Forma patvirtinta

Mykolo Romerio universiteto rektoriaus

2025 m. birželio d. įsakymu Nr. 1I-

6 priedas

**MYKOLO ROMERIO UNIVERSITETO**

*(padalinio pavadinimas)*

*(pareigos, vardas ir pavardė)*

\_\_\_\_\_\_\_\_\_\_\_

*(data)*

**PRAŠYMO ATITIKTIES MOKSLINIŲ TYRIMŲ ETIKAI KOMITETUI 7 KLAUSIMO PRIEDAS**

Kontrolinis klausimynas parengtas pagal ALTAI (*Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*) sistemą, pritaikytą vertinti dirbtinio intelekto sistemas ir užtikrinti, kad jos atitiktų etinius ir patikimo dirbtinio intelekto principus.

|  |  |
| --- | --- |
| **1.** **Žmogiškasis veiksnys ir neteisingo vartojimo prevencija** *[Human Agency and Oversight]* | |
| 1.1. Ar DI sistema palaiko žmogaus sprendimų priėmimą ir savarankiškumą? *[Does the AI system support human agency and decision-making?]* | * Taip* * Ne* |
| 1.2. Ar galutiniai vartotojai informuojami, kad sąveikauja su DI sistema? *[Are end-users informed they are interacting with an AI system?]* | * Taip* * Ne* |
| 1.3. Ar yra įdiegti mechanizmai, kurie padeda išvengti per didelio pasitikėjimo DI sistema? *[Are there mechanisms to avoid over-reliance on the AI system?]* | * Taip* * Ne* |
| 1.4. Ar yra įdiegta neteisingo DI vartojimo priežiūra (pvz. „žmogus cikle“ - DI sistema generuoja pasiūlymus ar rezultatus, o žmogus juos peržiūri ir priima galutinį sprendimą; „žmogus stebi ciklą“ - DI sistema veikia savarankiškai, tačiau žmogus stebi jos veiklą ir gali įsikišti, jei pastebi klaidas ar netikslumus; žmogus nėra įtrauktas į kiekvieną sprendimą, bet turi galimybę koreguoti sistemos veikimą prireikus; „žmogus vadovauja“ - žmogus turi aukščiausią kontrolę ir atsakomybę už DI sistemos veiklą. Jis nustato tikslus, parametrus ir gali bet kada pakeisti arba sustabdyti sistemos veikimą)? *[Is there human oversight integrated, such as human-in-the-loop, human-on-the-loop, or human-in-command approaches?]* | * Taip* * Ne* |
| 1.5. Ar yra „stabdymo“ mechanizmai, leidžiantys saugiai nutraukti DI veikimą esant būtinybei? *[Are there "stop" mechanisms to abort AI operations when necessary?]* | * Taip* * Ne* |
| **2. Techninis patikimumas ir saugumas** *[Technical Robustness and Safety]* | |
| 2.1. Ar DI sistema buvo išbandyta dėl atsparumo galimoms atakoms ar netinkamam naudojimui (pvz., duomenų klastojimui, modelio apgaulės)? *[Has the AI system been tested for resilience against potential attacks or misuse (e.g., data poisoning, model evasion)?]* | * Taip* * Ne* |
| 2.2. Ar yra priemonės, užtikrinančios DI sistemos patikimumą ir tikslumą? *[Are measures in place to ensure the reliability and accuracy of the AI system?]* | * Taip* * Ne* |
| 2.3. Ar parengti atsarginiai veiksmų planai sistemai sugedus ar suklydus? *[Are there fallback plans in case of failures or errors in the system?]* | * Taip* * Ne* |
| 2.4. Ar DI sistema sugeba palaikyti savo našumą kintančioje aplinkoje? *[Is the AI system capable of maintaining its performance in changing environments?]* | * Taip* * Ne* |
| **3. Privatumas ir duomenų valdymas** *[Privacy and Data Governance]* | |
| 3.1. Ar sistema atitinka BDAR (ar lygiaverčių duomenų apsaugos teisės aktų) reikalavimus? *[Does the system comply with GDPR or equivalent data protection regulations?]* | * Taip* * Ne* |
| 3.2. Ar taikomos duomenų minimizavimo, pseudonimizavimo ir šifravimo praktikos? *[Are data minimization, pseudonymization, and encryption practices applied?]* | * Taip* * Ne* |
| 3.3. Ar atliktas poveikio duomenų apsaugai vertinimas (DPIA), jei tai būtina? *[Is a Data Protection Impact Assessment (DPIA) conducted, if necessary?]* | * Taip* * Ne* |
| 3.4. Ar asmens duomenys yra saugiai tvarkomi ir naudojami tik pagal numatytą tikslą? *[Is personal data processed securely and used only for its intended purpose?]* | * Taip* * Ne* |
| **4. Skaidrumas** *[Transparency]* | |
