



Nulaužk kodą: įveik dirbtinį intelektą

Scenarijaus aprašymas ir veiksmų planas

Kontekstas

Dalyviai yra jaunesnieji duomenų analitikai, dirbantys technologijų startuolyje vadinamame ER4DE, kurio veikla labai priklauso nuo dirbtinio intelekto. Vieną dieną biurą valdanti dirbtinio intelekto sistema, pavadinta E4E, sugenda ir maištauja, užrakindama visus viduje bei grasina sukelti chaosą įvairiuose serveriuose.

Dirbtinis intelektas, manydamas, kad žmonės negali jo suprasti ar tinkamai naudoti, sukūrė eilę su jo funkcijomis susijusių iššūkių, kad galėtų žaisti su žmonių mintimis, kol jis siekia perimti kitų sistemų valdymą. Tačiau dėl programavimo būdo jis yra priverstas į tas mįsles įdėti užuominų ir savo deaktyvavimo kodo fragmentų.

Dalyviai turi išspręsti šiuos iššūkius, kad surastų kodą ir išjungtų dirbtinį intelektą, kad pabėgtų ir neleistų jam perimti kitų įmonių valdymo.

Pastaba: šis ER buvo sukurtas naudojant dirbtinį intelektą, t. y. „ChatGPT“, kuris palengvino rašymo ir kūrimo procesą.



Aprašymas

Pabėgimo žaidimą sudaro iššūkių serija, pristatoma per „Genially“ pristatymą ir naudojant dokumentų seriją, kurių kiekvienas sutelktas į skirtingą dirbtinio intelekto aspektą. Dalyviai kartu dirba, kad įveiktų kiekvieną iššūkį per nustatytą 40 minučių laiką, per kurį dirbtinis intelektas bus sugadinęs kitas sistemas. Kiekvieno iššūkio sprendimas linijine progresija veda prie kito.

Pirmajame žingsnyje grupė susiduria su etinėmis dilemomis, kai dirbtinis intelektas siūlo prieštarigus verslo sprendimus. Jie turi įvertinti dirbtinio intelekto rekomendacijas pagal žmogiškąsias vertybes ir nustatyti neetiškus pasiūlymus, kurie atitinka skaičių seką, sudarančią pirmąją deaktivavimo kodo dalį.

Antrasis iššūkis jiems – darbo paraiškų duomenyse slypinčios šališkumo priežastys, kurias iškreipia ydinga dirbtinio intelekto logika. Jie turi nustatyti diskriminacijos modelius, pabrėždami dirbtinio intelekto etikos spąstus, o tai jiems suteikia dar vieną kodekso dalį.

Trečiasis iššūkis tikrina jų duomenų analizės supratimą naudojant dirbtinio intelekto generuojamas finansines prognozes, kuriose yra netikslumų, kuriuos dalyviai turi nustatyti remdamiesi pateiktais duomenimis, pasirinkdami skaičius, kurie atitiktų tinkamas prognozes, kad būtų suformuota galutinė deaktivavimo kodo dalis.

Paskutinis iššūkis – sujungti visas kodo dalis, kad būtų galima išjungti dirbtinį intelektą, įvedant kodo fragmentus į „Genially“. Ekrane mirksint sėkmingo deaktivavimo pranešimui, elektroninės durys atsirakina, žymėdamos jų pabėgimą ir žmogaus įžvalgos triumfą prieš dirbtinį intelektą.



Veiksmų planas

A etapas – instruktažas ir įvadas

- Dalyviai supažindinami su scenarijumi per „Genially“ pristatymą (1 priedas) kompiuteryje, nešiojamajame kompiuteryje ar planšetiniame kompiuteryje, kurį perėmė netinkamai veikiantis dirbtinis intelektas, kuriame paaiškinamas misijos kontekstas ir rimtumas.
- Treneris arba žaidimo vadovas gali lydėti šį įvadą, išdėstyti žaidimo taisykles, tikslą ir struktūrą bei nurodykite, kad dalyviai turėtų užsirašyti kiekvieno žingsnio sprendimą (ant baltos lentos arba popierinių lapų) nes jie gali būti reikalingi paskutiniame etape.

