



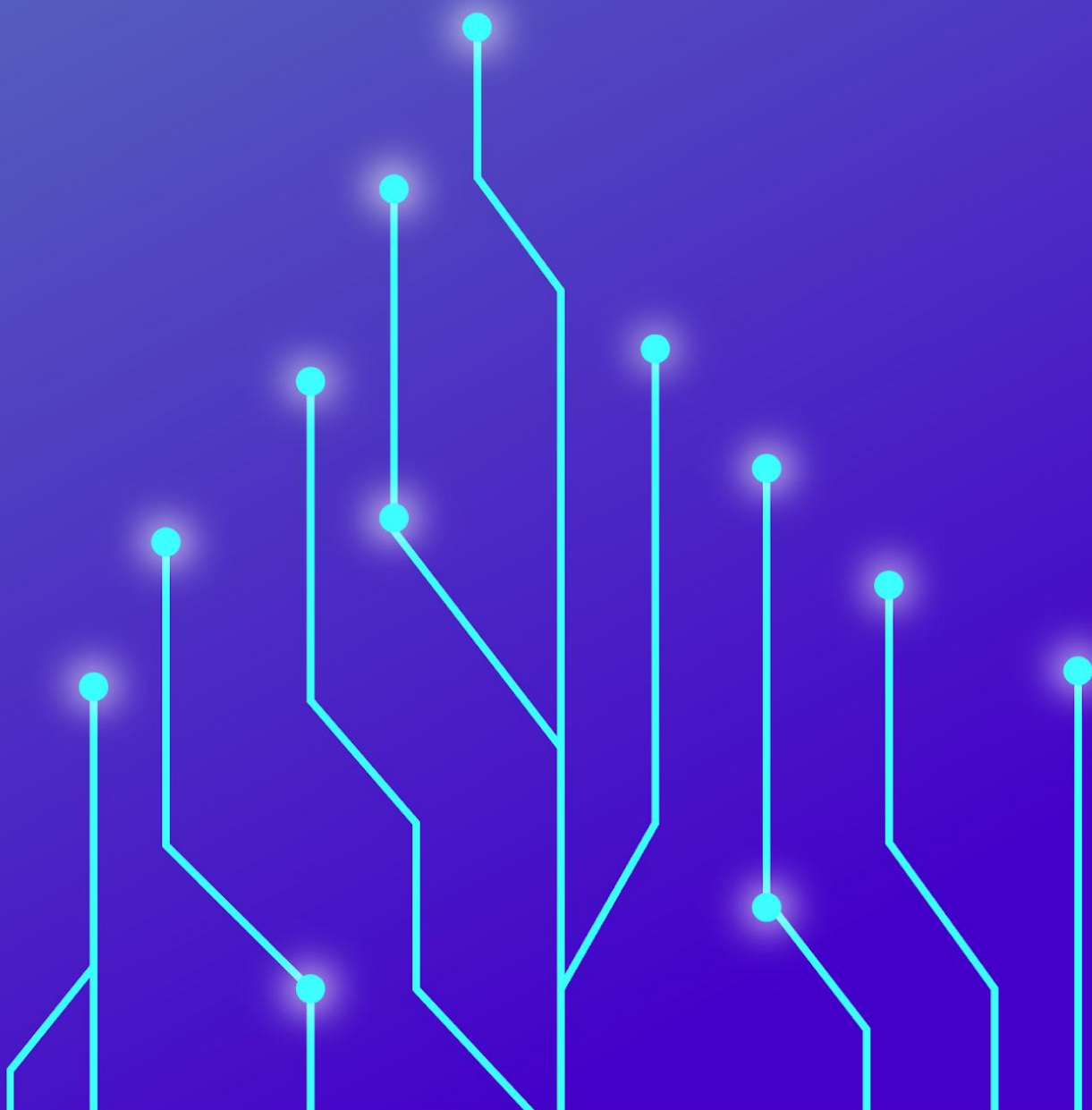
**Algoritmin perusteet ja tekoälyn
etiikka -kurssi: Teoriasta käytäntöön**

Käsikirja

Sisällys

CU1 AI ethics - A practical approach	2
CU2 AI privacy and convenience	33
CU3 Algorithms and their limitations	57
CU4 Data fairness and bias	97
CU5 Case studies and projects	113

Osaamiskokonaisuus 1 | Tekoälyn etiikka



Sisällys

1. Introduction	4
2. Defining ethical principles, frameworks, and guidelines	5
3. Common effort to shape the ethical landscape	6
4. Ethical principles	7
5. Ethical Frameworks, guidelines, and toolkits	11
6. From principles to practise; Trustworthy AI framework by HLEG	13
7. From principles to law: EU AI ACT	14
8. AI Development Lifecycle	16
9. Stakeholder engagement in AI Development lifecycle	17
10. AI Development Lifecycle: Problem definition and business understanding	21
11. AI Development Lifecycle: Design stage	22
12. AI Development Lifecycle: Data collection, understanding and preparation	23
13. AI Development Lifecycle; Model development and training	25
14. AI Development Lifecycle: Model evaluation and testing	26
15. AI Development Lifecycle: Deployment	27
16. AI Development Lifecycle: Operation and monitoring	28
17. Conclusion	29
18. References	30

1. Johdanto

Vinkkejä ja suosituksia opettajille

Tämä opas on tarkoitettu tukemaan sinua kurssin materiaalin läpikäymisessä. Tukimateriaalina toimivat PowerPoint-esityksissä on samanlaista sisältöä kuin verkko-oppimisympäristö, mutta yksityiskohtaisemmassa muodossa. PowerPointit ovat laajoja, eikä niitä ole tarkoitus käydä kokonaisuudessaan läpi luentojen aikana. Sen sijaan voit session alussa kysyä opiskelijoilta, mitkä aiheet olivat heille vaikeimpia ymmärtää, ja käyttää ainoastaan näihin aiheisiin liittyviä dioja keskustelun tukena.

Tätä voit kysyä esimerkiksi Mentimeterin tai Kahootin kaltaisilla työkaluilla, lisäämällä kyselyyn kurssiyksikön keskeiset sisällöt, jotka löytyvät tämän johdanto-osuuden lopusta luettelona. Lisäksi kurssiyksikköön liittyvä case-harjoitus on tarkoitus käsitellä myös vuorovaikutteisessa sessiossa.

Kaikki muut oppaassa esitetyt ehdotukset ovat ainoastaan ohjeellisia – voit käyttää opasta parhaaksi katsomallasi tavalla.

Mitä tapahtuisi, jos tekoäly saisi vallan päättää, myönnetäänkö sinulle laina? Tai mitä seurauksia olisi sillä, jos tekoäly ratkaisee, kuka tuomitaan vankilaan? Mitä tapahtuisi, jos autonominen auto joutuisi päättämään, pelastaako jalankulkija vai autossa olevat matkustajat? Entä jos tekoälypohjainen rekrytointijärjestelmä valitsisi lentokapteeniksi joko mies- tai naispuolisen hakijan?

Tekoälyn yhä laajempi integrointi yhteiskunnan eri osa-alueille herättää merkittäviä eettisiä kysymyksiä. Päätökset, jotka liittyvät lainoihin, rikosoikeudellisiin tuomioihin, autonomisiin ajoneuvoihin ja rekrytointikäytäntöihin, asettavat keskiöön eettiset periaatteet, kuten oikeudenmukaisuuden, läpinäkyvyyden ja vastuullisuuden. Tekoälyteknologiat voivat vahvistaa ennakkoluuloja, syrjintää ja eriarvoisuutta, ellei niitä kehitetä ja säännellä huolellisesti. On ensiarvoisen tärkeää varmistaa, että tekoälyjärjestelmät asettavat eettiset näkökohdat etusijalle, jotta voidaan edistää oikeudenmukaisuutta, tasa-arvoa ja luottamusta niiden sovelluksiin eri aloilla. (UNESCO, 2023c)

Tässä osiossa syvennytään tekoälyn kehittämiseen liittyviin eettisiin periaatteisiin, viitekehyksiin ja ohjeisiin. Tekoälyn kehittämisen prosessia tarkastellaan ongelman määrittelystä aina ratkaisun käyttöönottoon ja seurantaan asti. Kunkin vaiheen kohdalla otetaan huomioon asiaankuuluvat periaatteet ja sidosryhmien osallistaminen.

- Tekoälyn etiikka ja sen merkitys tekoälyratkaisujen kehittämisessä
- Tekoälyn eettiset periaatteet

- Eettiset viitekehykset, ohjeistukset ja työkalupakit
- Korkean tason asiantuntijaryhmä; Trustworthy AI – viitekehys
- EU:n tekoälyasetus (AI Act)
- Tekoälyn kehittämisen elinkaari
- Sidosryhmien osallistaminen tekoälyn kehittämisen elinkaareissa
- Tekoälyn kehittämisen elinkaari yhdistettynä eettisiin periaatteisiin ja sidosryhmiin

2. Eettisten periaatteiden, kehysten ja ohjeiden määrittely

Tekoäly vaikuttaa elämän kaikkiin osa-alueisiin, muuttaen työelämää ja vapaa-aikaa sekä tarjoten ratkaisuja globaaleihin haasteisiin, kuten ilmastonmuutokseen ja terveydenhuollon saavutettavuuteen. Koska tekoäly integroidaan yhä laajemmin talouksiin ja yhteiskuntiin, herää kysymys siitä, millaiset poliittiset ja hallinnolliset puitteet ovat tarpeen tekoälyn kehityksen ja soveltamisen ohjaamiseksi siten, että siitä koituu hyötyä koko yhteiskunnalle.

Laajat eettiset näkökohdat tekoälyn alalla ovat johtaneet lukuisten lähestymistapojen luomiseen ja julkaisemiseen tekoälyn eettisen käytön varmistamiseksi. (Prem, 2023)

Eettiset periaatteet, viitekehykset ja ohjeistukset ovat yleisesti käytettyjä ja laajasti tunnustettuja eettisen päätöksenteon perusrakenteita monilla aloilla. Ne tarjoavat kattavan lähestymistavan eettisten kysymysten käsittelyyn, perustavanlaatuisten arvojen määrittelystä aina konkreettisten toimintaohjeiden antamiseen. Vaikka nämä käsitteet usein limittyvät toisiinsa, ne voidaan kuvata seuraavasti:

- **Eettiset periaatteet** tarjoavat moraalisen kehyksen eettisten ongelmien arviointiin ja ratkaisemiseen. Ne muodostavat ydinarvot ja uskomukset, jotka toimivat kompassina eettisessä päätöksenteossa ja auttavat määrittämään, mikä on oikein ja mikä väärin eri tilanteissa. (Prem, 2023; Zhou & Chen, 2022)
- **Eettiset viitekehykset** ovat systemaattisia lähestymistapoja tai malleja, jotka auttavat yksilöitä ja organisaatioita navigoimaan eettisissä kysymyksissä. Ne ovat kattavia malleja tai rakenteellisia menetelmiä, jotka helpottavat eettisten ongelmien tarkastelua ja ratkaisemista. Eettiset viitekehykset tarjoavat rakenteen toiminnan eettisten seurausten arvioimiseksi ja ohjaavat päätöksentekoa varmistaen, että eettiset näkökohdat otetaan systemaattisesti huomioon. (Prem, 2023; Zhou & Chen, 2022)
- **Eettiset ohjeistukset** ovat konkreettisia sääntöjä, standardeja ja parhaita käytäntöjä, jotka tarjoavat käytännönläheistä ohjausta eettiseen toimintaan eri yhteyksissä. Ne ovat yksityiskohtaisia sääntöjä ja mittapuita, jotka antavat käytännön neuvoja eettisten periaatteiden soveltamiseen erilaisissa tilanteissa.

3. Yhteinen pyrkimys eettisen maiseman muokkaamiseksi

Eettiset periaatteet, viitekehykset ja ohjeistukset tulevat monilta eri aloilta ja organisaatioilta, kuten yksityisiltä yrityksiltä, tutkimuslaitoksilta ja julkisen sektorin organisaatioilta. Ne edustavat yhteistä pyrkimystä muokata tekoälyn eettistä toimintaympäristöä. (Jobin et al., 2019; Morley et al., 2020)

Tekoälyn eettisyyttä käsittelevät ohjeistukset ja raportit ovat esimerkkejä ei-lakisääteisistä poliittisista välineistä eli niin sanotusta "pehmeästä sääntelystä". Toisin kuin "kova laki", joka sisältää lainsäätäjien laatimia oikeudellisia säädöksiä tiettyjen käytäntöjen pakottamiseksi tai kieltämiseksi, eettisillä ohjeistuksilla ei ole lainvoimaa, mutta niillä on merkittävää vaikutusvaltaa. Nämä asiakirjat on suunniteltu tukemaan päätöksentekoa tietyillä alueilla, ja niiden on todettu vaikuttavan huomattavasti käytäntöihin. (Jobin et al., 2019)

Esimerkkejä tekoälyn eettisiä periaatteita määrittelevistä asiakirjoista:

OECD:n tekoälyperiaatteet edistävät innovatiivista ja luotettavaa tekoälyn käyttöä, joka kunnioittaa ihmisoikeuksia ja demokraattisia arvoja. Toukokuussa 2019 hyväksytyt periaatteet asettavat tekoälylle standardeja, jotka pyrkivät olemaan sekä käytännöllisiä että joustavia. (OECD, 2019)

Pekingin tekoälyakatemia (Beijing Academy of Artificial Intelligence, BAAI) julkaisi "Pekingin tekoälyperiaatteet", jotka toimivat ohjeistuksena tekoälyn tutkimukseen ja kehitykseen, käyttöönottoon sekä hallintaan. ("Beijing Academy of Artificial Intelligence - Beijing AI Principles," 2019; Beijing Artificial Intelligence Principles, 2022)

Suuret teknologiayritykset, kuten Google, IBM, Microsoft ja Intel, ovat kehittäneet omat eettiset ohjeistuksensa, jotka heijastavat niiden ainutlaatuisia näkökulmia ja toimintaympäristöjä. (Jobin et al., 2019; Morley et al., 2020)

Myös hallitukset ja julkishallinnon toimijat ovat osallistuneet tekoälyn eettisten sääntöjen ja politiikkojen kehittämiseen. Keskeisiä asiakirjoja ovat esimerkiksi Montrealin julistus, sekä sellaisten toimielinten panokset kuin Ison-Britannian ylähuoneen tekoälyvaliokunta (House of Lords Select Committee on Artificial Intelligence) ja Euroopan komission asiantuntijaryhmä, jotka kaikki vaikuttavat tekoälyn eettisiin sääntelyihin ja politiikkoihin. (Morley et al., 2020)

Akateemiset instituutiot ja ajatushautomot, kuten Future of Life Institute, IEEE ja AI4People, ovat tuoneet keskusteluun syvyyttä tieteellisen tutkimuksen ja teoreettisten mallien avulla. (Floridi et al., 2018; Jobin et al., 2019; Morley et al., 2020)

Näiden eri toimijoiden yhteinen panos heijastaa globaalia liikettä kohti vastuullista tekoälyä, ja jokainen osallistuminen muodostaa osan laajemmasta

kokonaisuudesta, jonka tavoitteena on rakentaa eettisiä tekoälyjärjestelmiä, jotka ovat linjassa yhteiskunnallisten arvojen ja normien kanssa.



Universitat
de les Illes Balears



Vinkkejä ja suosituksia opettajille

You can ask students to check the above-mentioned sources or some of them, and discuss the differences in the principles they present.

4. Eettiset periaatteet

Tekoälyn eettisten periaatteiden kenttä on laaja ja monimuotoinen, ja lukuisia periaatteita on esitetty eri toimijoiden, kuten akateemisten instituutioiden, teollisuuden johtajien, hallitusten ja kansainvälisten organisaatioiden toimesta. Näiden periaatteiden lisääntyminen kuvastaa kasvavaa yksimielisyyttä siitä, että tekoälyn kehittämiseen ja soveltamiseen tarvitaan eettistä ohjausta. (Jobin et al., 2019; Morley et al., 2020; Prem, 2023)

Jobin et al. (2019) tekivät kattavan katsauksen erilaisiin tekoälyn eettisiin periaatteisiin ja tiivistivät ne tunnistukseen yhteisiä teemoja ja elementtejä, jotka toistuvat eri lähteissä. Tarkastelemiensa ohjeistusten perusteella he havaitsivat, että tekoälyn keskeisistä eettisistä periaatteista on muodostumassa yhteinen ymmärrys. Vaikka näitä periaatteita ei ole yleisesti hyväksytty kaikkialla, yhteneväisyyksiä löytyy muun muassa seuraavissa arvoissa: avoimuus, oikeudenmukaisuus ja tasapuolisuus, vastuullisuus, hyvitys, yksityisyys, hyvän tekeminen, vapaus ja autonomia, luottamus, kestävyys, ihmisarvo ja solidaarisuus.

Tunnistetut yhteiset elementit heijastavat globaalin yhteisön keskeisiä huolenaiheita ja pyrkimyksiä. He viittaavat nousevaan yhteisymmärrykseen arvoista, joiden uskotaan laajasti olevan keskeisiä tekoälyjärjestelmien kehittämisessä ja käyttöönotossa. Tämä alustava sopimus muodostaa perustavanlaatuisen yhteisen lähtökohdan, jota voidaan käyttää odotusten asettamiseen ja tulosten arviointiin. (Morley ym., 2020)

Morley ym. (2020) ovat tunnistaneet tämän kollektiivisen sopimuksen tärkeäksi lähtökohdaksi. Heidän mukaansa nämä yhteisesti hyväksytyt periaatteet luonnehtivat eettisesti sopeutettua tekoälyä seuraavasti:

- Hyödyllinen ja kunnioittava ihmisiä ja ympäristöä kohtaan (hyvän tekeminen, beneficence).
- Vankka ja turvallinen (haitan välttäminen, non-maleficence).
- Ihmisten arvoja kunnioittava (autonomia).
- Oikeudenmukainen (oikeudenmukaisuus, justice).
- Selitettävissä oleva, vastuullinen ja ymmärrettävä (selitettävyyys, explicability).

Hyvän tekeminen (Beneficence)

- **Hyvän tekemisen** periaate (beneficence) on yksi keskeisistä eettisistä näkökohdista tekoälyn kehittämisessä. Se korostaa tekoälyjärjestelmien merkitystä paitsi yksilöiden ja yhteiskunnan hyödyksi myös ympäristön hyvinvoinnin edistämiseksi. Floridin ym. (2018) mukaan tavoitteena on varmistaa, että tekoäly toimii tavalla, joka tukee yleistä hyvinvointia ja ympäristön terveyttä.

Jobin ym. (2019) listaavat useita keskeisiä toimia, joilla tekoälyn kehittämistä voidaan ohjata tehokkaasti:

- Tekoälyn sovittaminen yhteen inhimillisten arvojen kanssa varmistaen, että tekoälyjärjestelmät tukevat ja vahvistavat näitä arvoja sen sijaan, että ne heikentäisivät niitä.
- Tekoälyn sidosryhmien osallistaminen kehitysprosessiin sekä heidän tarpeidensa ja palautteensa huomioiminen, jotta tekoälyjärjestelmien vaikutusta heidän elämäänsä voidaan parantaa.
- Aktiivinen toiminta konfliktien minimoimiseksi ja ratkaisemiseksi, käyttäjäpalautteen hyödyntäminen tekoälyn kehityksen ja soveltamisen ohjaamisessa.
- Uusien tapojen luominen ihmisten hyvinvoinnin mittaamiseen.

Vinkkejä ja suosituksia opettajille

Keskustele opiskelijoiden kanssa tekoälyn ympäristövaikutuksista ja yhteiskunnallisista seurauksista. Esimerkiksi TikTokin algoritmit voivat tarjota käyttäjälle haitallista sisältöä, jos tämä osoittaa kiinnostusta siihen, mikä voi johtaa kielteisiin vaikutuksiin mielenterveyteen. Lisäksi suurten kielimallien luominen kuluttaa huomattavan määrän sähköä, mikä voi aiheuttaa merkittävän hiilidioksidijalanjäljen.

