# Tekoäly- ja koneoppimismallien anonymiteettilomake

## 1 Mallin kuvaus

## Onko mallin opettamisessa käytetty ohjattua vai ohjaamatonta oppimista?

## ohjaamaton

## ohjattu

## muu, mikä? Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## Onko malli generatiivinen, luokittelu vai regressio malli?

## generatiivinen malli

## luokittelumalli

## regressiomalli

## muu, mikä? Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## Minkä tyyppinen malli on kyseessä (Esim. Onko kyseessä neuroverkko? Minkä tyyppinen neuroverkko on kyseessä ja kuinka monta piilokerrosta se sisältää? Jos kyseessä ei ole neuroverkko, miten luokittelisit mallin arkkitehtuurin?)

## Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## Onko käytetty esikoulutettua pohjamallia? Mikä pohjamalli on kyseessä? Kuinka suuri osa mallin pohjamallin kerroksista on vapautettu koulutusvaiheessa? Kuinka kuvailisit pohjamallin ja kehitetyn mallin eroja?

## Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## Anna lyhyt tiivistelmä mallin toimintaperiaatteesta.

## Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## Ilmoita mallin parametrien määrä.

## Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## Miten mallin yleistämiskyky ja ylisovittamisen (overfitting) minimointi on huomioitu mallin kehityksessä (esim. datan augmentointi, regularisointimenetelmät)?

## Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## Missä muodossa malli on tallennettu, ja aiotaan viedä ulos? Mitä tiedostoja malliin liittyen viedään ulos?

## Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## 2 Yksityisyyttä turvaavat teknologiat

Onko mallin kehityksessä käytetty differentiaalista yksityisyyttä tai muita yksityisyyden suojaa parantavia tekniikoita? Kuvaile, kuinka nämä menetelmät ja/tai mallin anonymiteetti muuten huomioitiin kehityksen aikana?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Mitä parametreja yksityisyyttä turvaavien teknologioiden osalta on käytetty? Ilmoita myös hyperparametrien arvot.

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## 3 Mallin käyttötarkoitus

Kenen käyttöön malli on tarkoitettu?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Missä mallia säilytetään, ja kenellä on oikeudet käsitellä mallia?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Kenellä on oikeudet nähdä tai käsitellä mallin tuloksia?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Mitä tietoja mallin arkkitehtuurista tai toimintaperiaatteesta on tarkoitus julkaista?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Anna arvio mallin uudelleenkäytön riskistä haitallisiin tarkoituksiin, erityisesti opetusdatassa esiintyvien yksilöitä koskevien tietojen paljastumiseen.

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

## 4 Opetusdatan kuvaus

Millä aineistolla malli on opetettu? Onko opetusdataa käsitelty ennen käyttöä mallin opetuksessa? Kyllä jos, kerro tarkemmin, kuinka opetusmateriaali on käsitelty.

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Onko opetukseen käytetty yksilötasoista aineistoa?

kyllä

ei

Sisältääkö opetukseen käytetty aineisto suoria tunnisteita?

kyllä

ei

Kuinka suurella määrällä dataa mallia on koulutettu? Ilmoita opetusdatan instanssien määrä sekä siinä esiintyvien yksilöiden määrä.

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Missä muodossa data on syötetty mallille?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Jos kyseessä on luokittelumalli, kuinka moneen luokkaan opetusdata on luokiteltu?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Ovatko luokat määrällisesti tasapainossa? Jos luokat eivät ole määrällisesti tasapainossa, kuvaa luokitellun epätasapainoisuutta.

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Miten mallin suoriutumista on arvioitu?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Jos mallin suoriutumista on arvioitu erillisellä testausdatalla (edellyttäen, että kyseistä dataa ei ole käytetty mallin opettamisessa), kuinka paljon testidataa on käytetty?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Millä mittareilla mallin suoriutumista ja yleistämiskykyä on arvioitu? Ja mitä tuloksia mittarit tuottivat?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

|  |  |
| --- | --- |
| *Arviointimittari (esim. tarkkuus)* | *Tulos* |
| Kirjoita tekstiä napsauttamalla. |  |
| Kirjoita tekstiä napsauttamalla. |  |
| Kirjoita tekstiä napsauttamalla. |  |
| Kirjoita tekstiä napsauttamalla. |  |
| Kirjoita tekstiä napsauttamalla. |  |

## 6 Mallin anonymiteetin arviointi ja tarkastus

Onko mallin resistenssiä testattu opetusdataan sisältyvien tietojen paljastumiseen tähtääviä hyökkäyksiä vastaan (mm. membership inference attacks, reconstruction attacks)? Jos kyllä, kuvaile prosessia ja sen lopputuloksia.

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Millä perustein voidaan olettaa, että malli on anonyymi?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.

Miten mallin anonymiteetti voidaan tarkistaa?

Kirjoita tekstiä napsauttamalla tai napauttamalla tätä.