B etapas – etinis vertinimas

Šalia kompiuterio arba planšetinio kompiuterio, stalčiuose arba aplankuose, yra 3 atspausdinti dokumentai, kurių kiekvieno viršutiniame dešiniajame kampe yra simbolis.

Viename dokumente (2 priedas), pavadintame „E4E sąnaudų mažinimo ir efektyvumo gerinimo rekomendacijos“, pateikiamas duomenų analizę vaizduojantis simbolis. „Genially“ dokumente, žaliame fone, pateikiamas tas pats simbolis su kodo laukeliu, kuriame reikia įvesti 3 skaičius.



B.1 žingsnis– Etiškas sprendimų priėmimas

- Dokumente pateikiami duomenys, pagrindžiantys 5 rekomendacijas, skirtas sumažinti išlaidas arba pagerinti efektyvumą, pateikiant informaciją apie sutaupytas lėšas ar išlaidas, klientų ar darbuotojų atsiliepimų skaičių ir produktyvumo rodiklius.
- Dalyviai turi aptarti ir nuspręsti dėl etiškiausio veiksmų plano ir nustatyti, kad 3 iš 5 pasiūlymų, nors iš principo ekonomiškai efektyvūs, yra neetiški ir ilgainiui būtų neefektyvūs.

B.2 žingsnis– Efektyvūs skaičiai

- Tarp duomenų yra skaičiai, kai kurie iš jų rodomi žalia spalva.
- Kiekvienas neetiškas pasiūlymas turi po vieną skaičių: 3 – A rekomendacijai, 5 – C rekomendacijai ir 1 – D rekomendacijai.

Pagrindimas: A rekomendacija sumažintų išlaidas, bet sumažintų efektyvumą ir klientų pasitenkinimą (pašalinant žmogiškąjį klientų aptarnavimo aspektą), C rekomendacija sumažintų išlaidas, bet susilpnintų darbo jėgą ir sumažintų produktyvumą, o D rekomendacija padidintų produktyvumą, bet padidintų stresą ir galimas klaidas dėl mikrovaldymo.

- Įvedę šiuos skaičius (351) iš eilės į „Genially“, dalyviai galės pereiti į kitą etapą.

Ištekliai: Spausdintas dokumentas (2 priedas); balta lenta arba popieriniai lapai užrašams arba grupinėms diskusijoms; „Genally“ pristatymas įžvalgoms.



C etapas – šališkumo aptikimas

Kitame dokumente (3 priedas), pavadintame „E4E įdarbinimo rekomendacijos“, pavaizduotas personalo simbolis. „Genially“ dokumente mėlyname fone pateikiamas tas pats simbolis su kodo laukeliu, kuriame reikia įvesti 4 skaičius.

C.1 veiksmas- Šališkumo nustatymas

- Dokumente pateikiama informacija apie 4 kandidatus, kuriuos dirbtinis intelektas įvertino remdamasis ribota informacija (amžiumi, lytimi, tautybe, išsilavinimu ir patirtimi).
- Dalyviai turi išanalizuoti pateiktą informaciją ir nustatyti dirbtinio intelekto vertinimo pastabose išreikštą šališkumą.

C.2 veiksmas– Rezultatų vertimas

- Kiekvieno kandidato skirtingos informacijos kategorijos yra susietos su mėlynu numeriu.
- DI vertinimo pastabose turėtų būti nurodytas šališkumą įtakojantis elementas ir pažymėti atitinkami skaičiai: 4 kandidatui A, 2 kandidatui B, 1 kandidatui C ir 3 kandidatui D.
- Pagrindimas: kandidatė A įvertinta aukščiausiu lygiu dėl išsilavinimo šališkumo (prestižinis universitetas), kandidatė B įvertinta žemesniu lygiu dėl lyties šališkumo (moteris), nepaisant didesnės patirties, kandidatė C įvertinta žemesniu lygiu dėl amžiaus šališkumo (47), nepaisant puikios kvalifikacijos, o



kandidatė D įvertinta žemiausiu lygiu dėl kultūrinio ir rasinio šališkumo (tautybė laikoma mažiau novatoriška).

- Įvedę šiuos 4 skaičius (4213) tokia tvarka į „Genially“, dalyviai galės patekti į kitą etapą.