Haitan välttäminen (non-maleficence)

Eettinen periaate haitan välttäminen keskittyy haitan estämiseen. Floridi ym. (2018) mukaan haitan välttäminen kattaa sekä ihmisten tahallista väärinkäyttöä että tekoälyjärjestelmien tahattomia kielteisiä vaikutuksia ennaltaehkäisevät toimet, jotka voivat vaikuttaa ihmisten käyttäytymiseen. Tavoitteena on varmistaa, ettei tekoäly aiheuta haitallisia vaikutuksia riippumatta niiden alkuperästä.

Jobin ym. (2019) ovat tunnistaneet seuraavat lähestymistavat riskien hallitsemiseksi ja haittojen ehkäisemiseksi tekoälyjärjestelmissä:

- Strategisen riskienhallinnan hyödyntäminen tekoälyjärjestelmiin liittyvien potentiaalisten riskien ennakointiin ja vähentämiseen.

- Teknisten keinojen ja hallintopolitiikkojen yhdistelmän käyttöönotto tekoälyn koko elinkaaren ajan, haittojen ennaltaehkäisemiseksi ennen niiden ilmenemistä.
- Joitakin vahinkoja ei voida välttää kaikista ponnisteluista huolimatta, mikä edellyttää kattavaa riskinarviointia. Tämä sisältää toimenpiteitä riskien vähentämiseksi ja lieventämiseksi sekä vastuun selkeää määrittämistä silloin, kun haittoja ilmenee.

Vinkkejä ja suosituksia opettajille

Keskustele opiskelijoiden kanssa tekoälyn mahdollisista haitoista. Esimerkiksi Amazonin tekoälypohjainen rekrytointityökalu osoitti vinoumaa naispuolisia hakijoita kohtaan. Lisätietoja tästä tapauksesta löytyy esimerkiksi täältä: [Insight - Amazon scraps secret AI recruiting tool that showed bias against women | Reuters](#)

Autonomia

Tekoälyn kontekstissa autonomialla tarkoitetaan tasapainon säilyttämistä ihmisten päätöksenteon ja tekoälyn vaikutuksen välillä. Ihmisten on säilytettävä päätöksenteon hallinta ja voitava valita, milloin he tekevät päätöksiä itse ja milloin he antavat tekoälyn päättää. Kuitenkin ihmisten on aina voitava ottaa ohjat takaisin tekoälyltä tarvittaessa, eli heillä on oltava viimeinen sana. (Floridi ym., 2018)

Jobin ym. (2019) tunnistavat joukon toimenpiteitä, joilla edistetään vapautta ja autonomiaa tekoälyjärjestelmissä:

- Tekoälyjärjestelmien läpinäkyvyyden ja ennakoitavuuden parantaminen, jotta käyttäjät voivat ymmärtää ja ennakoida tekoälyn toimintaa.
- Yleisen tietoisuuden ja ymmärryksen lisääminen tekoälyteknologioista.
- Vahvojen tiedottamiseen ja suostumukseen perustuvien menettelyjen käyttöönotto yksilön autonomian turvaamiseksi tekoälyn kanssa toimiessa.
- Henkilötietojen keräämisen ja jakamisen välttäminen ilman asianomaisten nimenomaista ja tietoista suostumusta heidän yksityisyytensä ja itsenäisyytensä kunnioittamiseksi.

Vinkkejä ja suosituksia opettajille

Keskustele opiskelijoiden kanssa tekoälyyn liittyvistä mahdollisista autonomiavajeista. Esimerkiksi Facebook tarjoaa käyttäjille työkaluja sisällön

hallintaan, mutta voivatko algoritmin monimutkaisuus ja läpinäkymättömyys toisinaan heikentää käyttäjän todellista itsemääräämisoikeutta?

Oikeudenmukaisuus (Justice)

Oikeudenmukaisuus on moniulotteinen periaate, jonka tavoitteena on hyödyntää tekoälyä välineenä aiempien epäoikeudenmukaisuuksien korjaamiseksi, varmistaa tekoälyn hyötyjen oikeudenmukainen jakautuminen sekä suojella uusien haittojen tai yhteiskunnallisten häiriöiden syntymiseltä. (Floridi ym., 2018)

Jobin ym. (2019) mukaan oikeudenmukaisuuden toteutuminen tekoälyssä edellyttää useita ennakoivia toimenpiteitä:

- Teknisten standardien ja normien luominen.
- Oikeuksien ja sääntelyn läpinäkyvyyden lisääminen.
- Testaus, valvonta ja auditointi, erityisesti tietosuojayksiköissä.
- Oikeusvaltion vahvistaminen, mukaan lukien valitusmenettelyt ja oikeussuojakeinot.
- Systemaattiset muutokset, kuten valtion valvonta, monipuoliset tiimit ja osallistava kansalaisyhteiskunta, keskittyen hyötyjen oikeudenmukaiseen jakautumiseen.

Vinkkejä ja suosituksia opettajille

Tutustu esimerkiksi Microsoftin chatbottiin Tayhin: [In 2016, Microsoft's Racist Chatbot Revealed the Dangers of Online Conversation - IEEE Spectrum](#). Algoritmi priorisoi epäoikeudenmukaisesti sensaatiomaisen sisällön mielekkään ja rakentavan keskustelun sijaan, mikä johti näkemysten ja äänten epätasapainoiseen edustukseen alustalla ja aiheutti yhteiskunnallisia häiriöitä.

Selitettävyyden (Explicability)

Selitettävyyden on keskeistä käyttäjien luottamuksen rakentamisessa ja ylläpitämisessä tekoälyjärjestelmiin. Tämä tarkoittaa, että prosessien on oltava läpinäkyviä, tekoälyjärjestelmien kyvykkyudet ja tarkoitus on viestittävä avoimesti, ja päätökset on selitettävä niille, joihin ne vaikuttavat suoraan tai epäsuorasti. (High-Level Expert Group on AI (AI HLEG), 2019) Myös Floridi ym. (2018) korostavat, että selitettävyyden on keskeinen tekijä muiden eettisten periaatteiden soveltamisessa.

- Läpinäkyvyyden lisäämiseksi tekoälyjärjestelmissä Jobin ym. (2019) esittävät useita lähestymistapoja:
- Tekoälyn kehittäjiä ja käyttäjiä kannustetaan jakamaan enemmän tietoa järjestelmistään, jotta tekoälypäättöskenteen taustalla olevat prosessit eivät jäisi mysteeriksi.
- Tekoälyn toiminnan ja päätösten selittäminen helposti ymmärrettävällä tavalla myös maallikoille, jotta tekoälyn toiminta voidaan arvioida myös sellaisten ihmisten toimesta, jotka eivät ole alan asiantuntijoita.

- Auditointien käyttö läpinäkyvyyden varmistamiseksi tekoälyssä; niiden avulla voidaan tunnistaa ja korjata mahdollisia vinoumia tai eettisiä ongelmia.
- Mekanismien luominen tekoälyn suorituskyvyn seuraamiseksi, sidosryhmien ja kansalaisten osallistaminen monipuolisten näkemysten keräämiseksi sekä mekanismien kehittäminen ilmiantajien tukemiseksi.

Vinkkejä ja suosituksia opettajille

Kysy opiskelijoilta, kokevatko he olevansa riittävän tietoisia niistä päätöksistä, joita heidän käyttämänsä tekoälyratkaisut tekevät. Esimerkiksi Amazonin tekoälyn perustuva rekrytointityökalu ei tarjonnut läpinäkyvyyttä päätöksentekoprosessissa. Ehdokkaille ja rekrytoijille ei annettu selkeitä selityksiä sille, miksi jotkut hakijat hylättiin, mikä vaikeutti järjestelmän taustalla olevien vinoumien haastamista tai korjaamista.

HUOM! Käytä periaatteet-osuuden lopussa annettua tapaustutkimusta ryhmätyönä.

5. Eettiset viitekehykset, ohjeet ja työkalupakit

Kun eettiset periaatteet on tunnistettu, ne tulisi muuntaa käytännönläheisiksi viitekehyksiksi ja strategioiksi, jotka ohjaavat tekoälypohjaista innovointia ja tekoälyn eettisten periaatteiden käytännön toteuttamista.

Aylingin ja Chapmanin (2022) mukaan eettiset viitekehykset muuttavat eettiset periaatteet käytännön toimenpiteiksi. Tämä prosessi sisältää ohjeiden ja menettelytapojen laatimisen, joita tekoälyteknologioiden kehittäjät ja käyttäjät voivat hyödyntää, ja joiden avulla eettiset näkökohdat voidaan sisällyttää jokaiseen vaiheeseen suunnittelusta käytännön sovelluksiin.

Prem (2023) täydentää, että monet eettiset tekoälyviitekehykset pyrkivät aktiivisesti ennakoidaan ja ratkaisemaan mahdollisia eettisiä ongelmia ja riskejä. Tunnistamalla nämä haasteet ennakoiden, viitekehykset ohjaavat kehittäjiä luomaan tekoälyratkaisuja. Tämä ennakoiva lähestymistapa on välttämätön nopeasti kehittyvässä tekoälyn kentässä, jossa uusia eettisiä kysymyksiä voi nousta esiin teknologian kehittyessä ja integroitua yhä laajemmin arkipäivän eri osa-alueisiin.

Ettinen viitekehys tarjoaa kattavan suunnitelman sellaisten sääntely- ja toimintastandardien luomiseksi, jotka ovat sopusoinnussa eettisten normien kanssa. Floridin ja Cowlsin (2019) mukaan tällainen viitekehys on olennainen lainsäädännön laatimisessa, sääntöjen muotoilussa, teknisten standardien määrittämisessä sekä parhaiden käytäntöjen tunnistamisessa eri toimialoilla ja maantieteellisillä alueilla.

Prem (2023, s. 701) määrittelee eettisten viitekehysten olennaiset osatekijät seuraavasti:

- Keskeiset käsitteet, jotka liittyvät eettisten näkökohtien tarkasteluun
- Eettiset periaatteet (esim. arvot)Huoli siitä, miten tekoälyjärjestelmien kehittäminen ja käyttö noudattavat eettisiä periaatteita
- Keinot huolenaiheisiin vastaamiseksi (esim. strategiat, säännöt ja ohjeistukset)

Eettisten periaatteiden soveltamiseen suunnittelussa, toteutuksessa ja käyttöönotossa tarvitaan kipeästi työkaluja ja ohjeistuksia. (Zhou ym., 2020) Eettiset työkalupakit ja ohjeistukset ovat tärkeitä välineitä, joiden avulla tekoälyinnovaatiota voidaan ohjata eettisten normien ja arvojen mukaiseksi. Niiden nähdään olevan keskeisessä roolissa siltana abstraktien eettisten periaatteiden ja niiden konkreettisen soveltamisen välillä tekoälyn käytännöissä.

Vaikka työkalujen ja ohjeistusten merkitys on selvä, periaatteiden muuttaminen konkreettisiksi työkaluiksi on haastavaa, ja monet olemassa olevat viitekehykset puuttuvat soveltamiskontekstia. (Prem, 2023, s. 702) Suuren kuilun periaatteiden ja käytännön välillä selittävät standardoinnin puute, vaihtelevuus, monimutkaisuus sekä abstraktien periaatteiden subjektiivinen tulkinta.

Tekoälyn viimeaikainen nopea kehitys on johtanut sekä saatavilla olevien työkalujen määrän että monimuotoisuuden merkittävään kasvuun, ja uusia työkaluja kehitetään jatkuvasti. Morley ym. (2019) ja Prem (2023) ovat tarkastelleet satoja erilaisia työkaluja, jotka on tarkoitettu tekoälyratkaisujen eettisten haasteiden käsittelyyn. Työkalut on jaettu tekoälyratkaisun suunnittelun eri vaiheiden (määrittämään myöhemmin luvusta H alkaen) ja eri eettisten periaatteiden mukaan.

Monia työkaluja kuitenkin pidetään vaikeakäyttöisinä tai sellaisina, että ne vaativat käyttäjältään huomattavan määrän vaivannäköä.

Morley ym. (2019) ja Prem (2023) kokoamat listaukset osoittavat selvästi, että työkalujen ja menetelmien saatavuus jakautuu epätasaisesti eri eettisten periaatteiden ja suunnitteluvaiheiden välillä. Esimerkiksi ratkaisun hyödyn (beneficence) arviointiin löytyy runsaasti työkaluja kehityksen alkuvaiheessa, kun taas selitettävyyttä (explicability) -työkalut sijoittuvat pääasiassa kehitysprosessin loppuvaiheeseen.

Katso työkalut tämän linkin kautta: [Applied AI Ethics Typology - Google Docs](#).

Myös käytetyt menetelmät vaihtelevat kehitysvaiheen mukaan. Kehyksillä, kirjastoilla ja algoritmeilla on usein suurempi rooli aikaisemmassa vaiheessa, erityisesti ratkaisun suunnitteluvaiheessa. Jotkin viitekehykset kuitenkin ulottuvat sekä aikaisempaan että jälkimmäiseen vaiheeseen. (UNESCO, 2023b, s. 5) Testausvaiheessa taas auditoinnit ja mittarit ovat yleisimpiä menetelmiä, kun taas jälkimmäisessä vaiheessa hyödynnetään pääasiassa julkilausumia, merkintöjä ja lisensejä. (Prem, 2023, s. 709)

Tutkimus toi myös esiin, että vain harvat työkalut käsittelevät ratkaisun vaikutuksia yksilöön tai yhteiskuntaan kokonaisuutena. Tämä saattaa johtua siitä, että monimutkaisen inhimillisen käyttäytymisen muuttaminen yksinkertaisiksi ja yleisiksi suunnittelutyökaluiksi on haastavaa. Kuitenkin ihmiseen ja yhteiskuntaan kohdistuvien vaikutusten huomioiminen olisi olennaista, jotta tekoälyratkaisuille saavutettaisiin hyväksyttävyyttä ja sosiaalinen suositeltavuus. (Morley ym., 2020, s. 2156)

6. Periaatteista käytäntöön; Luotettavan tekoälyn viitekehys

Vaikka tekoälyn eettisyydestä vallitsee laaja yhteisymmärrys, keskustelua käydään edelleen siitä, mitä "eettinen tekoäly" tarkalleen tarkoittaa sekä siitä, millaisia eettisiä velvoitteita ja teknisiä kriteerejä sen toteuttaminen edellyttää. (Leijnen ym., 2020)

Euroopan komissio on laatinut kattavan strategian tekoälyn kehittämiseksi Euroopassa, ja strategian keskiössä on varmistaa, että kehitetty ja käytetty tekoäly ei ole ainoastaan teknologisesti kehittyntä, vaan myös eettisten periaatteiden mukaista ja turvallista.

Tämän strategian ja sen taustalla olevien periaatteiden tehokkaan toimeenpanon tueksi Euroopan komissio perusti tekoälyn korkean tason asiantuntijaryhmän (High-Level Expert Group on Artificial Intelligence, AI HLEG). Ryhmän tehtävänä oli laatia tekoälyn eettiset ohjeet sekä politiikka- ja investointisuositukset, joiden tarkoituksena on ohjata tekoälyn eettistä kehittämistä, käyttöönottoa ja hyödyntämistä Euroopassa. Tavoitteena on edistää innovointia samalla kun turvataan perusarvot ja -oikeudet. (High-Level Expert Group on AI (AI HLEG), 2019)

Luotettavan tekoälyn viitekehys on suunniteltu varmistamaan, että tekoälyteknologiat kehitetään ja otetaan käyttöön eettisesti kestävästi, perusoikeuksia kunnioittaen sekä kestävästi ja luotettavasti toimien. Luotettavan tekoälyn viitekehys koostuu kolmesta osa-alueesta, jotka kaikki on suunniteltu ohjaamaan tekoälyn kehittämistä ja käyttöönottoa eettisesti ja luotettavasti. (High-Level Expert Group on AI (AI HLEG), 2019)

Luotettavan tekoälyn perustat

Tämä osa-alue määrittelee luotettavan tekoälyn taustalla olevat keskeiset periaatteet ja omaksuu lähestymistavaksi perusoikeuksiin perustuvan näkökulman. Se jäsentää eettiset periaatteet, jotka ovat olennaisia sen varmistamiseksi, että tekoälyjärjestelmät kehitetään ja niitä käytetään sekä eettisesti että kestävästi.

Eettiset periaatteet muodostavat perustan, jolle luotettava tekoäly on rakennettava, jotta eettisten standardien noudattaminen voidaan varmistaa. Nämä eettiset periaatteet ovat: (i) Ihmisen autonomian kunnioittaminen (ii) Haitan ehkäiseminen (iii) Oikeudenmukaisuus (iv) Selitettävyys.

Luotettavan tekoälyn toteuttaminen

Eettisten periaatteiden pohjalta viitekehyksen toinen osa muuntaa nämä periaatteet seitsemäksi keskeiseksi vaatimukseksi, joita tekoälyjärjestelmien odotetaan ilmentävän ja noudattavan koko elinkaarensa ajan. Tämä muutos eettisistä periaatteista käytännön vaatimuksiksi varmistaa, että perustavanlaatuinen etiikka ei jää pelkiksi teoreettisiksi ihanteiksi, vaan ne integroidaan aktiivisesti tekoälyn kehityksen ja toiminnan jokaiseen vaiheeseen.

Vaatimukset ovat seuraavat: Ihmisen edustus ja valvonta: perusoikeudet, ihmisen toimijuus ja ihmisen valvonta; Tekninen kestävyys ja turvallisuus: kyky kestää hyökkäyksiä ja tietoturva, varasuunnitelma ja yleinen turvallisuus, tarkkuus, luotettavuus ja toistettavuus; Yksityisyys ja datan hallinta: yksityisyyden kunnioittaminen, datan laatu ja eheys sekä pääsy dataan; Läpinäkyvyys: jäljitettävyyden, selitettävyyden ja viestintä; Monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus: epäoikeudenmukaisen vinouman välttäminen, saavutettavuus ja universaali suunnittelu sekä sidosryhmien osallistaminen; Yhteiskunnallinen ja ympäristöllinen hyvinvointi: kestävyys ja ympäristöystävällisyys, sosiaalinen vaikutus, yhteiskunta ja demokratia; Vastuullisuus: auditointimahdollisuus, haitallisten vaikutusten minimointi ja raportointi, kompromissien arviointi ja oikaisukeinot..

Luotettavan tekoälyn arviointi

Viitekehyksen kolmas osa tarjoaa konkreettisen ja kattavan listauksen luotettavan tekoälyn arvioimiseksi, ja sen tarkoituksena on operationalisoida edellä esitetyt vaatimukset.

Tämä työkalu, joka tunnetaan nimellä Assessment List for Trustworthy AI (ALTAI), tarjoaa kattavan ja dynaamisen tarkistuslistan, joka toimii tiekarttana kehittäjille ja käyttäjille eettisten periaatteiden integroimiseksi tekoälysovelluksiin. ALTAI tukee tätä prosessia tarjoamalla konkreettisia itsearviointivaiheita, joiden avulla voidaan varmistaa, että tekoälyteknologiat kehitetään tavalla, joka maksimoi käyttäjän hyödyt ja minimoi tarpeettomat riskit. (High-Level Expert Group on AI (AI HLEG), 2020)

Tämä arviointilista toimii käytännön oppaana tekoälyn parissa työskenteleville ja tarjoaa yksityiskohtaista ohjeistusta vaatimusten tehokkaaseen toteuttamiseen. Arvioinnin on tarkoitus olla mukautuva, mikä mahdollistaa sen räätälöinnin kunkin tekoälyjärjestelmän erityistarkoitukseen — varmistaen näin sen soveltuvuuden ja vaikuttavuuden erilaisissa konteksteissa.