| 4.1. Ar galima atsekti DI sistemos priimtus sprendimus iki jų šaltinių duomenų ir algoritmų? *[Can decisions made by the AI system be traced back to their source data and algorithms?]* | * Taip* * Ne* |
| 4.2. Ar vartotojams yra aiškiai paaiškintos sistemos ribos, tokios kaip tikslumo ar klaidų rodikliai? *[Are the system’s limitations, such as accuracy and error rates, communicated to users?]* | * Taip* * Ne* |
| 4.3. Ar galima paaiškinti DI sprendimų priėmimo procesą paveiktoms šalims? *[Is the reasoning behind AI decisions explainable to affected parties?]* | * Taip* * Ne* |
| 4.4. Ar įdiegtos tinkamos žurnalų vedimo praktikos audito tikslams? *[Are clear logging practices established for auditability?]* | * Taip* * Ne* |
| **5. Įvairovė, nediskriminavimas ir sąžiningumas** *[Diversity, Non-discrimination, and Fairness]* | |
| 5.1. Ar DI sistema buvo patikrinta dėl šališkumo ir sąžiningumo, atsižvelgiant į įvairias grupes? *[Has the AI system been tested for bias and fairness across diverse groups?]* | * Taip* * Ne* |
| 5.2. Ar apibrėžti ir stebimi sąžiningumo rodikliai per visą DI gyvavimo ciklą? *[Are fairness metrics defined and monitored throughout the AI lifecycle?]* | * Taip* * Ne* |
| 5.3. Ar taikomi visuotinio dizaino principai, siekiant užtikrinti, kad sistema būtų prieinama visiems, įskaitant neįgaliuosius? (pvz., ar užtikrinta, kad DI technologijos būtų įtraukiamosios ir nesukurtų barjerų jokiai vartotojų grupei.) *[Are Universal Design principles applied to ensure accessibility for all users, including those with disabilities?]* | * Taip* * Ne* |
| **6. Socialinė ir aplinkos gerovė** *[Societal and Environmental Well-being]* | |
| 6.1. Ar DI sistema siekia sumažinti savo poveikį aplinkai (pvz., energijos suvartojimą)? *[Does the AI system aim to minimize its environmental impact (e.g., energy consumption)?]* | * Taip* * Ne* |
| 6.2. Ar vertinamas sistemos poveikis visuomenei, įskaitant demokratiją ir visuomenės gerovę? *[Are societal impacts, including effects on democracy and public welfare, assessed?]* | * Taip* * Ne* |
| 6.3. Ar sistema skatina naujų įgūdžių įgijimą ir prasmingą darbą, o ne darbuotojų kvalifikacijos mažėjimą? *[Does the system promote new skills and meaningful work opportunities, rather than de-skilling?]* | * Taip* * Ne* |
| **7. Atsakomybė** *[Accountability]* | |
| 7.1. Ar įdiegti mechanizmai, leidžiantys trečiosioms šalims atlikti sistemos auditą? [*Are mechanisms in place for third-party audits of the AI system?]* | * Taip* * Ne* |
| 7.2. Ar yra procesai, leidžiantys pranešti ir spręsti pažeidžiamumus ar rizikas? *[Is there a process for reporting and addressing vulnerabilities or risks?]* | * Taip* * Ne* |
| 7.3. Ar aiškiai apibrėžtos DI kūrėjų, naudotojų ir kitų suinteresuotų šalių atsakomybės? *[Are the roles and responsibilities of AI developers, users, and stakeholders clearly defined?]* | * Taip* * Ne* |
| 7.4. Ar numatytos priemonės, leidžiančios pašalinti neigiamą poveikį, atsiradusį dėl DI sistemos veiklos? *[Are measures in place to provide redress for any adverse effects caused by the AI system?]* | * Taip* * Ne* |
| **8. Dokumentacija ir įrodymai** *[Documentation and Evidence]* | |
| 8.1. Ar tinkamai registruojami visi vertinimai ir auditai? *[Is a clear record of all assessments and audits maintained?]* | * Taip* * Ne* |
| 8.2. Ar pridėti visi teisinio ir etinio atitikimo dokumentai? *[Are all regulatory and ethical compliance documents attached?]* | * Taip* * Ne* |

*Patvirtinu, kad Tyrimas bus atliekamas remiantis šioje paraiškoje pateikta informacija ir Tyrimo vykdymo metu nebus pažeidžiami galiojantys teisės aktai.*

*.......................................* *............................* *....................................*

*(pareigos)* *(parašas)* *(vardas ir pavardė*