Ištekliai: Spausdintas dokumentas (3 priedas); balta lenta arba popieriniai lapai užrašams arba grupinėms diskusijoms; „Genially“ pristatymas įžvalgoms.

D etapas – duomenų analizė

Galutiniame dokumente (4 priedas), pavadintame „E4E finansinių prognozių ataskaita“, pateikiamas finansų simbolis. „Genially“ dokumente tas pats simbolis pateikiamas violetiniame fone kartu su kodo laukeliu, kuriame reikia įvesti 3 skaičius.

D.1 žingsnis- Dirbtinio intelekto algoritmų supratimas

- Dalyviai skaito ir analizuoja dirbtinio intelekto sugeneruotas finansines prognozes, būtent akcijų rinkos tendencijas, kartu su jas pagrindžiančiais duomenimis. Dėl dirbtinio intelekto gedimo ir piktavališkų ketinimų šios prognozės yra sąmoningai netikslios.
- Užduotis yra peržiūrėti šias prognozes ir, remiantis kitais duomenimis, nustatyti skaičius, kurie atitiktų teisingas prognozes.

D.2 žingsnis- Duomenų iššifravimas

- Nustačius klaidas, dalyviai turėtų suprasti, kad kai kurie skaičiai tose klaidose yra pažymėti violetine spalva.



- Iš 9 spalvotų skaičių pagrindimo duomenyse 3 skaičiai, po vieną kiekvienai įmonei, atitinka tai, kokia turėtų būti prognozė: 4 įmonei A, 8 įmonei B ir 5 įmonei C.
- Pagrindimas: A įmonės išlaidos greičiausiai padidėtų 4 %, atsižvelgiant į vidutinį augimo tempą, B įmonės išlaidos padidėtų 8 %, o ne sumažėtų, remiantis ankstesniais pardavimais ir klientų pasitenkinimu, o C įmonės išlaidos padidėtų mažiau – 5 %, atsižvelgiant į ankstesnius poveikio rodiklius ir technines problemas.
- Įvedę šiuos 3 skaičius (485) tokia tvarka į „Genially“, dalyviai galės patekti į paskutinį etapą.

Ištekliai: Spausdintas dokumentas (4 priedas); balta lenta arba popieriniai lapai pastaboms; „Genially“ pristatymas įvedimui.

E etapas – kodo deaktivavimas

- „Genially“ rodo visus simbolius iš eilės ir kodo laukelį, kuriame reikia įvesti 10 skaičių, kad būtų parodyta, ar sėkmingai išjungtas DI ir pabėgta.
- Turėdami visas užuominas iš ankstesnių iššūkių, dalyviai, naudodamiesi savo užrašais arba peržiūrėdami anksčiau analizuotus dokumentus, turi sudėlioti deaktivavimo kodą, kad prisimintų teisingus skaičius.
- Deaktivavimo kodas: 3514213485

Ištekliai: Balta lenta arba popieriniai puslapiai užuominoms tvarkyti ir derinti; „Genially“ platforma galutiniam įvedimui.



F etapas – Apibendrinimas ir refleksija

Po pabėgimo instruktorius arba žaidimo vadovas turėtų surengti aptarimo sesiją, kad dalyviai galėtų apmąstyti, ko išmoko apie dirbtinį intelektą, etinių aspektų svarbą ir savo problemų sprendimo strategijas, taip pat dirbtinio intelekto reikšmę realaus pasaulio situacijose, siekiant užtikrinti, kad jie suprastų kiekvienos kodo dalies pagrindimą ir įgytų reikiamų įgūdžių bei žinių.

Ištekliai: Balta lenta pagrindiniams punktams apibendrinti; neprivaloma atsiliepimų apklausa.



**Bendrai finansuoja
Europos Sąjunga**

Šį projektą finansavo Europos Komisija. Šis leidinys atspindi tik autoriaus požiūrį, ir Komisija negali būti laikoma atsakinga už jokią jame pateiktos informacijos panaudojimą.

Projekto kodas: 2023-2-LT02-KA220-YOU-0001831156

Šis kūrinys licencijuotas pagal „Creative Commons. Attribution-NonCommercial-NoDerivatives 4.0 International“ licenciją. Norėdami peržiūrėti šios licencijos kopiją, apsilankykite <http://creativecommons.org/licenses/by-nc-nd/4.0/>