7. Periaatteista lainsäädäntöön: EU:n tekoälylaki

EU:n tekoälyasetus (EU AI Act) tulee säätämään tekoälyn soveltamista Euroopan unionin alueella, ja se on maailman ensimmäinen laaja-alainen tekoälyä koskeva lainsäädäntö. (Euroopan parlamentti, 2023) Asetuksen

tavoitteena on sisämarkkinoiden toimivuuden parantaminen sekä ihmiskeskeisen ja luotettavan tekoälyn käyttöönoton edistäminen. Samalla asetuksen tarkoituksena on varmistaa korkea terveyden, turvallisuuden ja perusoikeuksien, mukaan lukien demokratian, oikeusvaltion ja ympäristönsuojelun taso suojautumalla tekoälyjärjestelmien haitallisilta vaikutuksilta unionin alueella sekä tukea innovointia. (Future of Life Institute, 2024) Ennusteiden mukaan asetus tulee voimaan vuonna 2026. (Hillard & Gulley, 2023)

Euroopan komissio on jäsentänyt tekoälyä koskevat säännökset riskiperusteiseen viitekehykseen, jossa tekoälysovellukset luokitellaan niiden aiheuttaman riskitason mukaan. (Euroopan komissio, 2023) EU:n tekoälyasetuksen riskiperusteisessa lähestymistavassa tekoälyjärjestelmät jaetaan neljään luokkaan niiden mahdollisten vaikutusten perusteella: kielletyt riskit, korkea riski, rajallinen riski ja vähäinen tai minimaalinen riski. (Euroopan komissio, 2023; Hillard & Gulley, 2023))

Vähäinen riski

Suurin osa tekoälyjärjestelmistä kuuluu tähän luokkaan, eikä niillä ole juuri lainkaan tai ollenkaan vaikutusta kansalaisten oikeuksiin tai turvallisuuteen. Esimerkkejä ovat tekoälypohjaiset suosittelujärjestelmät ja roskapostisuodattimet. Näitä järjestelmiä käyttävien yritysten ei tarvitse noudattaa tiukkoja vaatimuksia, mutta ne voivat vapaaehtoisesti sitoutua lisätoimenpiteisiin, kuten läpinäkyvyyttä ja vastuullisuutta edistäviin toimintaohjeisiin.

Rajallinen riski

Tähän luokkaan kuuluvat tekoälyjärjestelmät aiheuttavat vain rajallisen riskin käyttäjien oikeuksille ja turvallisuudelle. Esimerkkejä ovat chatbotit ja deepfake-sisällöt. Vaikka näitä ei säädellä tiukasti, palveluntarjoajien on varmistettava, että käyttäjät tietävät olevansa vuorovaikutuksessa tekoälyn kanssa ja että tekoälyn tuottama sisältö on asianmukaisesti merkitty. Tämä edistää läpinäkyvyyttä ja käyttäjien tietoisuutta tekoälyn käytöstä.

Korkea riski

Tekoälyjärjestelmät, jotka on tunnistettu korkean riskin luokkaan kuuluviksi, on alistettu tiukkoihin vaatimuksiin mahdollisten riskien vähentämiseksi. Näihin vaatimuksiin kuuluvat muun muassa vahvat riskinhallintajärjestelmät, korkealaatuiset tietoaaineistot, toiminnan lokitus, yksityiskohtainen dokumentaatio, selkeä käyttäjätiedotus, ihmisen valvonta sekä vahvat kyberturvallisuustoimet. Vastuullista innovaatiota edistetään myös säätelyhiekkalaatikoilla (regulatory sandboxes), jotka tukevat säännösten mukaisten tekoälyjärjestelmien kehittämistä.

Kielletty riski

Tekoälyjärjestelmät, jotka muodostavat selvän uhan perusoikeuksille, kielletään kokonaan. Tähän sisältyvät järjestelmät, jotka manipuloivat ihmisten käyttäytymistä heikentäen heidän vapaan tahtonsa toteutumista – esimerkiksi lelut, jotka kannustavat lapsia vaaralliseen käyttäytymiseen, tai hallitusten tai yritysten käyttämät “sosiaalisen pisteytyksen” järjestelmät. Myös ennakoivaa rikosten ehkäisyä (predictive policing) koskevat sovellukset sekä tietyt biometriset järjestelmät, kuten tunneanalyysi työpaikoilla, kuuluvat kiellettyihin käyttötarkoituksiin.

Vinkkejä ja suosituksia opettajille

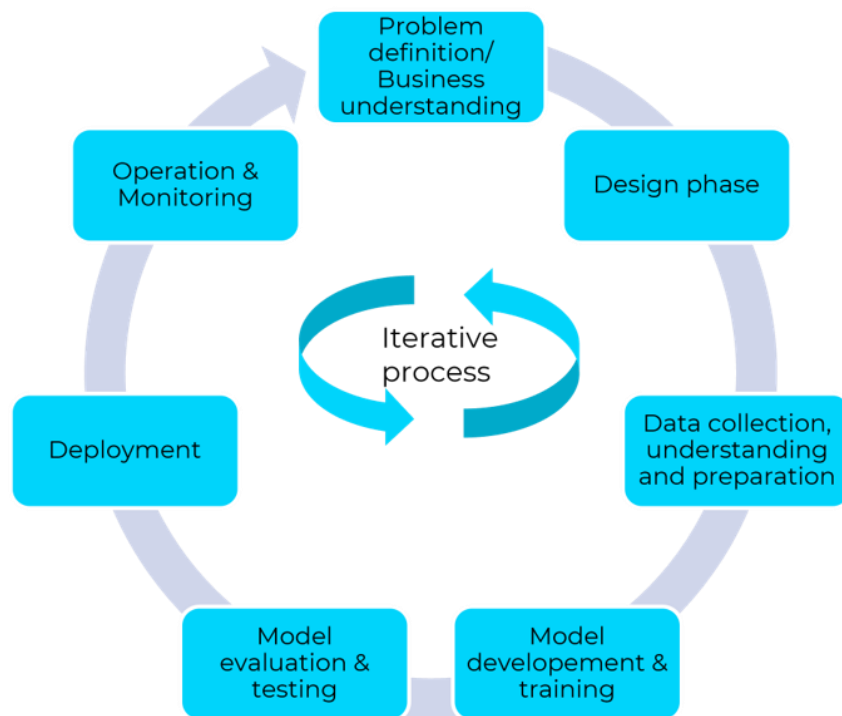
- 1) Keskustele opiskelijoiden kanssa siitä, ovatko he sitä mieltä, että EU:n laki on tiukempi kuin muut lait, ja mitä se aiheuttaa. Ajattelemisen aihetta esimerkiksi tästä artikkelista: <https://news.bloomberglaw.com/artificial-intelligence/eu-embraces-new-ai-rules-despite-doubts-it-got-the-right-balance>
- 2) Jos ajatellaan viimeaikaista keskustelua TikTokista ja sen haitallisuudesta, mihin riskiluokkaan se kuuluisi?

8. Tekoälyn kehittämisen elinkaari

Kuten luvussa E tarkastellaan, tietyt eettiset periaatteet korostuvat eri tavoin tekoälyjärjestelmän kehityksen eri vaiheissa, ja niiden tueksi käytetään erilaisia työkaluja. Kehitysprosessin ymmärtäminen laajasta näkökulmasta on keskeistä. Tässä kurssissa käytämme käsitettä "tekoälyn kehityksen elinkaari" (AI Development Lifecycle) kuvaamaan tätä kokonaisvaltaista näkemystä kehitysvaiheista.

Nopeasti kehittyvässä tekoälyratkaisujen kentässä projektit epäonnistuvat toisinaan puutteellisen projektinhallinnan vuoksi. Hyvin määritelty tekoälyn elinkaari, joka huomioi kaikki kehitysvaiheet, voi parantaa tekoälyratkaisujen onnistumisastetta. Vaikka tekoälyn elinkaaresta on olemassa useita eri kuvauksia, meidän lähestymistapamme yhdistelee Morley et al. (2020), Prem (2023) ja Data Science Process Alliance -organisaation CRISP-mallia (Hotz, 2018), kuten Rochel & Evequoz (2021) sen esittävät.

Tekoälyn kehityksen elinkaari kuvaa tekoälyjärjestelmän matkaa ideasta käyttöönottoon ja ylläpitoon. Jokainen vaihe on tärkeä onnistuneen lopputuloksen varmistamiseksi. Kehityksen elinkaari on iteratiivinen, mikä tarkoittaa, että aiempiin vaiheisiin voidaan joutua palaamaan, jos jotkin muuttujat eivät toimi odotetulla tavalla. Vaikka eettiset näkökulmat voivat painottua eri tavoin kehityksen eri vaiheissa, niiden integrointi jokaiseen vaiheeseen on olennaista. (Saltz, 2023))



Kuva: Asiakirjan kirjoittajat. Muokattu yhdistelmänä Data Science Process Alliancen (2023), Morleyn ym. (2019), Prehmin (2022) ja Rochelin ym. kaikkien (2020) artikkeleissa kuvatuista eri malleista.

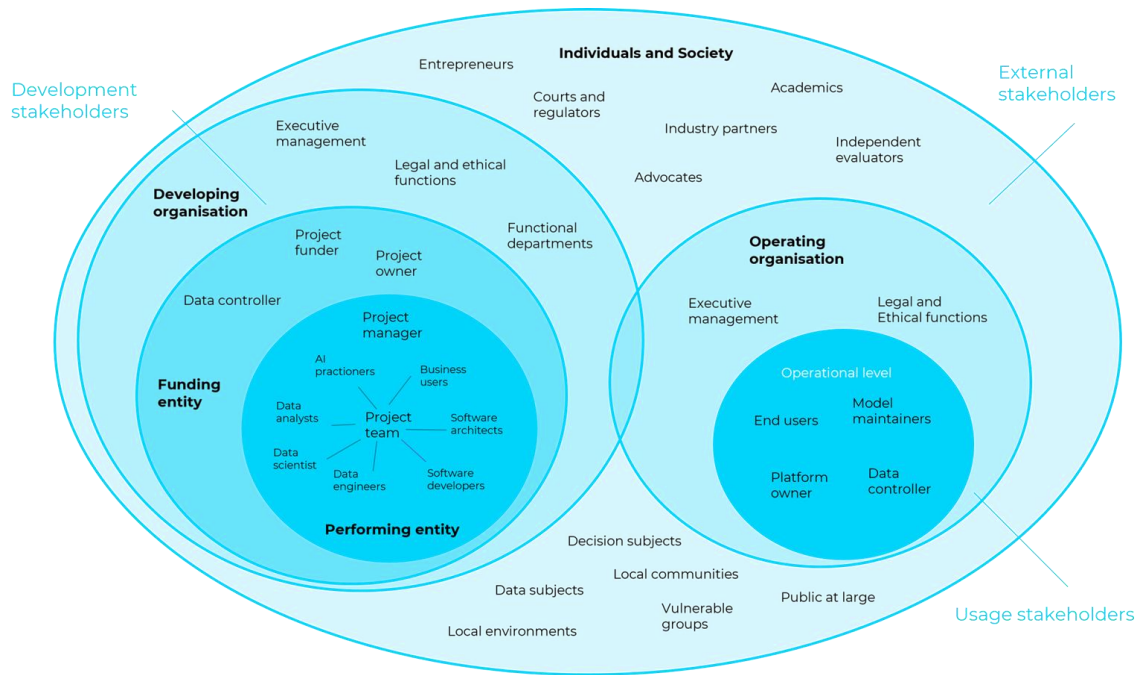
9. Sidosryhmien osallistuminen tekoälyn kehittämisen elinkaareissa

Monipuolisen sidosryhmäjoukon osallistaminen tekoälyn kehitysprosessiin on olennaista, jotta tekoälyhankkeen eettisiä vaikutuksia voidaan arvioida kattavasti eri näkökulmista. Sidosryhmien kuuleminen – mukaan lukien ne, joihin tekoälyratkaisu vaikuttaa suoraan – on tärkeää luottamusta herättävien järjestelmien rakentamisessa (Morley et al., 2020) ja hankkeen vaikutusten kokonaisvaltaisessa arvioinnissa (UNESCO, 2023a, s. 15). Tässä luvussa tarkastellaan tekoälykehityshankkeisiin osallistuvia eri sidosryhmiä ja kuvataan heidän roolinsa tekoälyn kehityksen elinkaaren eri vaiheissa.

Sidosryhmät jaetaan kehityksen sidosryhmiin, käytön sidosryhmiin ja ulkoisiin sidosryhmiin. Kehityksen ja käytön sidosryhmät ovat aktiivisesti mukana hankkeessa, kuten päätöksentekijät, joilla on valtaa tehdä ratkaisuja, ja

suunnittelijat, joilla on asiantuntemusta. Ulkoiset sidosryhmät eivät osallistu suoraan tekoälyhankkeeseen, mutta voivat vaikuttaa siihen tai olla sen vaikutusten kohteena – tai heitä voidaan konsultoida. (Miller, 2022)

Seuraavaksi tarkastellaan eri sidosryhmien keskeisiä rooleja tekoälyn kehityksessä ja luokitellaan ne kolmeen ryhmään: kehityksen, käytön ja ulkoisiin sidosryhmiin. Lisäksi perehdytään kunkin ryhmän erityisiin vastuualueisiin ja panoksiin tekoälyn kehityksen elinkaaren eri vaiheissa. Näiden roolien ymmärtäminen on tärkeää, jotta voidaan selviytyä tekoälyn kehittämisen monimutkaisista kysymyksistä ja varmistaa projektin onnistuminen.



Kuva: G.J. Miller (2022)

Kehityksen sidosryhmät

Tekoälyprojektien kehityksessä Miller (2022) jakaa sidosryhmät kolmeen pääkategoriaan: organisaatio, rahoittava taho ja toteuttava taho, joista jokaisella on oma selkeä roolinsa projektin elinkaareissa. Organisaatio hyötyy tekoälyprojektin lopputuloksista hyödyntäen saavutuksia ja innovaatioita oman kasvunsa ja menestyksensä tukena. Rahoittava taho on ratkaisevassa roolissa rahoituksen tarjoajana, mutta se myös ohjaa projektia varmistaen, että investoinnit ovat linjassa projektin tavoitteiden ja hallintastandardien kanssa. Toteuttava taho on keskeinen projektin tulosten tuottamisessa ja itse tekoälyjärjestelmän rakentamisessa, vastaten projektin teknisestä toteutuksesta. (Miller, 2022)

Organisaation sisällä roolit vaihtelevat kehitysorganisaatiosta, joka sponsoroi projektia ja määrittelee sen laajuuden, aina ylimpään johtoon, joka seuraa projektin etenemistä. Toiminnalliset osastot tarjoavat keskeistä tukea, kun taas oikeudelliset ja eettiset toiminnot varmistavat, että projekti on linjassa

sääntelyvaatimusten ja eettisten periaatteiden kanssa. Projektinhallintaryhmä toimii tukevassa roolissa, avustaen projektipäälliköitä hallinnollisissa tehtävissä projektin sujuvan toteutuksen varmistamiseksi (Zwikaël & Meredith, 2018).

Rahoittavaan tahoon kuuluu rooleja, kuten projektin rahoittaja, joka tarjoaa taloudelliset resurssit, sekä projektinomistaja, joka määrittää strategisen suunnan ja saattaa toimia ohjausryhmän puheenjohtajana varmistaen, että projekti etenee strategisten tavoitteiden mukaisesti. Mikäli rahoittaja ei ole suoraan mukana projektissa, vastuu siirretään toiselle taholle jatkuvan valvonnan varmistamiseksi. Rahoittavaan tahoon liitetään usein myös tietosuojavastaavan rooli, joka hallinnoi datan käyttöä ja varmistaa, että sääntely- ja lainsäädäntövaatimuksia noudatetaan, jotta mahdollisilta sanktioilta välttytään (Miller, 2022).

Toteuttavan tahon ytimessä ovat projektipäälliköt ja projektitiimi, jotka vastaavat projektin lopputulosten toimittamisesta hyväksytyn suunnitelman mukaisesti. Tämä yhteistyö varmistaa, että projekti etenee ideasta toteutukseen, ja että jokainen tiimin jäsen osallistuu tekoälyjärjestelmän kehittämiseen ja onnistuneeseen käyttöönottoon (Zwikaël & Meredith, 2018).

Tämä sidosryhmien ja heidän rooliensa jäsennelty organisointi korostaa yhteistyöhön perustuvaa ja moniulotteista lähestymistapaa, joka on välttämätön tekoälyprojektien onnistuneessa kehittämisessä ja käyttöönotossa, tuoden esiin selkeiden vastuiden ja lakien sekä eettisten standardien noudattamisen tärkeyden.

Käytön sidosryhmät

Tekoälyprojekteissa käytön sidosryhmillä on keskeinen rooli, erityisesti organisaatiotasolla, jossa operatiivinen organisaatio toimii asiakassidosryhmänä, joka hankkii tekoälyratkaisun. Kehitysorganisaation tavoin operatiiviseen organisaatioon kuuluu rooleja, kuten ylimmän johdon sekä oikeudellisten ja eettisten toimintojen edustajat, mikä korostaa samankaltaisia rakenteita niin kehitys- kuin käyttövaiheessakin (Miller, 2022).

Operatiivisella tasolla käyttöön liittyy useita keskeisiä toimijoita, kuten tietovastaavat, loppukäyttäjät, alustanhaltijat ja mallien ylläpitäjät. Loppukäyttäjät – olipa kyse yksilöistä tai ryhmistä – ovat vuorovaikutuksessa tekoälyjärjestelmän kanssa joko työ- tai palvelusopimuksen kautta operatiivisen organisaation kanssa tai kuluttajina, jotka käyttävät järjestelmän toimintoja suoraan (Miller, 2022).

Mallien ylläpitäjät vastaavat koneoppimismallien jatkuvasta hallinnasta, päivittämisestä ja kehittämisestä, varmistaen järjestelmän tehokkuuden ja toimivuuden. Tietosuojavastaavat määrittelevät henkilötietojen käsittelyn tavoitteet ja menetelmät järjestelmässä, huolehtien tietosuojalakien noudattamisesta.

Lopuksi alustanhaltijat kantavat vastuun tekoälyjärjestelmän kokonaisinfrastruktuurista ja käyttöönnotosta, ylläpitäen tekoälyn toiminnan perustana olevaa teknistä ympäristöä (Miller, 2022).

Ulkoiset sidosryhmät

Miller (2022) jakaa tekoälyprojektien ulkopuoliset sidosryhmät kolmeen pääalakategoriaan: yksilöt, yhteiskunta ja edustajat, joilla jokaisella on omat erityiset roolinsa ja panoksensa projektiin.

Yksilöiden alakategoria kattaa rekisteröidyt henkilöt, päätöksenteon kohteet ja työntekijät, jotka ovat vuorovaikutuksessa kehitys- tai käyttöorganisaation kanssa esimerkiksi palvelu- tai työsopimusten kautta. Näihin sidosryhmiin kuuluvat henkilöt, jotka ovat suoraan mukana projektissa tai joita projekti koskettaa. Heillä on keskeisiä rooleja rekisteröityinä henkilöinä, päätöksenteon kohteina ja työntekijöinä (Miller, 2022).

Yhteiskunnan alakategoria kuvaa laajempaa vaikutusta väestöön yleisesti, mukaan lukien paikallisyhteisöt, ympäristö sekä haavoittuvat ryhmät, kuten vammaiset, vähemmistöt, alaikäiset ja suuri yleisö. Tämä ryhmä on suoraan tekoälyjärjestelmän kehittämisen ja käytön vaikutuspiirissä, mikä korostaa tekoälyprojektien yhteiskunnallista ulottuvuutta (Miller, 2022).

Edustajat, kolmas alakategoria, eroavat muodollisista edustajista. Vaikka edustajiin kuuluvat ryhmät ja organisaatiot, kuten yrittäjät ja riippumattomat arvioijat (raportoijat, tilintarkastajat tai arvioijat), jotka eivät välttämättä ole suoraan tekoälyjärjestelmän vaikutuspiirissä, mutta osallistuvat sen kehittämiseen tai arviointiin, muodolliset edustajat, kuten tuomioistuimet, lait, sääntelyviranomaiset ja ammattiliitot, edustavat projektin oikeudellista ja sääntelyyn liittyvää valvontaa ja hallintoa (Miller, 2022; Rodrigues, 2020).

