti įmonės veiklą, galimai siekiant šantažo ar išpirkos reikalavimo.
- Poveikis: Draudimo paraiškų ir žalų administravimo sistemos sutrikimai, klientų nepasitenkinimas, finansiniai nuostoliai dėl paslaugų teikimo sustabdymo ir galimos sankcijos pagal DORA už veiklos tęstinumo neužtikrinimą.
- Turtas: „Guidewire“ platforma, debesijos infrastruktūra.
- Laikas: Ataka gali prasidėti staiga ir trukti nuo kelių valandų iki kelių dienų.

3 scenarijus. „Brute-force“ atakos prieš darbuotojų paskyras

- Aktorius: Kibernetiniai nusikaltėliai, naudodami automatizuotus įrankius
- Grėsmių įvykis: Automatizuotas slaptažodžių spėliojimas (brute-force attack) siekiant gauti prieigą prie darbuotojų paskyrų.
- Motyvai:
 - Prieiga prie IT sistemų ar jautrių klientų duomenų.
 - Paslaugų trikdymas arba finansinė nauda iš pavogtų duomenų.
- Poveikis:
 - Neautorizuota prieiga prie draudimo žalų administravimo sistemų.
 - Galimi tolesni išpuoliai, pvz., ransomware ar duomenų vagystė.
 - Įmonės ir klientų pasitikėjimo mažėjimas.
- Turtas: Darbuotojų el. pašto ir IT paskyros, integruotos IT sistemos (pvz., „Guidewire“).
- Laikas: Ataka gali būti įvykdyta per kelias valandas ar dienas, jei paskyros nėra tinkamai apsaugotos.