

Algoritmien perusteet ja tekoälyn etiikka: teoriasta käytäntöön -kurssi.

Työkalupakki synkronisia istuntoja
varten

Projektipohjaiset harjoitukset

Projektipohjaiset harjoitukset



Projektipohjaiset harjoitukset

	Kuvaus
Tehtävän kuvaus	Arvioikaa 4-5 hengen ryhmissä yhtä kuudesta tekoälyratkaisusta viiden kurssikokonaisuuden aikana. Tähän kattavaan tehtävään kuuluu tekoälyratkaisun yksityiskohtainen analyysi useista eri näkökulmista kurssin rakenteen ja kussakin kurssikokonaisuudessa käsiteltyjen aiheiden mukaisesti.
Kuvaus tehtävän suorittamisesta	Ohjeet • Ryhmänne käsittelee yhtä tekoälyratkaisua sivulla 4 olevasta luettelosta. • Analysoi valitsemasi ratkaisu sivulla 6 esitettyjen viiden näkökulman perusteella. • Dokumentoi havaintosi jäsennellyssä raportissa, jossa käsittelet yksityiskohtaisesti kutakin arviointikriteeriä. Kukin kohta tulee esittää max. 2 diaa • Esitlee analyysisi luokalle viimeisellä synkronisella istunnolla ja korosta keskeiset havainnot ja suositukset ratkaisun parantamiseksi. • Esityksenne ryhmälle tulisi kestää 5-10 minuuttia.
Arvioitu aika tehtävän suorittamiseen	Noin 1-2 tuntia / CU (yhteensä 5-10 tuntia).
Lähde-ehdotuksia tehtävän tekemistä varten	Interaktiiviset vihkot ja muu kurssimateriaali + muut lähteet.
Yksityiskohtainen kuvaus tehtävän suorittamisesta	Katso seuraavat sivut.
Tiedot tehtävän toimittamisen määräajasta	Esitetään viimeisessä synkronisessa istunnossa.
Yhteystiedot tai miten epäselvyyksiä voi selventää	Opettajan on annettava yhteystieto (se voi olla sähköpostiosoite, puhelinnumero...).

Projektipohjaiset harjoitukset

Aikataulu ja ohjeet

HANKKEEN AIKATAULU

Projektin esittely ja ryhmien muodostaminen.

- Aloita projektin työstäminen mieluiten 1. synkronisen istunnon jälkeen, sovi tiimin kanssa tapaamisista projektin työstämiseksi.
- Vaihtoehtoisesti voit aloittaa projektin 2. synkronisen istunnon jälkeen, joka kattaa opintokokonaisuudet 1 ja 2.

Edistymisen tarkastelu

- 3. ja 4. synkronisella istunnolla on mahdollisuus tarkastella edistymistä ja keskustella mahdollisista haasteista opettajan kanssa.

Lopullinen esitys

- Esittäkää 5. synkronisella istunnolla projektiharjoitus ryhmänä.
 - Esitys 5-10 minuuttia.
 - Voit valita PowerPoint- tai Canva-asettelun, muista tehdä houkutteleva esitys.

HUOMAUTUS | Jos opiskelija ei voi osallistua ryhmätyöskentelyyn (tai ei osallistu aktiivisesti projektiin), projektitehtävä on toimitettava opettajalle yksitellen (10 diaa).

Muita ohjeita

- Aiheenne esitellään seuraavassa diassa. Pääsette tiiminä päättämään palvelunne nimen, tarkoituksen ja tarkemman sisällön. Hakemuksen tulee kuitenkin olla seuraavalla sivulla annetussa kentässä.
- Jatkakaa projektin työstämistä projektiryhmän kanssa sovittuun tahtiin, sisältö liittyy kuitenkin aina kurssikokonaisuuksiin, joten työskentely sen jälkeen, kun ryhmän jäsenet ovat saaneet valmiiksi aihealuekohtaiset verkko-opinnot, on suositeltavaa.
- Voit itse päättää, miten rakennat esityksesi; kysymykset ovat siis ohjeena, mutta sinun tulisi luoda sisältö, joka on yleisöä kiinnostava.

Projektipohjaiset harjoitukset

AI-ratkaisut jaettu ryhmäkohtaisesti

Tekoälyratkaisu, jonka parissa tiimisi tulisi työskennellä:

Tekoälykäyttöinen terveydenhuollon diagnoosityökalu

- Työkalu, joka avustaa lääkäreitä antamalla alustavia diagnooseja potilaan oireiden ja sairaushistorian perusteella.
 - XX ryhmät, jotka työskentelevät aiheen parissa (lisää tähän oppilaiden nimet, jotka opettaja on jakanut 4-5 hengen ryhmiin).