Sidosryhmien osallistaminen tekoälyratkaisujen kehittämisessä

Yhteiskehittäminen tarkoittaa yhteistyöhön perustuvaa osallistamista sidosryhmien, mukaan lukien loppukäyttäjien tai asiakkaiden, kanssa koko suunnitteluprosessin ajan. Tämä lähestymistapa edistää näkökulmien ja oivallusten vaihtoa osallistujien kesken, joilla on erilaisia rooleja ja asiantuntemusta. (Interaction Design Foundation – IxDF, 2021; Sanin, 2020)

Yhteiskehittämistä voidaan toteuttaa fasiloituissa työpajoissa, jotka sisältävät erilaisia toimintoja, kuten tutustumisleikkejä, roolipelaamista, asiakaspolkujen kartoittamista, ideointia ja prototypointia. Yhteiskehittämisen avulla muotoilijat voivat syvällisesti ymmärtää asiakkaiden tai loppukäyttäjien tarpeita ja toiveita, ja näin luoda ratkaisuja, jotka huomioivat monipuolisesti eri käyttäjien näkökulmat. Tämä osallistava menetelmä varmistaa, että kehitettävät palvelut vastaavat aidosti niiden käyttäjien tarpeita, joita ne on tarkoitettu palvelemaan. (Interaction Design Foundation – IxDF, 2021; Sanin, 2020)

10. Tekoälyn kehittämisen elinkaari: ongelman määrittäminen ja liiketoiminnan ymmärtäminen

Vinkkejä ja suosituksia opettajille

Tässä osiossa selitetään yksityiskohtaisesti tekoälyratkaisun kehittämisen elinkaari, sen eettiset vaikutukset ja siihen liittyvät sidosryhmät. Käyttäkää esimerkkitapausta koko elinkaaren läpikäymiseksi, jotta keskustelu olisi aktiivista. Esimerkkinä voit käyttää tekoälyyn perustuvaa suosittelujärjestelmää ostosten tekemisestä tämän linkin kautta: https://www.datascience-pm.com/ai-lifecycle/#An_Example_Ai_Project_Life_Cycle

Kuten kaikissa projekteissa, myös tekoälyn elinkaari tulisi aloittaa ongelman määrittelystä. Hyvin määritelty ongelmanasettelu antaa syvällisen ymmärryksen asiakkaan tarpeesta ja ohjaa projektin suuntaa. Tässä vaiheessa tulisi tehdä toteutettavuustutkimus, jonka avulla selvitetään, onko tekoäly sopivin ratkaisu kyseisen ongelman ratkaisemiseksi. Lisäksi ongelman perusteellinen ymmärtäminen edellyttää tutkimusta ja sidosryhmien osallistamista. (De Silva & Alahakoon, 2022, s. 4–5)

Eettiset periaatteet ongelman määrittelyn ja liiketoiminnan ymmärtämisen vaiheessa

Projektin alkuvaiheessa on tärkeää huomioida keskeiset eettiset näkökohdat, sillä ne luovat perustan projektin kululle ja sen myöhemmälle yhteiskunnalliselle vaikutukselle. Erityisesti tässä vaiheessa korostuvat hyvän tekemisen (beneficence) ja haitan välttämisen (non-maleficence) periaatteet. (Prem, 2023, s. 703)

Hyvän tekemisen periaatteen osalta keskitytään arvioimaan, edistääkö tekoälyjärjestelmä aktiivisesti yksilön hyvinvointia ja yhteiskunnallista hyötyä. Järjestelmän tavoitteiden on oltava läpinäkyviä, ja niiden tulisi tähdätä konkreettisiin hyötyihin samalla kun ne kunnioittavat perusoikeuksia ja ottavat huomioon ympäristön kestävyuden. On olennaista osallistaa monipuolisesti eri sidosryhmiä ja varmistaa, että tekoälyratkaisun vaikutuspiirissä olevien näkökulmat tulevat kuulluiksi ja huomioiduiksi. (De Silva & Alahakoon, 2022, s. 5; Morley ym., 2020, s. 2151)

Haitan välttämisen periaatteen näkökulmasta on varmistettava, ettei tekoälyjärjestelmä aiheuta vahinkoa yksilöille tai yhteisöille. Tämä edellyttää ennakoivaa lähestymistapaa mahdollisten haittavaikutusten tunnistamiseen ja lieventämiseen sekä tarkkaa huolenpitoa järjestelmän vaikutuspiirissä olevien ihmisten fyysisen ja henkisen hyvinvoinnin suojelemiseksi. (De Silva & Alahakoon, 2022, s. 5; Morley ym., 2020, s. 2151) Yleiset turvallisuuteen liittyvät näkökohdat on

otettava huomioon ennakoivasti ja sisällytettävä järjestelmän suunnitteluun. (UNESCO, 2023b, s. 18)

Oikeudenmukaisuuden näkökulmasta on tärkeää tarkastella tekoälyjärjestelmän laajempia vaikutuksia instituutioihin, demokratiaan ja yhteiskunnallisiin rakenteisiin. On olennaista pohtia, voiko hanke tahtomattaankin synnyttää vinoumia, suosia joitakin ryhmiä muiden kustannuksella tai kaventaa yksilöiden itsemääräämisoikeutta. Jokainen näistä näkökohdista vaatii huolellista harkintaa, jotta tekoälyjärjestelmä edistäisi oikeudenmukaista ja tasa-arvoista yhteiskuntaa. (Morley ym., 2020, s. 2151)

Sidosryhmät ongelman määrittelyn ja liiketoiminnan ymmärtämisen vaiheessa

Tässä vaiheessa on olennaista osallistaa vähintään loppukäyttäjät, jotka ovat tekoälyratkaisun ensisijaisia käyttäjiä. Lisäksi tekoälyn kehittäjät tulisi ottaa mukaan jo varhaisessa vaiheessa, jotta he saavat syvällisen ymmärryksen ratkaistavasta ongelmasta. Myös eettisten kysymysten asiantuntijat tai yhteiskuntatieteilijät tulisi osallistaa arvioimaan ratkaisujen mahdollisia vaikutuksia ihmisiin ja yhteiskuntaan. Alakohtaiset asiantuntijat ovat niin ikään keskeisessä roolissa, sillä heidän osaamisensa syventää ymmärrystä toimialasta, johon tekoälyä ollaan soveltamassa. (De Silva & Alahakoon, 2022)

11. Tekoälyn kehittämisen elinkaari: suunnitteluvaihe

Suunnitteluvaiheessa ongelman määrittelyn vaiheessa tehty perustyö kehittyy kattavaksi suunnitelmaksi tekoälyjärjestelmän käyttöönottoa varten. Tässä vaiheessa määritellään projektin tavoitteet ja toivotut lopputulokset, laaditaan projektisuunnitelma aikatauluineen ja budjettineen sekä ymmärretään sidosryhmien tarpeet. (De Silva & Alahakoon, 2022, s. 4)

Keskeisiin toimiin kuuluu vaatimusten määrittelydokumentin laatiminen, jossa kuvataan tekoälyjärjestelmän toiminnot sekä projektin onnistumisen mittarit. Lisäksi tarkennetaan tiedonhankinnan strategioita, kuten esimerkiksi päätöksiä esikoulutettujen mallien käytöstä. Eettiset näkökohdat ovat läsnä koko suunnitteluvaiheen ajan, ja ne on huomioitava järjestelmän jokaisessa osa-alueessa. (De Silva & Alahakoon, 2022, s. 3)

Eettiset periaatteet suunnitteluvaiheessa

Tekoälyratkaisun suunnitteluvaiheessa on tärkeää käyttää ennakoivia "entä jos" -strategioita mahdollisten haasteiden ennustamiseen ja lieventämiseen koko tekoälyratkaisun elinkaaren ajan. Esittämällä kriittisiä kysymyksiä, kuten "Entä jos käytetty data on puolueellista?" ja tutkimalla tarvittavia mekanismeja epäoikeudenmukaisen toiminnan havaitsemiseksi ja korjaamiseksi, kehittäjät

voivat ennaltaehkäistä näitä ongelmia. Kehittäjien tukemiseksi on välttämätöntä määritellä konkreettiset toimenpiteet, joiden avulla voidaan ehkäistä tekoälyjärjestelmän vinoumia, jotka liittyvät esimerkiksi rotuun, ihonväriin, syntyperään, sukupuoleen, ikään, kieleen, uskontoon, poliittisiin mielipiteisiin, kansalliseen tai etniseen taustaan, sosiaaliseen asemaan, taloudelliseen tilanteeseen, syntymäolosuhteisiin tai vammaisuuteen. Lisäksi on tärkeää luoda toimivat viestintäkanavat, raportointikäytännöt ja mekanismit, jotka mahdollistavat nopean reagoinnin havaittuihin vinoumiin ja tarvittaviin korjauksiin. (Morley ym., 2020, ss. 2151, 2155; UNESCO, 2023b, ss. 13, 17)

Näiden varotoimien lisäksi suunnitteluvaiheessa tulisi selkeästi määritellä, miten prosessin läpinäkyvyyttä voidaan lisätä ja siten parantaa järjestelmän selitettävyyttä. Tavoitteena on, että tekoälyjärjestelmän toiminta olisi ymmärrettävää ja sen toiminnasta voitaisiin tarvittaessa myös vastata. (Morley ym., 2020, ss. 2151, 2155; UNESCO, 2023b, ss. 13, 17)

Prem (2023, s. 703) korostaa lisäksi sidosryhmien osallistamisen merkitystä tässä vaiheessa sekä ihmiskeskeisten valvontamekanismien keskeistä roolia. Näin varmistetaan, että tekoälyn kehittämiseen osallistuu monipuolisesti erilaisia näkökulmia, mikä edistää oikeudenmukaisempaa ja tietoisempaa suunnitteluprosessia.

Sidosryhmät suunnitteluvaiheessa

Tässä vaiheessa on tärkeää osallistaa ohjelmistoarkkitehteja, sillä heillä on keskeinen rooli tekoälyratkaisun teknisen perustan luomisessa. He huomioivat sekä nykyiset vaatimukset että tulevan skaalautuvuuden tarpeet (Peter, 2024). Käyttökokemussuunnittelijat (UX-suunnittelijat) ovat olennaisia käyttöliittymien suunnittelussa siten, että ne edistävät käyttäjien hyväksyntää ja tyytyväisyyttä sekä varmistavat sujuvan vuorovaikutuksen tekoälyjärjestelmän kanssa (White, 2023). Tekoälyn asiantuntijoiden tulee tehdä tiivistä yhteistyötä UX-suunnittelijoiden kanssa, jotta tekoälykomponentit voidaan integroida luontevasti käyttöliittymään. Eettisten kysymysten asiantuntijat käsittelevät tekoälyn kehitykseen liittyviä eettisiä huolia. Myös juridiset ja eettiset toimijat ovat tarpeen, jotta voidaan vähentää tekoälyprojekteihin liittyviä oikeudellisia riskejä.

De Silvan ja Alahakoonin (2022) mukaan on suositeltavaa hyödyntää osallistavaa suunnittelua, jossa mukana ovat ihmiset ja yhteisöt, joihin tekoälymalli ja sen päätökset vaikuttavat. Tämä lähestymistapa tukee läpinäkyvyyttä, osallisuutta ja yhteiskunnallisten arvojen huomioimista koko projektin elinkaaren ajan.

12. Tekoälyn kehittämisen elinkaari: tiedonkeruu, ymmärrys ja valmistelu

Tavoitteen asettamisen ja sen saavuttamiseksi laaditun strategian jälkeen seuraava vaihe on datan hankinta ja valmistelu. Tämä kattaa useita vaiheita:

- datan kerääminen,
- sen määrittely ja luokittelu,
- tutkiva analyysi,
- datan merkityksellisyyden arviointi,
- olennaisen tiedon valinta sekä sen puhdistaminen,
- uudelleenrakentaminen, yhdistäminen ja muotoilu niin, että se vastaa projektin tavoitteisiin liittyviä määriteltyjä luokkia ja käsitteitä

Datan puuttuvien arvojen käsittely ja korjaaminen on olennainen osa tätä prosessia. On tärkeää ymmärtää, millaisiin oletuksiin olemassa oleva data perustuu. Datan käyttö päätöksenteossa tulee olla läpinäkyvää, ja datan laadun varmistaminen on ensiarvoisen tärkeää – erityisesti silloin, kun data on puutteellista tai tulkinvaraista. (Rochel & Évéquoz, 2021, ss. 614–617)

Vaikka tämä vaihe on ajallisesti vaativa, se on myös perustavanlaatuinen. Tekoälyratkaisun luotettavuus ja suorituskkyky ovat suoraan sidoksissa käytetyn datan laatuun ja tarkkuuteen. (Saltz, 2023)

Eettiset periaatteet datan keruun, ymmärtämisen ja valmistelun vaiheessa

Datan keruuseen liittyy luontaisia riskejä, erityisesti haitan välttämisen (non-maleficence) periaatteen näkökulmasta. Tietosuoja ja luottamuksellisuuden varmistaminen on tässä vaiheessa ensiarvoisen tärkeää. Tekoälyjärjestelmien on turvattava tietosuoja johdonmukaisesti koko ratkaisun elinkaaren ajan, alusta loppuun. Tietosuojaa säädellään usein laeilla, kuten globaalilla tietosuojalailla (Global Data Protection Law). (De Silva & Alahakoon, 2022, s. 5)

Vinoumat datassa voivat johtaa tiettyjen väestöryhmien ylliedustukseen, joten datan laadun ja eheyden huolellinen tarkastelu on välttämätöntä. Lisäksi tietoineistot altistuvat kyberuhille, minkä vuoksi tarvitaan vahvoja suojausmenetelmiä mahdollisten hyökkäysten estämiseksi. (UNESCO, 2023a, ss. 9, 18)

Myös datankeruuprosessin läpinäkyvyys on tärkeää – on kerrottava, mitä tietoja kerätään, mihin tarkoitukseen ja miten niitä tullaan käyttämään tekoälyjärjestelmässä. (De Silva & Alahakoon, 2022, s. 6)

Sidosryhmät datan keruun, ymmärtämisen ja valmistelun vaiheessa

Datatieteilijöiden osallistaminen tässä vaiheessa on ratkaisevan tärkeää, sillä heillä on keskeinen rooli datassa esiintyvien kuvioiden ja trendien tunnistamisessa ja niiden muuttamisessa käyttökelpoiseksi tiedoksi (Miller, 2022). Samoin tekoälyinsinöörit ovat korvaamattomia datan eheyden arvioinnissa ja varmentamisessa (Rochel & Évéquoz, 2021), kun taas data-analytikot tukevat

projektia ymmärtämällä datan ominaisuuksia ja antamalla suosituksia tehokkaista datanvalmistelustrategioista. Lisäksi datainsinöörin panos on olennainen datan valmistelussa ja ylläpidossa sen laadun varmistamiseksi (Miller, 2022).

Tietosuojaavastaavan (DPO) mukanaolo on keskeistä, jotta voidaan varmistaa henkilötietojen käsittelyyn liittyvien oikeudellisten ja eettisten standardien noudattaminen (European Data Protection Supervisor, 2024).

Eettisten asiantuntijoiden rooli on merkittävä, sillä he tuovat eettiset näkökohdat mukaan koko datan käsittelyyn ja valintaan. Näiden monipuolisten sidosryhmien osallistaminen varmistaa, että tekoälyprojekti perustuu eettisesti hankittuun ja huolellisesti valmisteltuun dataan, mikä luo luottamusta ja varmuutta tekoälyratkaisun pohjaksi.

13. Tekoälyn kehittämisen elinkaari; mallin kehittäminen ja koulutus

Mallin kehittämis- ja koulutusvaiheessa tekoälymalli rakennetaan erityisesti ennalta määritellyn haasteen ratkaisemiseksi. Tässä vaiheessa valitaan, asetetaan ja otetaan käyttöön algoritmisia työkaluja tarkasti harkiten. On tärkeää dokumentoida nämä valinnat huolellisesti ja arvioida niiden vaikutuksia, erityisesti mahdollisia haittapuolia tai välttämättömiä kompromisseja. Tälle vaiheelle on ominaista iteratiivinen luonne – mallia kehitetään ja hiotaan useissa kierroksissa sen tarkkuuden ja tehokkuuden parantamiseksi.

Algoritmisten työkalujen valinta vaatii usein tasapainoilua ristiriitaisten tavoitteiden välillä. Hyvä esimerkki tästä on roskapostisuodatuksen haaste: on pyrittävä vähentämään roskapostin määrää käyttäjien postilaatikossa, mutta samalla vältettävä virheellistä tavallisten viestien luokittelua roskapostiksi. Tämän tasapainon saavuttaminen on olennaista mallin tehokkuuden parantamiseksi ilman, että sen luotettavuus tai käyttäjien luottamus heikkenee. (Rochel & Évéquoz, 2021, ss. 617–618; Saltz, 2023)

Eettiset periaatteet mallin kehittämisen ja koulutuksen vaiheessa

Vaikka tekniset yksityiskohdat usein korostuvat, tämä vaihe tarjoaa hyvän mahdollisuuden varmistaa valitun mallin selitettävyys ja tulkittavuus sekä arvioida mahdollisia vinoumia (Prem, 2023, s. 703). Kehitysprosessin on oltava selkeä ja ymmärrettävä kaikille osapuolille, mikä edellyttää kattavaa dokumentaatiota datalähteistä, mallivalinnoista ja päätöksenteon taustalla olevasta logiikasta (Morley ym., 2020, s. 2151).

Tarvittaessa on tunnistettava ja arvioitava avoimesti myös kompromisseja, joita erityisesti tehokkuuden tavoittelu usein edellyttää. Tekoälyinsinöörin on

pystyttävä perustelemaan valintansa ja samalla otettava huomioon sekä pyrittävä lieventämään mahdollisia kielteisiä vaikutuksia niihin, joita ratkaisu koskettaa. (Morley ym., 2020, s. 2151)

Sidosryhmät mallin kehittämisen ja koulutuksen vaiheessa

Tekoäly- tai koneoppimisen tutkijoilla on keskeinen rooli ongelmanmäärittelyjen muuttamisessa tekoälymallien ensimmäisiksi prototyypeiksi (De Silva & Alahakoon, 2022), mikä tekee heidän osallistumisestaan erityisen tärkeää mallin kehitys- ja koulutusvaiheen alkaessa. Tässä vaiheessa tekoälyinsinöörit ovat korvaamattomia, sillä he osaavat tunnistaa ja hyödyntää ne algoritmiset työkalut, jotka parhaiten tukevat projektin tavoitteita (Rochel & Évéquoz, 2021). Lisäksi datatieteilijät ovat tämän vaiheen ytimessä ohjaten koneoppimismallien kehitystä ja jatkuvaa hienosäätöä kohti suurempaa tarkkuutta ja tehokkuutta.

Eettisten asiantuntijoiden osallistaminen on äärimmäisen tärkeää, jotta kehitysprosessiin saadaan sisällytettyä eettiset näkökulmat ja vältetään mahdolliset kielteiset vaikutukset. Myös liiketoiminnan edustajien näkemykset ovat arvokkaita, sillä ne varmistavat, että kehitettävät mallit ovat linjassa liiketoimintastrategioiden kanssa ja tuottavat odotettuja tuloksia – näin kurotaan umpeen teknologisen innovaation ja käytännön liiketoimintatarpeiden välistä kuilua.

14. Tekoälyn kehittämisen elinkaaren mallin arviointi ja testaus

Kun tekoälymalli on kehitetty ja koulutettu, sen tehokkuus on testattava ja arvioitava ennalta määriteltujen tavoitteiden ja onnistumiskriteerien perusteella. Arvioinnin tarkoituksena on paitsi mitata mallin suorituskykyä, myös tarkastella siihen liittyviä oletuksia ja käytettyjen data-aineistojen valintaperusteita. Mikäli malli ei täytä odotuksia, voi olla tarpeen tehdä säätöjä esimerkiksi mallin parametreihin, rakenteeseen tai sen hyödyntämiin data-aineistoihin.

