



Kaip etiškai ir adekvačiai panaudoti dirbtinį intelektą



Apibrėžimas

Dirbtinis intelektas (DI arba AI) yra kompiuterinių sistemų, galinčių atlikti užduotis, kurioms paprastai reikia žmogaus intelekto, kūrimas. Šios užduotys apima samprotavimus, modelių atpažinimą, mokymąsi iš duomenų ir sprendimų priėmimą. Etinis AI naudojimas yra skirtas dirbtinio intelekto technologijų kūrimui, kurios yra ne tik veiksmingos, bet ir atitinka moralines bei visuomenės vertybes. Tai apima etikos klausimus, tokius kaip sąžiningumas, skaidrumas, atskaitomybė ir privatumas.

Etiškas AI užtikrina, kad AI sistemos veiktų be šališkumo, gerbtų vartotojų privatumą ir veiktų skaidriai, kad suinteresuotosios šalys suprastų AI sprendimus. Be to, norint tinkamai naudoti dirbtinį intelektą, būtinas efektyvumas, etikos standartų laikymasis ir nuosekli stebėseną, siekiant sumažinti riziką, pvz., nenumatytas pasekmes ar netinkamą technologijų naudojimą.

Integravus šiuos principus, dirbtinis intelektas gali būti atsakingai panaudotas siekiant padidinti produktyvumą ir pagerinti gyvenimą, kartu išlaikant pasitikėjimą ir saugą nuo AI priklausančioje veikloje. Tam tikros šio darbalapio temos sutampa su tomis, kurios nagrinėjamos skyriuje „Skaitmeninio verslumo etikos problemos“. Norėdami atlikti išsamesnę šių aspektų analizę, skaitytojai raginami remtis tuo dokumentu.



Kodėl tai svarbu?

Etinės praktikos integravimas į dirbtinį intelektą (AI) yra labai svarbus siekiant užtikrinti, kad technologijų pažanga teigiamai prisidėtų prie visuomenės ir gerbtų asmens teises.

- **Skatina pasitikėjimą:** etiškas dirbtinis intelektas stiprina vartotojų pasitikėjimą, o tai būtina norint įdiegti ir veiksmingai naudoti technologijas švietime ir ne tik.
- **Skatina sąžiningumą:** sumažindamas šališkumą, etiškas AI užtikrina, kad visi asmenys turėtų lygias galimybes gauti naudos iš technologijų pažangos.
- **Saugo privatumą:** etiška AI praktika saugo asmens duomenis, išsaugant vartotojų informacijos privatumą ir saugumą.





- **Padidina patikimumą:** AI sistemos, sukurtos laikantis etikos gairių, sumažina klaidas ir neatitikimus, padidina jų patikimumą ir efektyvumą.
- **Palaiko įtraukimą:** etiškai sukurtas AI atsižvelgia į įvairius poreikius ir pagrindus, todėl skaitmeninės priemonės tampa prieinamos ir naudingos visiems.



Patarimai ir gudrybės

- ➔ **Supraskite ir pašalinkite šališkumą:** nustatykite ir sumažinkite duomenų ir algoritmų paklaidas, kad užtikrintumėte teisingus AI rezultatus, naudodami įvairius mokymo duomenis, kurie tiksliai atspindi skirtingus demografinius rodiklius.
- ➔ **Pirmenybė teikiama skaidrumui ir atskaitomybei:** tvarkykite aiškią AI procesų ir sprendimų dokumentaciją ir užtikrinkite atvirą bendravimą su suinteresuotosiomis šalimis apie tai, kaip veikia AI sistemos.
- ➔ **Įdiekite patikimą privatumo apsaugą:** naudokite tvirtus šifravimo ir saugius duomenų saugojimo sprendimus, kad apsaugotumėte naudotojų duomenis, integruodami privatumo aspektus AI programų kūrimo etape.
- ➔ **Užtikrinti žmogaus priežiūrą:** išlaikyti žmonių dalyvavimą svarbiuose sprendimų priėmimo procesuose, leidžiant prižiūrėti ir prireikus koreguoti AI sprendimus.
- ➔ **Skatinkite nuolatinį tobulėjimą:** skatinkite nuolatinį grįžtamąjį ryšį apie AI etinį našumą ir pritaikykite gaires bei praktiką, pagrįstą naujais iššūkiais ir technologine pažanga.
- ➔ **Įtraukite skirtingus požiūrius:** įsitikinkite, kad jūsų AI komanda yra įvairi. Žmonės iš skirtingų sluoksnių gali padėti pastebėti AI sistemų problemas ir šališkumą, kurie iš pradžių gali būti neaiškūs.
- ➔ **Įtraukite visus:** naudokite komandinį metodą, kai įtraukite ne tik technologijų žmones, bet ir etikos ekspertus bei grupių, kurias gali paveikti AI, atstovus. Įsitikinkite, kad kuriant ir naudojant AI atsižvelgiama į daugybę skirtingų požiūrių.



Naudingi ištekliai ir įrankiai

- ➔ **Supraskite ir pašalinkite šališkumą:**
 - **[AI Fairness 360 \(AIF360\)](#):** IBM įrankių rinkinys, skirtas AI modelių šališkumui tirti ir sumažinti.





- [„Google“ įrankis „What-If“](#): Siūloma lengvai naudojama sąsaja, skirta mašininio mokymosi modeliams tyrinėti, siekiant sąžiningumo ir kitų funkcijų, nereikia koduoti.



Pirmenybė teikiama skaidrumui ir atskaitomybei:

- [IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems \(pasaulinė autonominių ir intelektualių sistemų etikos iniciatyva\)](#): Pateikiami etikos standartai ir geriausios praktikos pavyzdžiai kuriant dirbtinio intelekto sistemas.
- [Partnership on AI](#): Suformuluoja geriausią AI technologijų praktiką ir gerina visuomenės supratimą



Įdiekite patikimą privatumo apsaugą:

- [edX Ethics in AI and Data Science \(AI etika ir duomenų mokslas sistemoje\)](#): Apima etinius AI klausimus ir pateikiamos privatumo ir saugumo problemų valdymo sistemos.



Skatinkite nuolatinį tobulėjimą:

- [AI For Everyone \(AI visiems\) „Coursera“](#): Šiame kurse, kurį vedė Andrew Ng, aptariami AI socialiniai padariniai ir skatinama laikytis etinės praktikos.



Knygos gilesnėms įžvalgoms:

- [„Weapons of Math Destruction“](#) (Matematikos naikavimo ginklai) Cathy O'Neil – Šioje knygoje nagrinėjama tamsioji didelių duomenų pusė ir šališki algoritmai, galintys sustiprinti nelygybę
- [„Ethics of Artificial Intelligence“](#) (Dirbtinio intelekto etika), redagavo S. Matthew Liao – žymiausių filosofų esė rinkinys, siūlantis filosofinį dirbtinio intelekto etinių pasekmių traktavimą.