Suoratoistopalvelujen suosittelujärjestelmä

- Sovellus, joka ehdottaa käyttäjille elokuvia ja sarjoja heidän katseluhistoriansa ja mieltymystensä perusteella.
 - XX ryhmää työskentelee aiheen parissa (lisää tähän oppilaiden nimet, opettajan mukaan jaettuna 4-5 hengen ryhmiin).

Älykkään kodin automaatiojärjestelmä

- Ratkaisu, joka automatisoi kodin laitteita, kuten valoja, termostaatteja ja turvajärjestelmiä, käyttäjän rutiinien ja mieltymysten perusteella.
 - XX ryhmät työskentelevät aiheen parissa (lisää tähän oppilaiden nimet, opettajan mukaan jaettuna 4-5 hengen ryhmiin).

Tekoälyohjattu finanssisijoitusneuvoja

- Neuvonantaja, joka antaa sijoitussuosituksia markkinatietojen ja käyttäjän taloudellisten tavoitteiden perusteella.
 - XX ryhmää työskentelee aiheen parissa (lisää tähän oppilaiden nimet, jaetaan opettajan mukaan 4-5 hengen ryhmiin).

Verkkokaupan työkalusovellus

- Sovellus, joka suosittelee käyttäjille tuotteita heidän selaus- ja ostohistoriansa perusteella.
 - XX ryhmää työskentelee aiheen parissa (lisää tähän oppilaiden nimet, jaetaan opettajan toimesta 4-5 hengen ryhmiin).

Henkilökohtainen verkko-oppimisalusta

- Alusta, joka räätälöi oppimispolkuja ja sisältöä oppilaille heidän oppimistyyliensä, edistymisensä ja mieltymystensä perusteella.
 - XX ryhmät työskentelevät aiheen parissa (lisää tähän oppilaiden nimet, jaettu opettajan mukaan 4-5 hengen ryhmiin).

Hankekohtaiset harjoitukset

CU 1:een liittyvä sisältö

Aloita määrittelemällä seuraavat asiat (tämä tehtävä ei saisi viedä yli kahta esitysdiaa).

Ongelman kuvaus: Minkä ongelman sovelluksesi ratkaisee?

Tavoite: Mikä on tekoälysovelluksesi tavoite? Mikä on sovelluksen nimi? Mitkä ovat ratkaisusi menestyskriteerit?

Luo sidosryhmäkartta, jossa on määriteltävä ainakin käytön sidosryhmät, ulkoiset sidosryhmät ja kehittämisen sidosryhmät (käytä CU1-materiaalia ohjeena).

Pohdi, miten hyvin seuraavat eettiset periaatteet täyttyvät sovelluksessasi; pääpainon tulisi olla yhteiskunnallisissa ja ympäristönäkökohdissa:

Hyödyllisyys

Varmista, että tekoälyratkaisu edistää yksilön hyvinvointia ja yhteiskunnallista hyvää. Keskeiset huomioon otettavat alueet.

- Perusoikeuksien suojele: turvataan yksilön oikeudet ja vahvistetaan eettistä integriteettiä.
- Ympäristön kestävyys: minimoi ympäristövaikutukset kestävien käytäntöjen avulla.
- Oikeutus: järjestelmän tarkoitus on määriteltävä selkeästi ja yhdistettävä konkreettisiin yhteiskunnallisiin hyötyihin.

Epäoikeudenmukaisuus (non-maleficence).

- Varmistetaan, että ongelman tunnistaminen ja ehdotettu ratkaisu eivät vahingoita yksilöitä tai yhteisöjä.
- Harkitse ja seuraa huolellisesti järjestelmän vaikutuksia fyysiseen ja henkiseen hyvinvointiin.
- Arvioidaan mahdolliset turvallisuusriskit ja strategiat niiden lieventämiseksi.

Oikeudenmukaisuus.

Arvioidaan järjestelmän vaikutus instituutioihin, demokratiaan ja yhteiskuntaan.

Hankekohtaiset harjoitukset

CU 1 (Ohjausdia)

Tämän taulukon avulla voit tarkastella yksityiskohtaisemmin kunkin eettisen periaatteen mukaisia mahdollisia näkökohtia, kuten Morley et. al. (2020) ovat yhdistäneet.