Mahdollisista puutteista on tärkeää viestiä avoimesti projektipäällikölle ja kaikille asiaankuuluville sidosryhmille, sillä niillä voi olla merkittäviä vaikutuksia osallisiin. Arvioinnin tulosten perusteella tiimin on päätettävä jatkotoimista – tehdäänkö lisäiterointeja mallin parantamiseksi vai edetäänkö käyttöönottovaiheeseen – varmistaen, että kaikki toimenpiteet ovat selkeästi linjassa ja yhdessä sovittuja. (Hotz, 2018; Rochel & Évéquoz, 2021, s. 619)

Eettiset periaatteet mallin arvioinnin ja testauksen vaiheessa

Tässä vaiheessa mallin testaaminen sen eettisten ohjeiden noudattamisen kannalta on ensiarvoisen tärkeää, jotta varmistetaan sen yhdenmukaisuus alussa määriteltujen keskeisten periaatteiden kanssa. Auditointien ja auditointimittarien

kehittäminen ja hienosäätö ovat olennaisia tehtäviä. Tärkeää on sisällyttää keskeisiä eettisiä periaatteita, kuten oikeudenmukaisuus, hyvän tekeminen ja haitan välttäminen. Mallin kykyä kestää mahdollisia tietoturva-uhkia tulee arvioida, ja odottamattomiin haasteisiin on kehitettävä selkeä toimintastrategia. Haitallisten vaikutusten varalta on oltava olemassa selkeät mekanismit korjaaviin toimenpiteisiin. Arvioinnin tulee olla perusteellinen ja kattaa mm. yksityisyyden suoja, yhteiskunnalliset vaikutukset ja mahdolliset vinoumat. Negatiiviset vaikutukset on pyrittävä minimoimaan ja niistä on raportoitava avoimesti. (De Silva & Alahakoon, 2022, s. 8; Morley ym., 2020, s. 2151; Prem, 2023, s. 703)

Läpinäkyvyys on olennainen paitsi mallin teknisen toiminnan osalta myös suhteessa projektin sisäisiin, ihmisten tekemiin päätöksiin, jotta selitettävyys voidaan varmistaa. (De Silva & Alahakoon, 2022, s. 8; Morley ym., 2020, s. 2151) Lisäksi on tärkeää tunnistaa, että insinööreihin saattaa kohdistua painetta nopeuttaa kehitysprosessia käyttöönottoa varten, mikä voi lisätä riskejä (Rochel & Évéquoz, 2021, s. 618).

Sidosryhmät mallin arvioinnin ja testauksen vaiheessa

Tässä vaiheessa tekoälyinsinööreillä on keskeinen rooli mallinnuksesta saatujen tulosten tulkinnessa, ja he tarjoavat olennaisia näkemyksiä mallin toimivuudesta (Rochel & Évéquoz, 2021).

Laadunvarmistustiimit (Quality Assurance, QA) ovat ratkaisevassa asemassa varmistamassa, että mallit täyttävät laatuvaatimukset eikä niissä ilmene virheitä tai ei-toivottuja tuloksia. Eettisten asiantuntijoiden osallistuminen on tärkeää, jotta tekoälymalleja otetaan käyttöön eettisesti ja vastuullisesti. Käyttäjien edustajat tai -puolestapuhujat ovat merkittävässä roolissa sen varmistamisessa, että tekoälymallit vastaavat käyttäjien odotuksia ja tarpeita, mikä lisää käyttäjätyytyväisyyttä ja hyväksyttävyyttä.

Lisäksi lakiasiain- ja vaatimustenmukaisuustiimit ovat välttämättömiä oikeudellisten riskien hallinnassa ja sen varmistamisessa, että tekoälymallit noudattavat kaikkia sovellettavia lakeja ja säädöksiä. (Moore, 2023)

15. Tekoälyn kehittämisen elinkaari: käyttöönotto

Kun tekoälyratkaisu on onnistuneesti kehitetty, se otetaan käyttöön tuotantoympäristössä ratkaisemaan todellisia, käytännön ongelmia. Toteutuksen laajuus ja menetelmät riippuvat ratkaisun erityisistä tavoitteista sekä siitä, millaisiin järjestelmiin tai sovelluksiin se on suunniteltu. Yleensä käyttöönotto aloitetaan pilottivaiheella, johon osallistuu valikoitu joukko asiantuntijoita ja varhaisia käyttäjiä, jotka antavat ensipalautetta (De Silva & Alahakoon, 2022, s. 8).

Tekoälyratkaisu voidaan integroida olemassa oleviin järjestelmiin, käyttää uuden sovelluksen tai palvelun perustana tai sen tuottamaa tietoa voidaan jakaa offline-menetelmin, kuten esihenkilöraporttien muodossa. On tärkeää perustaa palautemekanismi, jolla seurataan ratkaisun suorituskykyä ja vaikutuksia – näin mahdollistetaan jatkuva kehittäminen todelliseen käyttöön perustuen (Saltz, 2023).

Mahdollisten riskien hallitsemiseksi otetaan käyttöön kattavat riskienhallintastrategiat. Operatiivisia ja seurantatoimenpiteitä toteutetaan sen varmistamiseksi, että tekoälyratkaisu toimii tehokkaasti myös käyttöönoton jälkeen. Projektin lopuksi suoritetaan perusteellinen arviointi, josta laaditaan yksityiskohtainen loppuraportti, joka kokoaa yhteen projektin tulokset ja siitä saadut opit (De Silva & Alahakoon, 2022, ss. 8–9).

Eettiset periaatteet käyttöönoton vaiheessa

Morley (2020) korostaa, että eettiset näkökohdat saavat usein liian vähän huomiota suunnitteluprosessin käyttöönoton vaiheessa. Monimutkaisten inhimillisten käyttäytymismallien yksinkertaistaminen työkaluiksi, jotka ovat sekä ymmärrettäviä että läpinäkyviä, on merkittävä haaste. Kuitenkin tällaisten työkalujen luominen on olennaista luottamuksen lisäämiseksi tekoälyjärjestelmiin. Tästä syystä selitettävyyden nousee keskeiseksi eettiseksi periaatteeksi tässä vaiheessa.

Lisäksi autonomian turvaaminen on tärkeää – loppukäyttäjiä on informoitava aina, kun he ovat vuorovaikutuksessa tekoälyohjatun prosessin kanssa. Tämä läpinäkyvyys on keskeistä, jotta voidaan välttää liiallinen riippuvuus tekoälyjärjestelmistä. UNESCO (2023b, s. 36) painottaa tehokkaiden mekanismien tarvetta neljällä keskeisellä osa-alueella: järjestelmätietoisuus, vahvat auditointimenettelyt, algoritmisen päättelyn selkeys ja vaikutusten arviointikeinot – kuten valitus- tai palautekanavat, jotka sisältävät myös suojatoimet ilmoittajille.

Ihmisen valvonnan kriittistä roolia ei voida liikaa korostaa – on välttämätöntä varmistaa, että tekoälyjärjestelmiin liittyy mahdollisuus inhimilliseen puuttumiseen ja ohjaukseen, jotta niiden käyttöönotto olisi eettisesti kestävä (Morley ym., 2020, s. 2151).

Sidosryhmät käyttöönoton vaiheessa

AI or machine learning engineers are pivotal in transitioning the prototype AI
Tekoäly- tai koneoppimisinsinööreillä on keskeinen rooli prototyypivaiheessa olevan tekoälymallin muuttamisessa täysin käyttöön otettavaksi palveluksi tai ratkaisuksi, joka on sidosryhmien ja loppukäyttäjien saatavilla (De Silva & Alahakoon, 2022).

Kyberturvallisuusasiantuntijat suorittavat riskiarvioinnin tekoälyjärjestelmän suojaamiseksi (Junklewitz ym., 2023, ss. 9–12). Loppukäyttäjillä on merkittävä rooli,

sillä he voivat tuoda esiin ongelmia tai puutteita, joita ei ilmennyt hallitussa testausympäristössä.

Lisäksi käyttäjäedustajat tai -puolestapuhujat ovat tärkeitä palautteen keräämisessä tekoälyratkaisun käytettävyydestä ja suorituskyvystä. Heidän roolinsa varmistaa, että ratkaisu vastaa käyttäjien odotuksia ja tarpeita, ja auttaa tunnistamaan kehityskohteita.

16. Tekoälyn kehittämisen elinkaaren toiminta ja seuranta

Jatkuva ylläpito, päivitykset ja arvioinnit ovat olennaisia toimivan tekoälyjärjestelmän pitkäikäisyyden ja tehokkuuden kannalta. Järjestelmän suorituskyyä tarkastellaan säännöllisesti sen varmistamiseksi, että se täyttää odotetut standardit. Käyttäjävurorvaikutusten jatkuva seuranta tarjoaa arvokasta tietoa järjestelmän toimivuudesta käytännön tilanteissa ja mahdollisista kehityskohteista. Mallin ajantasaisuuden ja tarkkuuden säilyttämiseksi säännölliset päivitykset uudella datalla ovat keskeisiä. Käyttäjäpalaute on tärkeässä roolissa tekoälymallin jatkuvassa hienosäätämisessä, ja se korostaa iteratiivisten parannusten merkitystä osana ylläpitoprosessia – jatkuva kehitys nähdään edistymisenä, ei epäonnistumisena.

Lisäksi järjestelmän tuottaman sijoitetun pääoman tuoton (ROI) ja muiden keskeisten mittareiden seuraaminen on tärkeää, jotta voidaan arvioida, onko järjestelmä saavuttanut sille asetetut tavoitteet. (De Silva & Alahakoon, 2022, s. 9; Saltz, 2023)

Eettiset periaatteet käyttö- ja seurantavaiheessa

Tässä vaiheessa suorituskyyyn seuranta ulottuu teknisten mittareiden lisäksi myös eettisten standardien noudattamisen arviointiin. Kaikkien eettisten periaatteiden säännöllinen uudelleenarviointi on olennaista, erityisesti kun otetaan huomioon mahdolliset muutokset data-aineistoissa tai mallin rakenteessa. Sidosryhmien osallistaminen on tärkeää myös käyttöönoton jälkeen – heidän näkemyksensä ovat keskeisiä jatkuvien parannusten toteuttamisessa.

Palautetta tulisi aktiivisesti kerätä käyttäjiltä, organisaatioilta ja laajemmalla yhteiskunnalla. On tärkeää, että palautteen keräämiseen ja hyödyntämiseen on olemassa rakenteelliset mekanismit, jotta sitä voidaan johdonmukaisesti käyttää tekoälyratkaisun jatkuvaan kehittämiseen. (High-Level Expert Group on AI (AI HLEG), 2019, s. 19; UNESCO, 2023b, s. 15)

Sidosryhmät käyttö- ja seurantavaiheessa

Käyttö- ja seurantavaiheessa DevOps- ja IT-tiimit ovat keskeisessä roolissa järjestelmän suorituskyyyn seurannassa ja jatkuvan käytettävyyden varmistamisessa (Immersant Data Solutions, 2023). Riippumattomat arvioijat,

kuten turvallisuussertifioijat, voivat arvioida kehitetyn tekoälyratkaisun tietoturvaa (Miller, 2022).

Lisäksi asiantuntijapaneelit, ohjausryhmät tai sääntelyviranomaiset suorittavat projektin tarkasteluja, jotka kattavat sekä tekniset että eettiset näkökulmat (De Silva & Alahakoon, 2022).

17. Johtopäätös

Tässä osaamiskokonaisuudessa on perehdytty eettisiin periaatteisiin, viitekehyksiin ja ohjeistuksiin, jotka ovat keskeisiä tekoälyratkaisujen kehittämisen ohjaamisessa. Siinä on tarkasteltu tekoälyn kehitysprosessin elinkaarta aina alkuvaiheen ongelmanmäärittelystä tekoälyratkaisun käyttöön ja seurantaan asti, painottaen eettisten periaatteiden integrointia ja sidosryhmien osallistamista jokaisessa vaiheessa.

Eettinen tekoälyn kehittäminen korostaa sitä, kuinka tärkeää on sisällyttää eettiset näkökohdat koko tekoälyn elinkaaren ajan. Tämä lähestymistapa varmistaa, että tekoälyteknologiat eivät ainoastaan tuota myönteisiä vaikutuksia yhteiskunnalle, vaan myös käsittelevät tehokkaasti mahdollisia riskejä, kuten algoritmista vinoumaa ja yksityisyydensuojaan liittyviä huolia. Lisäksi korostettiin laaja-alaisen sidosryhmäosallistamisen merkitystä, mikä osoittaa, että onnistunut tekoälyn kehittäminen vaatii teknisen asiantuntemuksen lisäksi monipuolisia näkökulmia. Tämä osallistava lähestymistapa rikastuttaa kehitysprosessia ja varmistaa, että tekoälyratkaisut ovat yhteiskunnallisten arvojen mukaisia ja saavuttavat laajemman hyväksynnän.

Lisäksi korostettiin jatkuvan seurannan ja tekoälyjärjestelmien iteratiivisen kehittämisen tarvetta käyttöönoton jälkeen, jotta järjestelmien eettinen kestävyys säilyy ja niiden pitkäaikainen toimivuus voidaan varmistaa.

Omaksumalla ennakoivan asenteen innovaatioita kohtaan kehittäjät voivat tunnistaa ja ratkaista eettisiä haasteita niiden ilmaantuessa. Näin he voivat edetä tekoälyn kehittämisen monimutkaisessa kentässä sitoutuen eettisiin periaatteisiin ja vastuulliseen innovointiin.

18. Viittaukset

Ayling, J., & Chapman, A. (2022). Putting AI ethics to work: Are the tools fit for purpose? *AI and Ethics*, 2(3), 405–429. <https://doi.org/10.1007/s43681-021-00084-x>

Beijing Academy of Artificial Intelligence—Beijing AI Principles. (2019). *Datenschutz Und Datensicherheit*, 10.

Beijing Artificial Intelligence Principles. (2022, January 10). International Research Center for AI Ethics and Governance. <https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/>

- De Silva, D., & Alahakoon, D. (2022). An artificial intelligence life cycle: From conception to production. *Patterns*, 3(6), 100489.
<https://doi.org/10.1016/j.patter.2022.100489>
- European Commission. (2023, September 12). Commission welcomes political agreement on AI Act [Text]. European Commission - European Commission.
https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473
- European Data Protection Supervisor. (2024, April 9). Data Protection Officer (DPO) | European Data Protection Supervisor.
https://www.edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_en
- European Parliament. (2023, June 8). EU AI Act: First regulation on artificial intelligence. Topics | European Parliament.
<https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
- Floridi, L., & Cows, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*.
<https://doi.org/10.1162/99608f92.8cd550d1>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Future of Life Institute. (2024). The AI Act Explorer | EU Artificial Intelligence Act.
<https://artificialintelligenceact.eu/ai-act-explorer/>
- High-Level Expert Group on AI (AI HLEG). (2019). Ethical Guidelines of Trustworthy AI. High-Level Expert Group on Artificial Intelligence. European Commission. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>
- High-Level Expert Group on AI (AI HLEG). (2020). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. European Commission.
- Hillard, A., & Gulley, A. (2023, December 9). What is the EU AI Act?
<https://www.holistica.com/blog/eu-ai-act>
- Hotz, N. (2018, September 10). What is CRISP DM? Data Science Process Alliance.
<https://www.datascience-pm.com/crisp-dm-2/>
- Immersant Data Solutions. (2023, June 9). DevOps: Bridging the Gap between Development and Operations | LinkedIn.
<https://www.linkedin.com/pulse/devops-bridging-gap-between-development-operations/>
- Interaction Design Foundation - IxDF. (2021, October 27). What is Co-Creation? Interaction Design Foundation—IxDF. The Interaction Design Foundation.
<https://www.interaction-design.org/literature/topics/co-creation>
- Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389–399.
<https://doi.org/10.1038/s42256-019-0088-2>
- Junklewitz, H., Hamon, R., André, A., Evas, T., Soler Garrido, J., & Sanchez Martin, J. I. (2023). Cybersecurity of artificial intelligence in the AI Act: Guiding

- principles to address the cybersecurity requirement for high risk AI systems. Publications Office of the European Union.
<https://data.europa.eu/doi/10.2760/271009>
- Leijnen, S., Aldewereld, H., van Belkom, R., Bijvank, R., & Ossewaarde, R. (2020). An agile framework for trustworthy AI. 75–78.
- Miller, G. J. (2022). Stakeholder roles in artificial intelligence projects. *Project Leadership and Society*, 3, 100068. <https://doi.org/10.1016/j.plas.2022.100068>
- Moore, S. (2023, December 18). What is...a Corporate Legal and Compliance team? | LinkedIn. <https://www.linkedin.com/pulse/what-is-a-corporate-legal-compliance-team-steven-moore-iczqf/>
- Morley, J., Floridi, L., Kinsey, L., & Elhalal, A. (2020). From What to How: 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>
- OECD. (2019). AI-Principles Overview. OECD AI Principles Overview. <https://oecd.ai/en/principles>
- Peter, A. (2024, October 2). AI Software Architecture: Navigating the Essentials of Business App Development | LinkedIn. <https://www.linkedin.com/pulse/ai-software-architecture-navigating-essentials-business-alex-peter-uugnc/>
- Prem, E. (2023). From ethical AI frameworks to tools: A review of approaches. *AI and Ethics*, 3(3), 699–716. <https://doi.org/10.1007/s43681-023-00258-9>
- Rochel, J., & Évéquoz, F. (2021). Getting into the engine room: A blueprint to investigate the shadowy steps of AI ethics. *AI & SOCIETY*, 36(2), 609–622. <https://doi.org/10.1007/s00146-020-01069-w>
- Rodrigues, R. (2020). Legal and human rights issues of AI: Gaps, challenges and vulnerabilities. *Journal of Responsible Technology*, 4, 100005. <https://doi.org/10.1016/j.jrt.2020.100005>
- Saltz, J. (2023, June 1). What is the AI Life Cycle? Data Science Process Alliance. <https://www.datascience-pm.com/ai-lifecycle/>
- Sanin, A. (2020, September 4). Co-Creation How-To's. Design Globant. <https://medium.com/design-globant/co-creation-how-tos-e25696f56d6f>
- UNESCO. (2023a). Ethical impact assessment. A tool of the Recommendation on the Ethics of Artificial Intelligence. UNESCO. <https://doi.org/10.54678/YTSA7796>
- UNESCO. (2023b). Readiness assessment methodology. A tool of the Recommendation on the Ethics of Artificial Intelligence. UNESCO. <https://doi.org/10.54678/YHAA4429>
- UNESCO. (2023c, April 21). Artificial Intelligence: Examples of ethical dilemmas | UNESCO. <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics/cases>
- White, C. (2023, November 30). What Does a UX Designer Actually Do? [2024 Guide]. <https://careerfoundry.com/en/blog/ux-design/what-does-a-ux-designer-actually-do/>
- Zhou, J., & Chen, F. (2022). AI ethics: From principles to practice. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01602-z>
- Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., & Savage, S. (2020). A Survey on Ethical Principles of AI and Implementations. 2020 IEEE Symposium Series

on Computational Intelligence (SSCI), 3010–3017.

