Nuo efektyvumo iki pažeidžiamumo: kaip išvengti dirbtinio intelekto keliamų rizikų?

1. AI aktas.

Nors galutinis AI akto tekstas dar derinamas esminius bruožus žinome jau dabar:

- **Reguliavimo esmė** – AI sprendimų skirstymas pagal keliamas potencialias rizikas. Nuo to, į kurią „lentynėlę“ sprendimas patenka, priklauso kokios taisyklės jam bus taikomos. Jei rizika itin aukšta (pvz., socialinio reitingo sistemos), tokių sprendimų naudojimas iš viso nėra galimas, jei rizika didelė – taikomi aukšti skaidrumo bei atskaitomybės reikalavimai, jei rizika maža – taikomi skaidrumo reikalavimai bus šiek tiek švelnesni;
- Atskiras dėmesys skiriamas tokioms generatyvinėms AI sistemoms kaip *ChatGPT*;
- Sankcijos daugiau nei įspūdingos - nuo 35 mln. eurų arba 7 % pasaulinės apyvartos iki 7,5 mln. eurų arba 1,5 % apyvartos;
- Aktui įsigaliojus, turėsime 2 metus pasiruošti jo taikymui.

2. Kai AI naudoja Jūsų komanda.

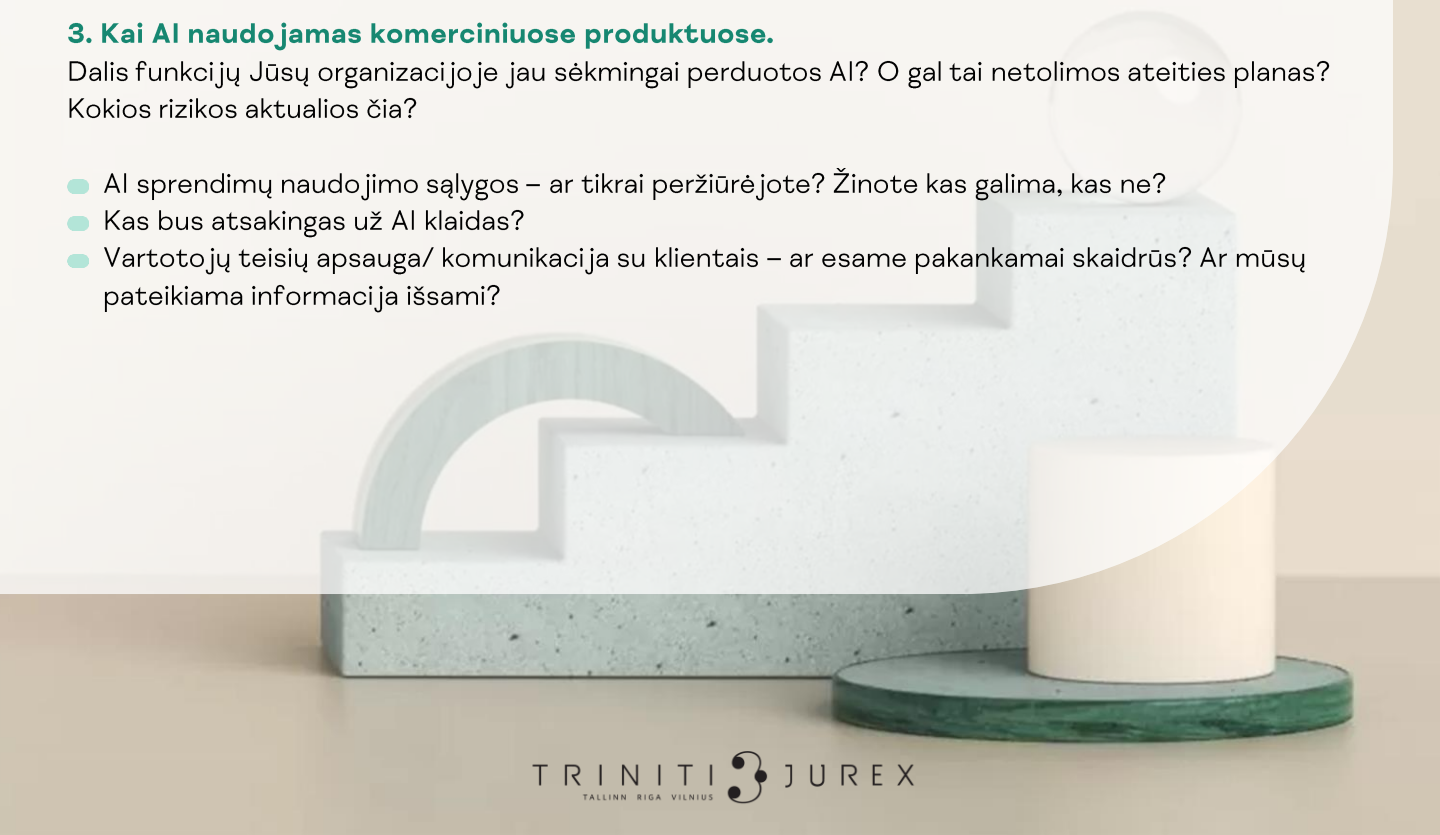
Labai tikėtina, kad Jūsų komanda AI sprendimus kasdienėms funkcijoms vykdyti jau kurį laiką naudoja ir ši tendencija tik stiprės. Su jūsų žinia ir pritarimu, ar ne. Kokias rizikas tai kelia?