Table 5 showing the connection between high-level ethical principles and tangible system requirements as adapted from the methodology outlined in Chapter II of the European Commission’s “Ethics Guidelines for Trustworthy AI”

Principle	Beneficence	Non-Maleficence	Autonomy	Justice	Explicability
Requirements	Stakeholder participation: to develop systems that are trustworthy and support human flourishing, those who will be affected by the system should be consulted Protection of fundamental rights Sustainable and environmentally friendly AI: the system’s supply chain should be assessed for resource usage and energy consumption Justification: the purpose for building the system must be clear and linked to a clear benefit—system’s should not be built for the sake of it	Resilience to attack and security: AI systems should be protected against vulnerabilities that can allow them to be exploited by adversaries. Fallback plan and general safety: AI systems should have safeguards that enable a fallback plan in case of problems. Accuracy: for example, documentation that demonstrates evaluation of whether the system is properly classifying results. Privacy and Data Protection: AI systems should guarantee privacy and data protection throughout a system’s entire lifecycle. Reliability and Reproducibility: does the system work the same way in a variety of different scenarios. Quality and integrity of the data: when data is gathered it may contain socially constructed biases, inaccuracies, errors and mistakes—this needs to be addressed Social Impact: the effects of system’s on people’s physical and mental wellbeing should be carefully considered and monitored	Human agency: users should be able to make informed autonomous decisions regarding AI systems Human oversight: may be achieved through governance mechanisms such as human-on-the-loop, human-in-the-loop, human-in-command	Avoidance of unfair bias Accessibility and universal design Society and democracy: the impact of the system on institutions, democracy and society at large should be considered Auditability: the enablement of the assessment of algorithms, data and design processes. Minimisation and reporting of negative impacts: measures should be taken to identify, assess, document, minimise and respond to potential negative impacts of AI systems Trade-offs: when trade-offs between requirements are necessary, a process should be put in place to explicitly acknowledge the trade-off, and evaluate it transparently Redress: mechanism should be in place to respond when things go wrong	Traceability: the data sets and the processes that yield the AI system’s decision should be documented Explainability: the ability to explain both the technical processes of an AI system and the related human decisions Interpretability

LÄHTEET | Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). Mistä miten: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices. Science and Engineering Ethics, 26(4), 2141-2168. <https://doi.org/10.1007/s11948-019-00165-5>

Projektipohjaiset harjoitukset

CU:n aiheeseen liittyvä sisältö

CU 2

Yksityisyyden suojaan liittyvät huolenaiheet

- Tunnista ratkaisuun liittyvät yksityisyyden suojaan liittyvät ongelmat.
- Mieti, miten tietoja kerätään, tallennetaan ja käytetään.
- Ehdota toimenpiteitä käyttäjien yksityisyyden suojaamiseksi ja tietoturvan varmistamiseksi.

CU 3

Käytetty algoritmi

- Kuvaile ratkaisun käyttövoimana oleva algoritmi.
- Selitä, miten algoritmi on rakennettu ja koulutettu kerättyjen tietojen perusteella (cu2).
- Keskustele kaikista teknisistä yksityiskohdista, jotka ovat tärkeitä ratkaisun ymmärtämisen kannalta.

CU 4

Mahdolliset vääristymät

- Tunnista järjestelmässä mahdollisesti esiintyvät vääristymät.
- Keskustele siitä, mikä voi aiheuttaa näitä vääristymiä (esim. koulutusdata, algoritmin suunnittelu).
- Ehdota keinoja näiden vääristymien estämiseksi tai lieventämiseksi.

CU 5

- TU delft on luonut tarkistuslistan eettistä tekoälyn kehittämistä varten. Tutustu työkaluun tämän [linkin](#) kautta ja hyödynnä sivuilla 12-13 olevaa AI APP -projektin tarkistuslistaa ja täytä sivulla 13 oleva tarkistuslista. Esitle tarkistuslistasi projektiesityksesi lopussa.
- Viimeisessä diassa tulisi esittää, mitä te tiiminä opitte ja mitä uusia oivalluksia saitte tämän esityksen tekemisen myötä.

KIITOS

Hankkeen numero: 2022-1-ES01-KA220-HED-000085257.



Universitat
de les Illes Balears



**Euroopan unionin
osarahoittama**

Euroopan komission tuki tämän julkaisun tuottamiselle ei tarkoita sitä, että sen sisältö vastaa ainoastaan sen kirjoittajien näkemyksiä, eikä komissio ole vastuussa sen sisältämien tietojen mahdollisesta käytöstä.