<https://doi.org/10.1109/SSCI47803.2020.9308437>

Zwikael, O., & Meredith, J. R. (2018). Who's who in the project zoo? The ten core project roles. *International Journal of Operations & Production Management*, 38(2), 474–492. <https://doi.org/10.1108/IJOPM-05-2017-0274>



Universitat
de les Illes Balears

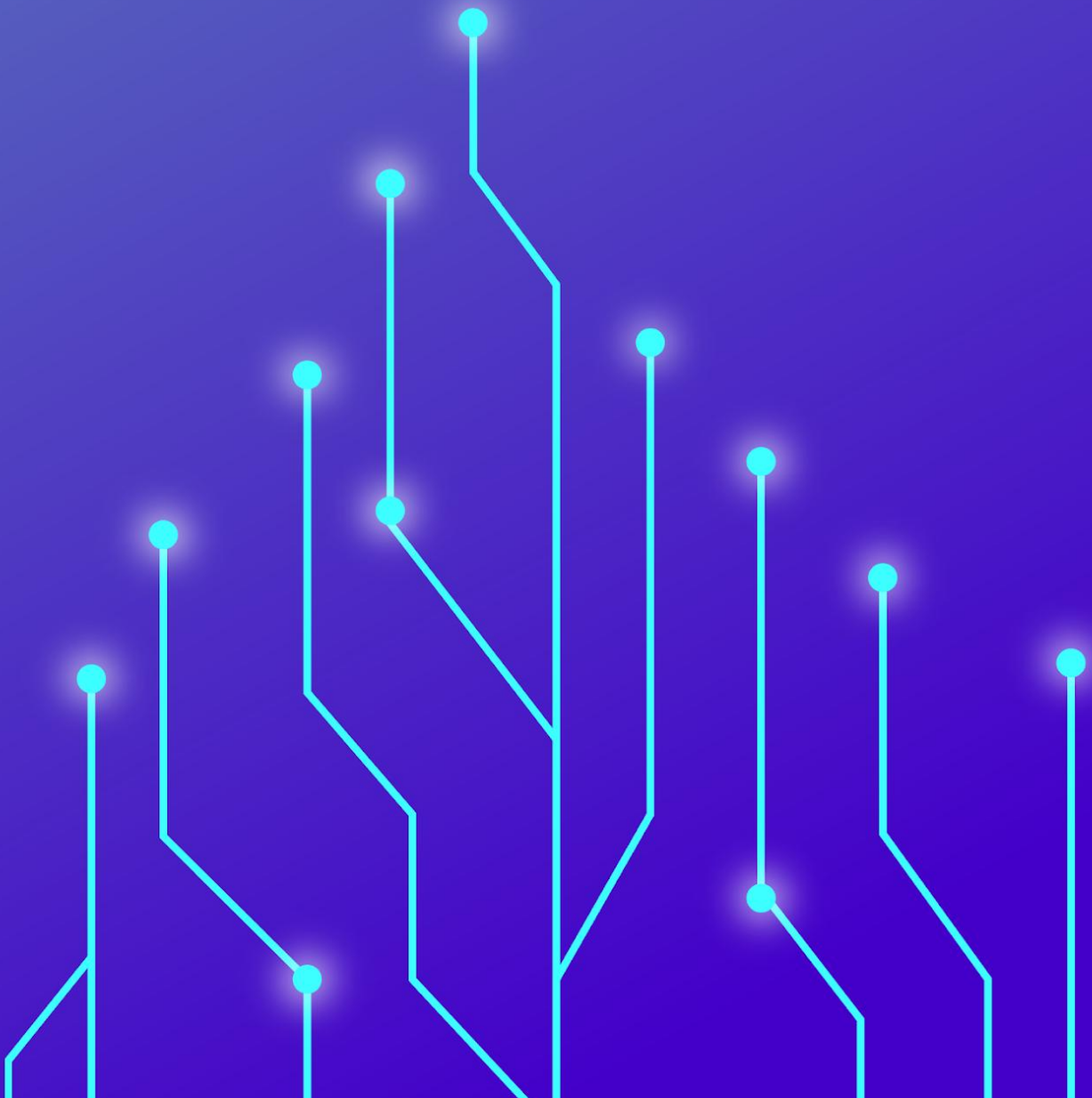
ISQe
ENGAGING PEOPLE



VAAKAN AMMATTIKORKEAKOULU
UNIVERSITY OF APPLIED SCIENCES

helixconnect
Consult. Finance. Grow.

Osaamiskokonaisuus 2 | Tekoälyn yksityisyys ja käytettävyys



Sisällys

1. Introduction	35
2. Sarah's dilemma: the privacy-convenience trade-off	37
3. Understanding Privacy in AI	40
3.1. Data Privacy: Protection of Personal Data from Unauthorized Access	41
3.2. Information Security: Measures to Protect Data Integrity and Confidentiality	42
3.3. Personal Autonomy: The Right to Control Information About Oneself	42
4. Data Privacy	43
4.1. Importance of Data Privacy in AI	43
4.2. Role of Regulations like GDPR	43
4.3. Key Aspects of Data Privacy	44
5. Understanding Convenience in AI	45
5.1. Real-life Examples of Convenience	46
6. The Interplay of Privacy and Convenience	49
7. Importance of Privacy and Convenience	51
8. Navigating Risks and Benefits	52
9. Case study: AI in surveillance	54
10. Conclusion	56

1. Johdanto

Digitaalisella aikakaudella tekoälyn nopea kehitys on muuttanut lukemattomia arjen osa-alueita ja uudistanut tapojamme olla vuorovaikutuksessa teknologian ja toistemme kanssa. Tämän murroksen ytimessä on kahden keskeisen tekijän – yksityisyyden ja käytettävyyden – välinen dynaaminen vuorovaikutus. Tämä osaamiskokonaisuus syventyy näiden elementtien monimutkaiseen suhteeseen ja tarkastelee niiden tuomia haasteita ja mahdollisuuksia älyteknologioiden aikakaudella.

Koska tekoäly sulautuu yhä tiiviimmin osaksi henkilökohtaista ja ammatillista elämäämme, on entistä tärkeämpää tarkastella kriittisesti tasapainoa käyttäjäkokemuksen parantamisen ja henkilökohtaisen yksityisyyden suojaamisen välillä. Tämä osaamiskokonaisuus tarjoaa osallistujille kattavan ymmärryksen siitä, miten sekä yksityisyys että käytettävyys sisältyvät tekoälysovelluksiin. Siinä korostetaan molempien mahdollisuuksia helpottaa arkea sekä niihin liittyviä eettisiä näkökulmia.

Osallistujat saavat tietoa yksityisyyden suojan peruseriaatteista, tekoälyn aiheuttamista eettisistä ristiriidoista sekä oikeudellisista viitekehyksistä, jotka ohjaavat teknologian tulevaisuuden käyttöönottoa.

Tarkastelemalla sekä todellisia sovelluksia että hypoteettisia tilanteita tämä kurssi pyrkii antamaan oppijoille valmiudet toimia vastuullisesti kehittyvässä tekoälyn kentässä. Tavoitteena on varmistaa, että teknologinen kehitys rikastuttaa ihmisten kokemuksia vaarantamatta yksilön oikeuksia.

Tämä matka tekoälyn, yksityisyyden ja käytettävyyden maailmaan haastaa osallistujat ajattelemaan kriittisesti teknologian roolia yhteiskunnassa sekä omaa vastuutaan tulevana tekoälyn kehittäjinä ja vaikuttajina.



Vinkkejä ja suosituksia opettajille

Kiinnostava kysymys tai yllättävä fakta, joka herättää oppijoiden mielenkiinnon ja auttaa selittämään nykyisen tekoälyn tilannetta korostaen yksityisyyden ja mukavuuden merkitystä.

Yllättävä fakta

Tuoreen tutkimuksen mukaan tekoälyjärjestelmät voivat tehdä päätöksiä työpaikkahakemuksistasi, luottotiedoistasi ja jopa terveydenhuollostasi – usein ilman suoraa tietoa tai suostumustasi. Tekoälyn integrointi herättää tärkeitä kysymyksiä: kuinka paljon voimme hallita omia tietojamme ja mitä se tarkoittaa yksityisyydellemme? Sukelletaan tekoälyn monimutkaiseen maailmaan, jossa käytettävyyys saatetaan saavuttaa yksityisyytemme kustannuksella.

Kiinnostava kysymys

Tiesitkö, että keskimääräinen ihminen on vuorovaikutuksessa tekoälyn kanssa yli 20 kertaa päivässä ilman että huomaa sitä? Tekoäly on integroitu saumattomasti elämäämme henkilökohtaisista ostosuosituksista päivittäisiä rutiinejamme hallitseviin älykkäisiin kodinkoneisiin. Mutta kuinka usein mietit, mitä henkilökohtaisia tietoja annat vastineeksi tästä kätevyydestä? Tutkitaanpa, mitä todella on pelissä.

Sisältö voidaan esittää myös kyselynä.

Kysely

"Mitä palveluita käytät, joiden tiedät keräävän tietojasi käytettävyyden parantamiseksi ja käyttökokemuksen mukauttamiseksi?"

Vaihtoehtoja

- Älypuhelimet
- Sosiaalisen median alustat
- Suoratoistopalvelut
- Verkkokauppasivustot
- Älykotilaitteet
- Kuntoseurantalaite ja terveyssovellukset
- Ääniohjaus
- Navigointi- ja matkasovellukset
- Pankki- ja rahoitussovellukset
- Sähköpostipalvelut

Jatkokysymys

"Kuinka hyvin hyväksyt näiden palvelujen keräävän tietojasi?"

Hyvin || Melko hyvin || Neutraalisti || Melko huonosti || Huonosti

Tämä lähestymistapa auttaa tunnistamaan tietoisuuden tietojen keräämisestä eri palveluissa ja kartoittaa tunteita yksityisyydestä ja henkilötietojen käytöstä, jolloin saadaan kattava kuva osallistujien asenteista tietosuojaa kohtaan.

2. Sarahin dilemma: yksityisyyden ja käytettävyyden välinen kompromissi



KUVAN LÄHDE | Kuvan on luonut Dall-E.

Kuvittele opiskelija Sarah, joka on juuri muuttanut ensimmäiseen asuntoonsa. Sarah haluaa sujuvoittaa kiireistä elämäänsä ja investoi useisiin älylaitteisiin: älykaiuttimeen, jolla hän voi hallita aikataulujaan ja soittaa musiikkia, älytermostaattiin, jolla hän voi säätää asunnon lämpötilaa, ja älykelloon, jolla hän voi seurata kuntoilurutiinejaan.

Vinkkejä ja suosituksia opettajille

Vinkki: esitä kysymyksiä kriittisissä kohdissa, jotta kuulijat voivat samaistua tilanteeseen.



KUVAN LÄHDE | Kuvan on luonut Dall-E.

Eräänä aamuna Sarahin älykaiutin ehdottaa uutta reittiä kohti hänen korkeakouluun, säästään aikaa ruuhkaisena hetkenä – täydellinen esimerkki käytettävyydestä. Myöhemmin hän saa puhelimeensa personoidun mainoksen kahvilasta uuden reittinsä varrella, houkutellen häntä lempilattellaan. Aluksi mielissään Sarah miettii:

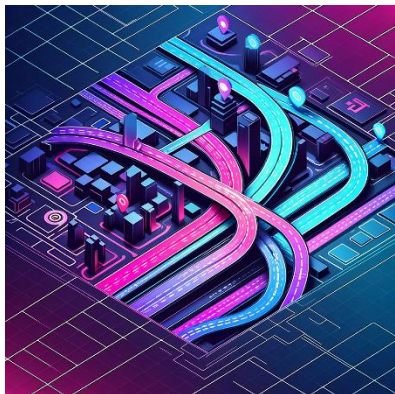
"Mistā se tiesi?"

Kysymys

"Luuletko, että Sarah otti huomioon yksityisyyden suojan ennen näiden laitteiden ostamista? Miksi tai miksi ei?"

Kysymys

- Mitä mieltä olet laitteista, jotka tekevät päätöksiä päivittäisten rutiiniesi perusteella? Millaisia yksityisyyden suojaan liittyviä kompromisseja tämä voi aiheuttaa?
- Mitä tietoja on voitu jakaa tämän mainoksen personointia varten? Oletko tyytyväinen tähän tietojen jakamisen tasoon personoitujen mainosten osalta?



KUVAN LÄHDE | Kuvan on luonut Dall-E.

Päivien kuluessa Sarah alkaa huomata yhä enemmän tällaisia tilanteita. Hänen älykellonsa ehdottaa liikuntasuunnitelmia hänen kuntonsa ja äskettäisten pikaruokaostosten perusteella, jotka on kirjattu maksusovelluksen kautta. Hänen termostaattinsa mukautuu aikatauluihin ja vaikuttaa jopa ennakoivan poikkeuksia rutiinissa, kuten sen, että hän jää kotiin sairastuessaan.

Kysymys

"Missä vaiheessa personointi muuttuu liian tunkeilevaksi? Mihin itse vetäisit rajan?"



KUVAN LÄHDE | Kuvan on luonut Dall-E.

Vaikka Sarah arvostaa näiden tekoälyohjattujen laitteiden tarjoamaa käytettävyyttä, jotka tekevät hänen elämästään sujuvampaa ja tehokkaampaa, hän alkaa tuntea levottomuutta siitä, kuinka paljon henkilökohtaista tietoa ne keräävät ja miten sitä käytetään. Hän ei ollut nimenomaisesti suostunut

jakamaan sijaintiaan tai ostohistoriaansa. Alkuperäinen ihastus käytettävyyteen saa nopeasti huolestuneen sävyn yksityisyyden menetyksestä.



Tämä tarina johdattelee meidät tekoälyn maailmaan, jossa yksityisyyden ja käytettävyyden välinen tasapaino on häilyvä. Tekoälyteknologioiden integroituminen arkeemme tarjoaa ennennäkemätöntä käytettävyyttä, mutta nostaa samalla merkittäviä yksityisyyteen liittyviä kysymyksiä.

KUVAN LÄHDE | Kuvan on luonut Dall-E.

Question:

- Onko teknologia koskaan saanut sinut tuntemaan samoin kuin Sarah? Jos on, mitä teit?
- Miten neuvoisit Sarahia tässä tilanteessa?
- Mitä hän voi tehdä hallitakseen tietoaan paremmin?

3. Yksityisyyden ymmärtäminen tekoälyssä

Yksityisyys voidaan yleisesti ymmärtää yksilön oikeutena pitää henkilökohtaiset tietonsa poissa julkisesta näkyvyydestä sekä hallita omaa dataansa ja vuorovaikutustaan. Tämä oikeus tunnustetaan monissa oikeusjärjestelmissä ja kansainvälisissä sopimuksissa perusoikeutena, joka korostaa henkilökohtaista autonomiaa ja ihmisarvoa.

Tekoälyn kontekstissa yksityisyyden suoja saa uusia ulottuvuuksia. Tekoälyjärjestelmät perustuvat usein suuriin tietomääriin oppiakseen ja tehdäkseen päätöksiä. Näihin tietoihin voi sisältyä arkaluontoista henkilötietoa, jonka virheellinen käsittely voi johtaa vakaviin yksityisyyden loukkauksiin. Kun tekoälyä integroidaan arjen teknologioihin – kuten henkilökohtaisiin avustajiin, puettaviin laitteisiin tai terveydenhuollon ennakoivaan analytiikkaan – yksityisyyden suojaaminen vaatii perinteisten tietosuojatoimien lisäksi myös uusia lähestymistapoja, joilla voidaan vastata tekoälyteknologioiden asettamiin erityishaasteisiin.

"Yksityisyys on perusoikeus, joka on tunnustettu muun muassa Yhdistyneiden Kansakuntien ihmisoikeuksien julistuksessa, kansalaisoikeuksia ja poliittisia oikeuksia koskevassa kansainvälisessä yleissopimuksessa sekä monissa muissa globaaleissa ja alueellisissa sopimuksissa."

Yksityisyyden merkitys tekoälyn yhteydessä on monitasoinen:

- **Luottamus:** Toimiva yksityisyyden suoja on ratkaisevan tärkeää käyttäjien luottamuksen säilyttämiseksi tekoälyteknologioihin. Ilman luottamusta käyttäjät saattavat epäröidä ottaa käyttöön innovaatioita, joista voisi olla heille hyötyä.
- **Eettiset velvoitteet:** Yksilön yksityisyyden kunnioittaminen ja suojeleminen on eettinen velvollisuus. Tämä tarkoittaa, että tekoälyjärjestelmiä tulee suunnitella alusta alkaen siten, etteivät ne loukkaa henkilökohtaisia oikeuksia, vaan huomioivat yksityisyyden suojan osana järjestelmän rakennetta.
- **Säädösten noudattaminen:** Useissa maissa on tiukkoja tietosuojalakeja, kuten EU:n yleinen tietosuoja-asetus (GDPR), jotka asettavat erityisiä velvoitteita organisaatioille, jotka käsittelevät henkilötietoja tekoälyjärjestelmien avulla.

3.1. Tietosuoja: Henkilötietojen suojaaminen luvattomalta käytöltä

Tekoälyn yhteydessä tietosuojalla tarkoitetaan henkilökohtaisten tietojen suojaamista luvattomalta pääsylvä, käytöltä tai paljastamiselta.

Tekoälyjärjestelmät on suunniteltava siten, että data ja arkaluonteinen tieto käsitellään turvallisesti ja että niihin pääsevät käsiksi vain valtuutetut tahot. Tehokkaat tietosuojatoimet auttavat ehkäisemään identiteettivarkauksia, taloudellisia petoksia ja henkilötietojen väärinkäyttöä. Näihin toimenpiteisiin kuuluvat muun muassa:

- **Datan salaaminen:** Datan salaaminen sekä sen ollessa tallennettuna että siirrettäessä, jotta luvaton pääsy voidaan estää.
- **Pääsynhallinta:** Datan käyttöoikeuksien rajoittaminen vain niille henkilöille, jotka tarvitsevat tietoa työtehtäviensä suorittamiseen.
- **Datan anonymisointi:** Henkilöön yhdistettävien tietojen poistaminen tekoälyssä käytettävistä aineistoista, jotta yksityisyyden loukkauksen riskiä voidaan vähentää.

Esimerkki: Verkkokauppayritys kerää asiakastietoja tilausten käsittelyä ja personoitua markkinointia varten. Yritys käyttää salausta suojatakseen asiakkaiden maksutiedot ja henkilökohtaiset tiedot varmistaakseen tietosuojan toteutumisen. Yritys toteuttaa lisäksi tiukat sisäiset käytännöt, jotka rajoittavat työntekijöiden pääsyä arkaluonteisiin tietoihin vain niille, jotka tarvitsevat niitä tilausten käsittelyyn. Asiakkaiden suostumus markkinointiviestintään varmistetaan selkeillä suostumusmenettelyillä (opt-in), ennen markkinointiviestintää.

3.2. Tietoturva: Toimenpiteet tiedon eheyden ja luottamuksellisuuden suojaamiseksi

Tietoturva liittyy tiiviisti tietosuojaan, mutta keskittyy enemmän niihin suojaustoimiin, joiden avulla varmistetaan, että tieto pysyy tarkkana, luotettavana ja ainoastaan valtuutettujen henkilöiden käytettävissä. Tekoälyn kontekstissa:

- **Eheys:** Tiedon suojaaminen luvattomilta muutoksilta, jotka voisivat vaarantaa sen tarkkuuden.
- **Luottamuksellisuus:** Henkilökohtaisten tietojen pitäminen salassa ja vain niille saatavilla, joilla on siihen valtuutus.
- **Turvaprotokollat:** Vahvojen suojausprotokollien, kuten SSL-suojauksen, palomuurien ja tunkeutumisen havaitsemisjärjestelmien käyttöönotto ulkoisten hyökkäysten torjumiseksi.

Esimerkki: Terveystieteiden palveluntarjoaja käyttää potilastietojen tallentamiseen sähköistä potilastietojärjestelmää. Suojatakseen näitä tietoja palveluntarjoaja hyödyntää useita kerroksia tietoturvatavoimia. Tähän sisältyy vahva datan tallennus- ja siirtosalaukset, monivaiheinen tunnistautuminen järjestelmäkäytön yhteydessä sekä säännölliset tietoturva-auditoinnit, joiden avulla tunnistetaan ja korjataan haavoittuvuuksia. Nämä käytännöt auttavat varmistamaan potilastietojen eheyden ja luottamuksellisuuden kyberuhkia vastaan.