- **Konfidencialumas** – ar esate tikri, kad Jūsų komanda su AI įrankiais nesidalina Jūsų ar Jūsų klientų konfidencialia informacija?
- **Asmens duomenų apsauga** – ar pasirūpinote, kad AI sprendimams skirtoms užduotims formuluoti būtų naudojami tik nuasmeninti tekstai/ medžiaga?
- **Klaidos ir šališki sprendimai** – ar tikrai su AI pagalba sugeneruotas turinys yra peržiūrimas žmogaus? Ar nėra pernelyg pasitikima tuo, ką mums atsako AI?
- **Intelektinė nuosavybė** – kam priklauso su AI pagalba sukurtas kūrinys? Ar mes neprarasime teisių į intelektinės nuosavybės objektus, jei juos kursime pasitelkdami AI?
- **Kibernetinis saugumas** – kaip išvengti pažeidžiamumo?

3. Kai AI naudojamas komerciniuose produktuose.

Dalis funkcijų Jūsų organizacijoje jau sėkmingai perduotos AI? O gal tai netolimos ateities planas? Kokios rizikos aktualios čia?

- AI sprendimų naudojimo sąlygos – ar tikrai peržiūrėjote? Žinote kas galima, kas ne?
- Kas bus atsakingas už AI klaidas?
- Vartotojų teisių apsauga/ komunikacija su klientais – ar esame pakankamai skaidrūs? Ar mūsų pateikiama informacija išsami?



Nuo efektyvumo iki pažeidžiamumo: kaip išvengti dirbtinio intelekto keliamų rizikų?

4. Kaip visas šias rizikas valdyti?

Svarbu iš anksto pasirengti bei turėti labai aiškias AI sprendimų naudojimo taisykles ir gaires, angl. AI policy.

AI policy paprastai aprašoma:

- Kokius AI sprendimus sutariame naudoti;
- Ką reiškia konfidenciali informacija? Kaip neprarasti jos konfidencialumo naudojantis AI;
- Ką reiškia asmens duomenys ir kaip veikia nuasmeninimas?
- Pareiga patikrinti rezultatą – galutinį sprendimą priima žmogus;
- Susitarimas dėl informavimo apie AI naudojimą.

Kas dar svarbu?

Be AI policy kiti svarbus dokumentai:

- **Privatumo politika svetainei**, ypatingai jei AI sprendimai naudojami klientų aptarnavime;
- **Vidinės asmens duomenų tvarkymo politikos** komandai turi būti aišku, kas yra asmens duomenys, kokių principų mes laikomės tvarkydami asmens duomenis ir ką reiškia, kad pasitelkiant AI sprendimus turi būti laikomasi privatumo ir kaip keliami duomenys turi būti pilnai nuasmeninti;
- **Konfidencialumo valdymo dokumentai,** tai gali būti konfidencialios informacijos sąrašas, susitarimai dėl jų apsaugos, valdymo tvarkos ir vidinės politikos;
- **Informacijos klasifikavimo ir saugos taisyklės,** komandai turi būti aišku, kokią informaciją yra saugu dalintis su AI įrankiais, o kur jau kelsime riziką konfidencialumo praradimui.

Būk žingsniu priekyje, palikdamas rizikas užnugaryje!

Jeigu jūsų įmonėje vis dar nėra numatytos AI sprendimų naudojimo taisyklės ir gairės, nors žinote, kad dalis darbuotojų naudojasi dirbtiniu intelektu, mes galime jums padėti.

Pranešimą pristatė
Aurelija Rutkauskaitė
aurelija.rutkauskaite@triniti jurex.lt