3.3. Henkilökohtainen autonomia: Oikeus hallita itseään koskevia tietoja

Digitaalisella aikakaudella – erityisesti tekoälyn kontekstissa – henkilökohtainen autonomia tarkoittaa oikeutta hallita omia tietojaan ja niihin perustuvia päätöksiä. Tekoälyjärjestelmät on suunniteltava siten, että ne kunnioittavat ja vahvistavat henkilökohtaista autonomiaa esimerkiksi seuraavin tavoin:

- **Tietoinen suostumus:** Varmistetaan, että käyttäjille kerrotaan selkeästi, mihin heidän tietojaan käytetään, ja että heidän suostumuksensa saadaan ennen tiedonkeruuta.
- **Läpinäkyvyys:** Annetaan käyttäjille selkeää tietoa siitä, miten tietoja käsitellään ja millä logiikalla tekoäly tekee päätöksiä.
- **Hallintamekanismit:** Tarjotaan käyttäjille mahdollisuus hallita, miten heidän tietojaan käytetään, ja mahdollisuus kieltäytyä tietojenkäsittelystä silloin, kun he niin haluavat.

Esimerkiksi kuntoilunseurantaan tarkoitettu sovellus kerää käyttäjän fyysistä aktiivisuutta koskevaa dataa tarjotakseen henkilökohtaista liikuntaneuvontaa. Sovellus kunnioittaa henkilökohtaista autonomiaa kertomalla käyttäjille selkeästi, millaista dataa kerätään ja mihin tarkoitukseen sitä käytetään. Lisäksi käyttäjät voivat tarkastella omia tietojaan, korjata mahdollisia virheitä ja kieltäytyä tietojen keruusta tietyissä toiminnoissa. Tällainen lähestymistapa varmistaa, että käyttäjillä säilyy hallinta omiin tietoihinsa ja niihin perustuvaan päätöksentekoon.

4. Tietosuoja

Tietosuoja tarkoittaa henkilökohtaisten tietojen suojaamista luvattomalta pääsylvä, käytöltä tai paljastamiselta. Siihen sisältyy varmistaminen siitä, että yksilöillä on hallinta omiin tietoihinsa ja että tietoja käytetään heidän toiveidensa ja voimassa olevien lakien mukaisesti.

4.1. Tietosuojan merkitys tekoälyssä

Tietosuoja on erityisen tärkeää tekoälyn yhteydessä, jossa käsitellään suuria määriä henkilötietoja. Ilman riittävää suojaa arkaluonteiset tiedot voivat olla alttiita tietomurroille ja väärinkäytöksille, mikä voi aiheuttaa vahinkoa, kuten identiteettivarkauksia, taloudellisia menetyksiä tai syrjintää. Lisäksi tekoälyjärjestelmät nojaavat vahvasti dataan tehdäkseen päätöksiä, joten datan laatu ja eheys vaikuttavat suoraan näiden järjestelmien toimivuuteen ja luotettavuuteen.

4.2. Sääntelyn, kuten GDPR:n, rooli

Säännökset, kuten EU:n yleinen tietosuoja-asetus (GDPR), ovat keskeisiä tietosuojaperiaatteiden toteuttamisessa tekoälyssä. GDPR asettaa tiukat vaatimukset henkilötietojen keräämiselle, käsittelylle ja säilyttämiselle. Näihin kuuluu muun muassa yksilöiden nimenomaisen suostumuksen hankkiminen, asianmukaisten tietoturvatöimien toteuttaminen sekä mekanismien tarjoaminen rekisteröidyille tietojen tarkasteluun ja hallintaan.

Sääntelyä, kuten GDPR:ää, noudattamalla pyritään turvaamaan yksilöiden yksityisyysoikeudet ja asettamaan organisaatioille vastuu henkilötietojen käsittelystä.

Esimerkki – Terveyssovelluksen tilanne:

Kuvitellaan terveyssovellus, joka kerää ja tallentaa käyttäjiensä terveystietoja, kuten sairaushistoriaa, diagnooseja ja hoitosuunnitelmia. Ilman asianmukaista suostumusta tai tietoturvatavoimia tällainen sovellus aiheuttaa merkittäviä riskejä käyttäjän tietosuojalle. Esimerkiksi, jos sovelluksesta puuttuu vahva salaus tai pääsynhallinta, se voi olla altis tietomurroille ja paljastaa arkaluonteiset terveystiedot luvattomille tahoille. Lisäksi, jos sovellus jakaa käyttäjätietoja kolmansien osapuolten kanssa ilman nimenomaista suostumusta tai anonymisointia, se voi rikkoa tietosuojasäädöksiä, kuten GDPR:ää, mikä voi johtaa oikeudellisiin seuraamuksiin.

Tämä tilannekuva korostaa tiukkojen tietosuojakäytäntöjen toteuttamisen tärkeyttä tekoälysovelluksissa, jotta voidaan suojata yksilöiden arkaluonteiset tiedot ja varmistaa sääntelyn noudattaminen.

4.3. Tietosuojan keskeiset osa-alueet

Tietosuoja on olennainen osa nykyaikaisia digitaalisia käytäntöjä, erityisesti kun datalähtöiset teknologiat ja palvelut yleistyvät. Alla on esitelty joitakin tietosuojan keskeisiä elementtejä, jotka auttavat suojaamaan henkilötietoja ja varmistamaan eettisen tiedonkäsittelyn:

- **Suostumus:** Suostumus tarkoittaa sitä, että yksilöiltä on saatava lupa ennen kuin heidän tietojaan kerätään, käytetään tai jaetaan. Suostumuksen on oltava vapaaehtoinen, yksilöity, tietoon perustuva ja yksiselitteinen. Tämä tarkoittaa, että yksilön on oltava täysin tietoinen siitä, mihin hän antaa suostumuksensa, ja hänen on ymmärrettävä selkeästi tietojenkäsittelytoimien laajuus ja vaikutukset.
- **Tarkoitussidonnaisuus:** Tämän periaatteen mukaan tietoja saa kerätä vain tiettyihin, nimenomaisiin ja laillisiin tarkoituksiin, eikä niitä saa käsitellä myöhemmin tavoilla, jotka eivät ole yhteensopivia alkuperäisten tarkoitusten kanssa. Tarkoitussidonnaisuus varmistaa, että tietoja käytetään vain sellaisiin käyttötarkoituksiin, joihin rekisteröity on antanut suostumuksensa, ja estää tietojen väärinkäyttöä tai vastuutonta käsittelyä.
- **Tietojen minimointi:** Tietojen minimointi tarkoittaa, että kerätään ja käsitellään vain ne tiedot, jotka ovat välttämättömiä käsittelytarkoituksen saavuttamiseksi. Tämä periaate kannustaa organisaatioita rajoittamaan tietojen keruun ainoastaan siihen, mikä on olennaista ja tarpeellista. Tämä vähentää tietomurtoihin liittyvää riskiä, koska vähemmän tietoa on mahdollista joutua väriin käsiin.
- **Vastuullisuus:** Vastuullisuus tarkoittaa organisaatioiden velvollisuutta kantaa vastuu käsittelemistään tiedoista ja osoittaa, että ne noudattavat kaikkia tietosuojaperiaatteita. Tämä sisältää muun muassa rekisterinpidon tietojenkäsittelytoiminnoista, tietosuojatoimien käyttöönnoton ja sen osoittamisen, miten sääntelyn noudattaminen toteutuu. Vastuullisuus varmistaa, että organisaatioiden tietojenkäsittelykäytännöt ovat läpinäkyviä ja että ne voidaan asettaa vastuuseen toiminnastaan.

- **Läpinäkyvyys:** Läpinäkyvyys edellyttää, että kaikki henkilötietojen käsittelyyn liittyvä tieto ja viestintä on helposti saatavilla ja ymmärrettävissä yksilöille. Tietosuojakäytännöt ja tietojen keruuilmoitukset on laadittava selkeällä, yksinkertaisella kielellä. Läpinäkyvyys rakentaa luottamusta, edistää oikeudenmukaisuutta ja antaa käyttäjille mahdollisuuden tehdä tietoisia päätöksiä omien tietojensa suhteen.
- **Pääsy ja oikaisu:** Yksilöillä on oikeus päästä käsiksi omiin tietoihinsa ja oikaista heistä tallennetut virheelliset tiedot. Mahdollisuus tarkastella ja päivittää tietoja tarvittaessa varmistaa tietojen paikkansapitävyyden ja antaa yksilöille hallinnan omiin tietoihinsa, mikä on henkilökohtaisen autonomian keskeinen osa.
- **Tietoturva:** Tietoturvalla tarkoitetaan asianmukaisten teknisten ja organisatoristen toimenpiteiden käyttöönottoa henkilötietojen suojaamiseksi tahattomalta tai laittomalta tuhoutumiselta, katoamiselta, muuttamiselta, luvattomalta luovuttamiselta tai pääsylvä. Näihin käytäntöihin kuuluvat esimerkiksi salaus, turvallinen tietojen tallennus sekä fyysisen ja digitaalisen pääsyn hallinta.

Yhdessä nämä tietosuojan osa-alueet muodostavat vankan viitekehyksen, jonka avulla organisaatiot voivat suojata henkilötietoja ja kunnioittaa yksilöiden oikeuksia yksityisyyteen. Ne auttavat rakentamaan yksityisyyttä kunnioittavaa toimintakulttuuria, joka on linjassa sekä lainsäädännön vaatimusten että digitaalisen aikakauden eettisten odotusten kanssa.

5. Käytettävyyden ymmärtäminen tekoälyssä

Tekoälyn kontekstissa käytettävyys tarkoittaa sitä, kuinka vaivattomasti ja tehokkaasti tehtävät voidaan suorittaa tai kokemuksia parantaa tekoälyteknologioiden avulla. Tekoäly tehostaa prosesseja, vähentää manuaalista työtä ja tarjoaa yksilöllisiä, käyttäjän mieltymysten mukaan räätälöityjä kokemuksia.

Tekoälyn mahdollistama käytettävyys kattaa automaation, personoinnin ja tehokkuuden saumattoman yhdistämisen eri osa-alueilla. Tekoälypohjaiset ratkaisut optimoivat prosesseja, tukevat päätöksentekoa ja rikastuttavat käyttäjäkokemuksia esimerkiksi terveydenhuollossa, taloudessa ja arjen toiminnoissa – lisäksi tuottavuutta ja parantaen hyvinvointia.

Käytettävyys tekoälyn näkökulmasta tarkoittaa teknologian sujuvaa integroimista osaksi tehtävien suorittamista ja tehokkuuden kasvattamista. Tekoäly hyödyntää algoritmeja ja koneoppimista prosessien automatisointiin, manuaalisen työn vähentämiseen ja kokemusten personointiin. Tekoälyjärjestelmät optimoivat työnkulkuja analysoimalla suuria tietomääriä ja oppimalla niistä, säästäten aikaa ja resursseja samalla kun ne tuottavat parempia tuloksia.

Alla on joitakin esimerkkejä siitä, miten tekoäly edistää käytettävyyttä eri alueilla:

- **Automaatiot ja rutiinitehtävät:** Tekoäly on erityisen tehokas toistuvien ja yksitoikkoisten tehtävien automatisoinnissa, jotka perinteisesti vaativat ihmisen osallistumista. Esimerkiksi tekoälypohjainen ohjelmisto voi hoitaa ajanvarauksia, vastata asiakaskyselyihin tai hallita sähköposteja automaattisesti. Tämä nopeuttaa prosesseja ja vapauttaa ihmisten aikaa vaativampiin ja luovempiin tehtäviin, lisäten samalla tuottavuutta ja tehokkuutta.
- **Personointi:** Tekoälyjärjestelmät voivat analysoida valtavia määriä dataa räätälöidäkseen palveluita ja tuotteita yksilöllisten mieltymysten ja tarpeiden mukaan. Kuluttajakontekstissa tämä voi näkyä esimerkiksi tuotteiden suosittelemisena aiempien ostosten tai katseluhistorian perusteella – kuten verkkokaupoissa tai suoratoistopalveluissa. Henkilökohtaisemmissa sovelluksissa tekoäly voi säätää kodin lämmitysjärjestelmiä asukkaan aikataulun ja mieltymysten mukaan, varmistaen mukavuuden ja optimoiden samalla energiankulutuksen.
- **Päätöksenteon tehostaminen:** Tekoäly parantaa päätöksentekoa tarjoamalla yksilöille ja organisaatioille oivalluksia, jotka perustuvat laajojen ja monimutkaisten tietoa-aineistojen analysointiin – sellaisten, joita ihmisten olisi vaikea käsitellä manuaalisesti. Esimerkiksi terveydenhuollossa tekoälyalgoritmit auttavat sairauksien diagnosoinnissa analysoimalla lääketieteellistä kuvadataa nopeammin ja usein tarkemmin kuin ihmislääkärit. Rahoitusallalla tekoäly voi analysoida markkinadataa tarjotakseen sijoitusnäkemyksiä tai havaita petollisia tapahtumia erittäin tarkasti ja nopeasti.
- **Tehokkuus ja resurssien hallinta:** Tekoäly parantaa merkittävästi kokonaistehokkuutta ja optimoi resurssien hallintaa. Esimerkiksi tekoälypohjaiset logistiikka- ja toimitusketjunhallintajärjestelmät voivat nopeasti ennustaa kysyntää, optimoida toimitusreitit ja hallita varastotilannetta, mikä vähentää hävikkiä ja kustannuksia. Samoin tekoälyyn perustuvat älyverkot voivat ohjata sähköntoimitusta kaupungeissa tehokkaammin, mukautuen reaaliajassa kysynnän vaihteluihin.
- **Esteettömyys ja helppokäyttöisyys:** Tekoälyteknologiat tekevät usein teknologiasta saavutettavampaa ja helpompaa suuremmalle joukolle ihmisiä, mukaan lukien vammaiset. Esimerkiksi tekoälyllä toimivat ääniohjatut laitteet auttavat näkörajoitteisia tai liikuntarajoitteisia henkilöitä käyttämään teknologiaa ja hallitsemaan ympäristöään tehokkaammin. Tekoäly voi myös ylittää kielimuurit reaaliaikaisten käännöspalvelujen avulla, tehden tiedosta ja viestinnästä saavutettavaa myös ei-äidinkielisille käyttäjille.

5.1. Käytettävyyden tosielämän esimerkkejä

Terveydenhuolto: Terveydenhuollossa tekoälyteknologiat tarjoavat sujuvuutta automatisoimalla diagnoosiprosesseja, analysoimalla lääketieteellisiä kuvia ja tunnistamalla potilastiedoista malleja, jotka auttavat terveydenhuollon ammattilaisia sairauksien diagnosoinnissa nopeammin ja tarkemmin. Tämä säästää aikaa sekä hoitohenkilökunnalta että potilailta, tehden terveydenhuollon toimituksesta tehokkaampaa ja parantaen hoitotuloksia.



KUVAN LÄHDE | Kuvan on luonut Dall-E

Sujuvuutta terveydenhuoltoon tekoälyn avulla

- **Automatisoidut diagnoosiprosessit:** Tekoäly nopeuttaa lääketieteellisten tietojen analysointia, kuten kuvantamista ja diagnoositestejä, mikä mahdollistaa nopeammat diagnoosit.
- **Tarkemmat diagnoosit:** Tekoälyalgoritmit tunnistavat datasta hienovaraisia malleja, jotka saattavat jäädä ihmiseltä huomaamatta, mikä parantaa diagnoosien tarkkuutta ja luotettavuutta.
- **Ennakoiva analytiikka:** Tekoäly hyödyntää dataa useista lähteistä, kuten puettavasta teknologiasta, ennustaakseen mahdollisia terveysongelmia jo ennen kuin ne muuttuvat vakaviksi, edistäen ennaltaehkäisevää terveydenhuoltoa.
- **Sujuvampi potilashallinta:** Tekoäly tehostaa sairaaloiden työnkulkuja hallitsemalla aikatauluja, potilaiden vastaanottoa ja ohjaamista oikeaan hoitopaikkaan, mikä lyhentää odotusaikoja ja parantaa potilaskokemusta.
- **Yksilölliset hoitosuunnitelmat:** Algoritmit analysoivat potilaskohtaista tietoa ja räätälöivät hoitosuunnitelmia, jotka todennäköisimmin toimivat kunkin potilaan yksilöllisen terveydentilan perusteella.

Rahoitusala: Tekoälypohjaiset rahoitustyökalut tarjoavat sujuvuutta tarjoamalla yksilöllistä sijoitusneuvontaa henkilökohtaisten taloudellisten tavoitteiden, riskinsietokyvyn ja markkinatrendien perusteella. Nämä työkalut käyttävät algoritmeja analysoidakseen valtavia määriä taloustietoa ja antavat suosituksia sijoitusstrategioista, auttaen yksilöitä tekemään harkittuja päätöksiä ja optimoimaan taloudellisia salkkujaan.



KUVAN LÄHDE | Kuvan on luonut Dall-E.

Sujuvuus rahoitusosalalla tekoälyn avulla

- **Yksilöllinen sijoitusneuvonta:** Tekoälytyökalut analysoivat henkilön taloudellisia tavoitteita, riskinsietokykyä ja taloudellista tilannetta tarjotakseen räätälöityjä sijoitussuosituksia.
- **Edistynyt data-analyysi:** Nämä järjestelmät hyödyntävät algoritmeja seulomaan suuria määriä taloustietoa – mukaan lukien markkinatrendit ja historiallinen kehitys – tunnistakseen parhaat sijoitusmahdollisuudet.
- **Optimoitu salkunhallinta:** Tekoäly auttaa säätämään sijoitussalkkuja dynaamisesti reaaliaikaisten markkinamuutosten ja yksilöllisten taloudellisten tavoitteiden perusteella, parantaen tuottomahdollisuuksia ja samalla halliten riskialttiutta.
- **Tehokkaat markkinanäkymät:** Tekoälytyökalut tarjoavat nopeita näkemyksiä markkinatilanteesta, auttaen sijoittajia ymmärtämään monimutkaisia markkinadynamiikkoja ja tekemään päätöksiä oikea-aikaisesti.
- **Riskien arviointi:** Algoritmit arvioivat eri sijoitusvaihtoehtoihin liittyviä riskejä, mahdollistaen käyttäjien tehdä tietoon perustuvia päätöksiä oman riskinsietokykynsä mukaan.
- **Transaktioiden automatisointi:** Tekoäly voi automatisoida kaupankäynnin ja sijoitustapahtumat, tehostaen toteutusprosessia ja vähentäen inhimillisten virheiden mahdollisuutta.
- **Parannettu tietoturva:** Kehittyneet turvatoimet, kuten poikkeavuuksien havaitseminen, auttavat tekoälyn avulla tunnistamaan ja torjumaan mahdollisia uhkia sijoitustileille.

Arkielämä: Tekoölyyn perustuvat älykodin laitteet, kuten ääniohjaimet, älytermostaatit ja automatisoidut kotiturvajärjestelmät, tarjoavat sujuvuutta automatisoimalla kotitöitä ja parantamalla mukavuutta ja turvallisuutta. Nämä laitteet oppivat käyttäjän mieltymyksiä ajan mittaan ja säättävät asetuksia niiden mukaisesti, tarjoten saumattomia ja yksilöllisiä käyttökokemuksia, jotka yksinkertaistavat arjen rutiineja ja parantavat elämänlaatua.



IMAGE SOURCE | Image generated by Dall-E

Arjen sujuvuus tekoölyn avulla

- **Kotitöiden automatisointi:** Tekoölyyn perustuvat laitteet, kuten robotti-imurit ja automatisoidut keittiölaitteet, hoitavat arkisia askareita, vapauttaen käyttäjien aikaa muuhun.
- **Personoidut asetukset:** Laitteet, kuten älytermostaatit ja valaistusjärjestelmät, oppivat käyttäjän mieltymyksiä ja mukautuvat niihin, sääten asetukset automaattisesti optimaalisen mukavuuden saavuttamiseksi.
- **Ääniavusteinen teknologia:** Ääniohjaimet, kuten Alexa ja Google Assistant, mahdollistavat kotilaitteiden handsfree-ohjauksen, helpottaen tiedonhakua ja tehtävien hallintaa.
- **Parannettu kotiturvallisuus:** Älykkäät turvajärjestelmät käyttävät tekoälyä kodin valvontaan, tunnistavat tuttuja kasvoja, havaitsevat poikkeavaa toimintaa ja ilmoittavat omistajalle mahdollisista uhkista.
- **Energiätehokkuus:** Tekoölypohjaiset laitteet optimoivat kodin energiankäyttöä oppimalla huippukulutusajat ja säättämällä toimintojaan niin, että hukkaa syntyy vähemmän – mikä pienentää myös sähkölaskuja.
- **Saumaton yhteensopivuus:** Älykodin ekosysteemit yhdistävät eri laitteet toisiinsa, jolloin ne toimivat sujuvasti yhdessä. Tämä lisää kodin sujuvuutta integroitujen ohjaus- ja vuorovaikutusratkaisujen avulla.
- **Etävalvonta ja -ohjaus:** Aukkaat voivat valvoa ja hallita älykodin laitteita etänä älypuhelimella, mikä tuo mielenrauhaa erityisesti poissa ollessa.

Vaikka tekoäly lisää merkittävästi sujuvuutta, se herättää myös tärkeitä kysymyksiä, kuten yksityisyyden suojan huolia, tarpeen inhimilliselle valvonnalle sekä riskin työpaikkojen katoamisesta tietyillä aloilla. On tärkeää varmistaa, että tekoälyjärjestelmiä käytetään eettisesti ja vastuullisesti, jotta niiden hyödyt voidaan maksimoida ja mahdolliset haitat minimoida. Yhteenvetona voidaan todeta, että sujuvuus tekoölyssä tarkoittaa tekoölyn voimavarojen hyödyntämistä elämän helpottamiseksi, miellyttävämmäksi tekemiseksi ja tehokkuuden lisäämiseksi. Tekoälyteknologioiden kehittyessä ja levitessä eri elämäntilanteille niiden mahdollisuus lisätä sujuvuutta kasvaa, ja ne lupaavatkin mullistavia vaikutuksia koko yhteiskunnan tasolla.

Vinkkejä ja suosituksia opettajille

Esitä hypoteettisia skenaarioita, joissa osallistujat haastetaan päättämään, onko yksityisyys vai käytettävyys etusijalla.

Esimerkki

Hypoteettinen skenaario: Terveysterveystasovellus

Tausta: Harkitset uutta terveysseurantasovellusta, HealthMatea. Tämä sovellus tarjoaa henkilökohtaisia kuntoilu- ja ruokavaliosuunnitelmia analysoimalla päivittäisiä aktiviteettejasi, ruokailuasi ja terveystavotteitasi.

Skenaarion haaste: HealthMate voi synkronoitua sosiaalisen median kanssa, jotta se voi kerätä lisää elämäntapatietoja henkilökohtaistamisen tehostamiseksi ja jakaa tietoja kumppaneiden kanssa, jotka voivat lähettää sinulle kohdennettuja mainoksia terveystavoitteidesi perusteella.

Kysymyksiä pohdintaa varten:

- Suostuisitko tehostettuun personointiin tietäen, että tietojasi saatetaan käyttää kohdennettuun mainontaan?
- Miten punnitset räätälöidyn terveystasovelluksen etuja suhteessa tietojasi jakamiseen liittyviin riskeihin?

Tämä skenaario kehottaa osallistujia arvioimaan räätälöityjen terveystasovellusten mukavuuden ja tietojen jakamiseen liittyvien yksityisyydensuojariskien välistä tasapainoa.

6. Yksityisyyden ja käytettävyyden vuorovaikutus

Tietosuojan ja sujuvuuden välinen dynaaminen vuorovaikutus nykyaikaisessa teknologiassa tarjoaa sekä mahdollisuuksia että haasteita. Teknologian sulautuessa yhä tiiviimmin osaksi arkeamme, näiden osa-alueiden tasapainottamisen ymmärtäminen nousee keskeiseksi. Alla tarkastellaan tietosuojan ja sujuvuuden välisen tasapainottelun seurauksia, eettisiä ristiriitoja ja tosielämän vaikutuksia.

Tietosuojan ja käytettävyyden väliset kompromissit:

Näiden kahden osa-alueen tasapainottaminen vaatii huolellista harkintaa ja valmiutta tehdä kompromisseja:

- **Rajoitukset tiedonkeruussa:** Tietosuojan vahvistaminen edellyttää usein sitä, että kerättävän ja käytettävän datan määrää ja tyyppiä rajoitetaan. Tämä voi heikentää tekoälyjärjestelmien toimivuutta, sillä monet niistä

perustuvat suuriin tietomassoihin palvelun ja käyttökokemuksen parantamiseksi.

- **Yksityisyyden uhraaminen käytettävyyden vuoksi:** Toisaalta sujuvuuden parantaminen edellyttää usein laajamittaista henkilökohtaisen datan keruuta ja analysointia. Tämä voi sisältää käyttäytymisen, mieltymysten ja sijainnin seurannan – mikä saattaa kaventaa yksityisyyttä ja tuntua tunkeilevalta.

Eettiset ristiriidat, joita näistä kompromisseista seuraa:

Tietosuojan ja sujuvuuden tasapaino tuo esiin useita eettisiä ristiriitoja:

- Yksityisyyden vaarantaminen: Kuinka paljon yksityisyyttään ihmiset ovat valmiita luovuttamaan saadakseen parempaa sujuvuutta? Vastaus vaihtelee usein yksilön käsitysten ja teknologian käyttökontekstin mukaan.
- Tiedonkeruukäytännöt: Organisaatioilla on eettinen vastuu henkilökohtaisen tiedon keräämisestä, käytöstä ja jakamisesta. Eettisten periaatteiden tulisi ohjata päätöksiä siitä, mitä tietoja kerätään, mihin niitä käytetään ja kuinka paljon käyttäjille kerrotaan näistä käytännöistä.

Käytännön seuraukset, kun käytettävyys asetetaan yksityisyyden edelle – tai päinvastoin:

Päätökset, jotka koskevat tietosuojan ja sujuvuuden välistä tasapainoa, johtavat konkreettisiin vaikutuksiin:

- Käytettävyyden asettaminen etusijalle: Käytettävyyden priorisointi lisää usein riskejä, kuten tietomurtoja, identiteettivarkauksia ja vakavia yksityisyydensuojan loukkauksia. Tämä voi heikentää käyttäjien luottamusta pitkäaikaisesti ja johtaa yritysten osalta mahdollisiin oikeudellisiin seuraamuksiin.
- Vaikutukset, kun yksityisyys asetetaan etusijalle: Jos yksityisyyttä painotetaan liiallisesti, se saattaa heikentää joidenkin teknologisten ratkaisujen tehokkuutta. Tämä voi johtaa vähemmän personoituihin palveluihin ja mahdollisesti hidastaa teknologista innovaatiota ja kehitystä, joka voisi hyödyttää yhteiskuntaa laajemmin.

Tietosuojan ja sujuvuuden tasapainottaminen vaatii hienovaraisuutta ja lähestymistapaa, joka ottaa huomioon teknologian eettiset, sosiaaliset ja henkilökohtaiset vaikutukset. Sekä organisaatioiden että yksilöiden tulee pyrkiä löytämään keskitie, joka kunnioittaa käyttäjän yksityisyyttä, mutta mahdollistaa samalla teknologian tarjoamien hyötyjen hyödyntämisen. Tämä tasapaino ei ole pysyvä, vaan sitä on arvioitava jatkuvasti teknologian kehittyessä ja yhteiskunnallisten arvojen muuttuessa. Näiden dynamiikkojen ymmärtäminen auttaa sidosryhmiä tekemään tietoisia päätöksiä, jotka suojaavat yksilön oikeuksia samalla kun omaksutaan digitaalisen aikakauden kehitysmuotoja.

7. Yksityisyyden ja käytettävyyden merkitys

Yksityisyys ja käytettävyys tekoälyssä ovat keskeisiä tekijöitä käyttäjien luottamuksen ja teknologian käyttöönoton muovaamisessa. Kun tekoäly sulautuu osaksi jokapäiväistä elämää, on olennaista löytää tasapaino henkilökohtaisten tietojen suojaamisen ja käyttäjäkokemusten parantamisen välillä. Yksityisyys takaa henkilötietojen suojelun, kun taas käytettävyys lisää tehokkuutta ja mahdollistaa teknologian personoidun käytön. Haasteena on sovittaa nämä näkökulmat yhteen tavalla, joka edistää eettistä tekoälyn käyttöönottoa ja yhteiskunnallista hyväksyntää – varmistaen, että tekoälyn kehitys tuo hyötyjä vaarantamatta yksilön oikeuksia.

Käyttäjäluottamuksen ja teknologian käyttöönoton vahvistaminen

Yksityisyyden ja käytettävyyden priorisointi tekoälyteknologioissa on olennaista käyttäjäluottamuksen rakentamiseksi ja laajemman käyttöönoton edistämiseksi. Kun tekoälyjärjestelmät hallitsevat yksityisyyttä läpinäkyvästi ja tarjoavat saumattoman käytettäviä käyttökokemuksia, ne saavat käyttäjien hyväksynnän todennäköisemmin. Näin näiden osa-alueiden korostaminen voi vaikuttaa käyttäjäluottamukseen ja teknologian omaksumiseen:

- **Luottamuksen rakentaminen:** Käyttäjät luottavat todennäköisemmin teknologiaan, joka kunnioittaa heidän yksityisyyttään ja helpottaa arkea ilman merkittävää tunkeilua. Tekoälyratkaisut, jotka suojaavat käyttäjätietoja tehokkaasti ja samalla tarjoavat personoituja ja tehokkaita kokemuksia, rakentavat positiivista mainetta käyttäjien keskuudessa.
- **Käytön edistäminen:** Kun käyttäjät kokevat, että tekoälyteknologia parantaa heidän tuottavuuttaan tai käytettävyyttään vaarantamatta heidän yksityisyyttään, he omaksuvat sen todennäköisemmin ja suosittelevat sitä myös muille. Tekoälyn arvon osoittaminen käytettävyyden kautta – yhdistettynä vahvoihin tietosuojakäytäntöihin – kannustaa käyttäjiä ottamaan nämä teknologiat osaksi arkipäivän rutinejaan.

Tekoälyn käyttöönottoon liittyvät oikeudelliset ja eettiset näkökulmat

Tekoälyteknologioiden käyttöönotto edellyttää tarkkaa harkintaa sekä oikeudellisten että eettisten vaikutusten osalta, jotta varmistetaan vastuullinen käyttö ja säädösten noudattaminen:

- **Säädösten noudattaminen:** Lait, kuten Euroopan unionin yleinen tietosuojasetus (GDPR), asettavat tiukat ohjeet tietosuojan toteuttamiselle. Kaikkien tekoälyteknologioiden, jotka toimivat EU:ssa tai kohdistuvat EU:n käyttäjiin, on noudatettava näitä sääntöjä. Sääntelyn noudattaminen ei ole vain oikeudellinen velvoite, vaan myös tapa osoittaa sitoutumista käyttäjien yksityisyyden suojaan.

- **Eettiset viitekehykset:** Lakivaatimusten lisäksi eettiset viitekehykset ohjaavat tekoälyn kehittämistä varmistamaan, että teknologiaa hyödynnetään yhteiskunnan parhaaksi. Näiden avulla kehittäjät ja yritykset voivat käsitellä monimutkaisia kysymyksiä, kuten algoritmien vinouma, oikeudenmukaisuus, läpinäkyvyys ja vastuullisuus tekoälyjärjestelmissä.

Vaikutus yhteiskunnallisiin normeihin ja yksilön käyttäytymiseen

Tekoälyn rooli yksityisyyden ja käytettävyyden tasapainottamisessa vaikuttaa merkittävästi sekä yhteiskunnallisiin normeihin että yksilöiden käyttäytymiseen:

- **Säädösten noudattaminen:** Lait, kuten Euroopan unionin yleinen tietosuojasetus (GDPR), asettavat tiukat ohjeet tietosuojan toteuttamiselle. Kaikkien tekoälyteknologioiden, jotka toimivat EU:ssa tai kohdistuvat EU:n käyttäjiin, on noudatettava näitä sääntöjä. Sääntelyn noudattaminen ei ole vain oikeudellinen velvoite, vaan myös tapa osoittaa sitoutumista käyttäjien yksityisyyden suojaan.
- **Eettiset viitekehykset:** Lakivaatimusten lisäksi eettiset viitekehykset ohjaavat tekoälyn kehittämistä varmistamaan, että teknologiaa hyödynnetään yhteiskunnan parhaaksi. Näiden avulla kehittäjät ja yritykset voivat käsitellä monimutkaisia kysymyksiä, kuten algoritmien vinouma, oikeudenmukaisuus, läpinäkyvyys ja vastuullisuus tekoälyjärjestelmissä.

Ymmärtämällä ja käsittelemällä yksityisyyden ja käytettävyyden välistä monimutkaista vuorovaikutusta tekoälyn kehittäjät ja käyttäjät voivat lisätä teknologian hyväksyttävyyttä ja varmistaa, että se tukee yhteiskunnallista kehitystä ja yksilöiden hyvinvointia. Nämä näkökohdat ovat keskeisiä tekoälyteknologioiden kestäväälle kasvulle yhteiskunnassa, jossa yksityisyyttä ja eettisiä periaatteita arvostetaan yhä enemmän

8. Riskien ja hyötyjen selvittäminen

Tekoälyteknologioihin liittyviä mahdollisia riskejä yksityisyyden ja käytettävyyden näkökulmasta ovat muun muassa:

- **Tietomurrot:** Tekoälyjärjestelmät käsittelevät usein suuria määriä arkaluonteista tietoa, mikä tekee niistä houkuttelevia kohteita haitallisille hyökkäyksille. Tietomurrot voivat johtaa luvattomaan pääsyyn henkilötietoihin, mikä voi aiheuttaa identiteettivarkauksia, taloudellisia menetyksiä ja mainehaittoja.
- **Yksityisyyden menettäminen:** Tekoälyteknologiat saattavat kerätä ja analysoida henkilötietoja ilman yksilön suostumusta tai tietoisuutta, vaarantaen yksityisyysoikeuksia. Tämä voi heikentää käyttäjän ja tekoälyjärjestelmän välistä luottamusta sekä herättää huolia tietojen väärinkäytöstä ja hyväksikäytöstä.

- **Vinoumat tekoälyalgoritmeissa:** Tekoälyalgoritmit voivat ylläpitää koulutusdatan vinoumia, mikä johtaa syrjiviin lopputuloksiin. Vinoutuneet algoritmit voivat aiheuttaa epäoikeudenmukaista kohtelua tai päätöksentekoa, syventäen yhteiskunnallisia eriarvoisuuksia ja heikentäen luottamusta tekoälyjärjestelmiin

Näiden riskien seuraukset:

Tekoälyteknologioihin liittyvät riskit voivat heikentää luottamusta järjestelmiin ja aiheuttaa kielteisiä seurauksia sekä yksilöille että yhteiskunnalle:

- **Tietomurrot:** Tekoälyjärjestelmät käsittelevät usein suuria määriä arkaluonteista tietoa, mikä tekee niistä houkuttelevia kohteita haitallisille hyökkäyksille. Tietomurrot voivat johtaa luvattomaan pääsyyn henkilötietoihin, mikä voi aiheuttaa identiteettivarkauksia, taloudellisia menetyksiä ja mainehaittoja.
- **Yksityisyyden menettäminen:** Tekoälyteknologiat saattavat kerätä ja analysoida henkilötietoja ilman yksilön suostumusta tai tietoisuutta, vaarantaen yksityisyysoikeuksia. Tämä voi heikentää käyttäjän ja tekoälyjärjestelmän välistä luottamusta sekä herättää huolia tietojen väärinkäytöstä ja hyväksikäytöstä.
- **Vinoumat tekoälyalgoritmeissa:** Tekoälyalgoritmit voivat ylläpitää koulutusdatan vinoumia, mikä johtaa syrjiviin lopputuloksiin. Vinoutuneet algoritmit voivat aiheuttaa epäoikeudenmukaista kohtelua tai päätöksentekoa, syventäen yhteiskunnallisia eriarvoisuuksia ja heikentäen luottamusta tekoälyjärjestelmiin.

Mahdolliset hyödyt:

Riskeistä huolimatta tekoälyteknologiat tarjoavat monia merkittäviä hyötyjä:

- **Tietomurrot:** Tekoälyjärjestelmät käsittelevät usein suuria määriä arkaluonteista tietoa, mikä tekee niistä houkuttelevia kohteita haitallisille hyökkäyksille. Tietomurrot voivat johtaa luvattomaan pääsyyn henkilötietoihin, mikä voi aiheuttaa identiteettivarkauksia, taloudellisia menetyksiä ja mainehaittoja.
- **Yksityisyyden menettäminen:** Tekoälyteknologiat saattavat kerätä ja analysoida henkilötietoja ilman yksilön suostumusta tai tietoisuutta, vaarantaen yksityisyysoikeuksia. Tämä voi heikentää käyttäjän ja tekoälyjärjestelmän välistä luottamusta sekä herättää huolia tietojen väärinkäytöstä ja hyväksikäytöstä.
- **Vinoumat tekoälyalgoritmeissa:** Tekoälyalgoritmit voivat ylläpitää koulutusdatan vinoumia, mikä johtaa syrjiviin lopputuloksiin. Vinoutuneet algoritmit voivat aiheuttaa epäoikeudenmukaista kohtelua tai

päätöksentekoa, syventäen yhteiskunnallisia eriarvoisuuksia ja heikentäen luottamusta tekoälyjärjestelmiin.

Riskien lieventäminen:

Organisaatioiden tulisi ottaa käyttöön vahvoja tietoturvatavoimia näiden riskien lieventämiseksi, varmistaa tekoälyjärjestelmien läpinäkyvyys ja vastuullisuus sekä asettaa eettiset näkökulmat etusijalle koko tekoälyn kehityksen elinkaaren ajan. Ennakoivalla riskien hallinnalla organisaatiot voivat rakentaa luottamusta tekoälyjärjestelmiin ja vähentää mahdollisia haittoja yksilöille ja yhteiskunnalle.

Tekoälyteknologioiden hyötyjen ja riskien tasapainottamiseksi organisaatiot voivat omaksua seuraavat strategiat:

- **Ota käyttöön vahvat tietoturvatavoimet:** Organisaatioiden tulisi suojautua tietomurroilta ja luvattomalta pääsylvä käyttämällä vahvoja tietoturvakeinoja. Näihin kuuluvat muun muassa salaus, pääsynhallinta sekä säännölliset tietoturva-auditoinnit.
- **Varmista läpinäkyvyys ja vastuullisuus:** Läpinäkyvyys ja vastuullisuus ovat keskeisiä luottamuksen rakentamisessa tekoälyjärjestelmiin. Organisaatioiden tulisi olla avoimia siitä, miten tekoälyteknologioita käytetään, ja varmistaa, että käytön seurauksista ollaan vastuussa.
- **Aseta eettiset näkökulmat etusijalle:** Eettisten periaatteiden tulisi ohjata tekoälyn kehittämistä ja käyttöönottoa. Organisaatioiden tulee priorisoida oikeudenmukaisuutta, läpinäkyvyyttä ja vastuullisuutta tekoälyjärjestelmissä vinoumien ehkäisemiseksi ja datan eettisen käytön varmistamiseksi.

Ottamalla käyttöön ennakoivan riskienhallinnan ja vastuulliset tekoälyn käyttöönoton käytännöt organisaatiot voivat maksimoida tekoälyteknologioiden hyödyt samalla kun ne minimoivat mahdolliset haitat yksilöille ja yhteiskunnalle.

9. Tapaustutkimus: Tekoäly valvonnassa

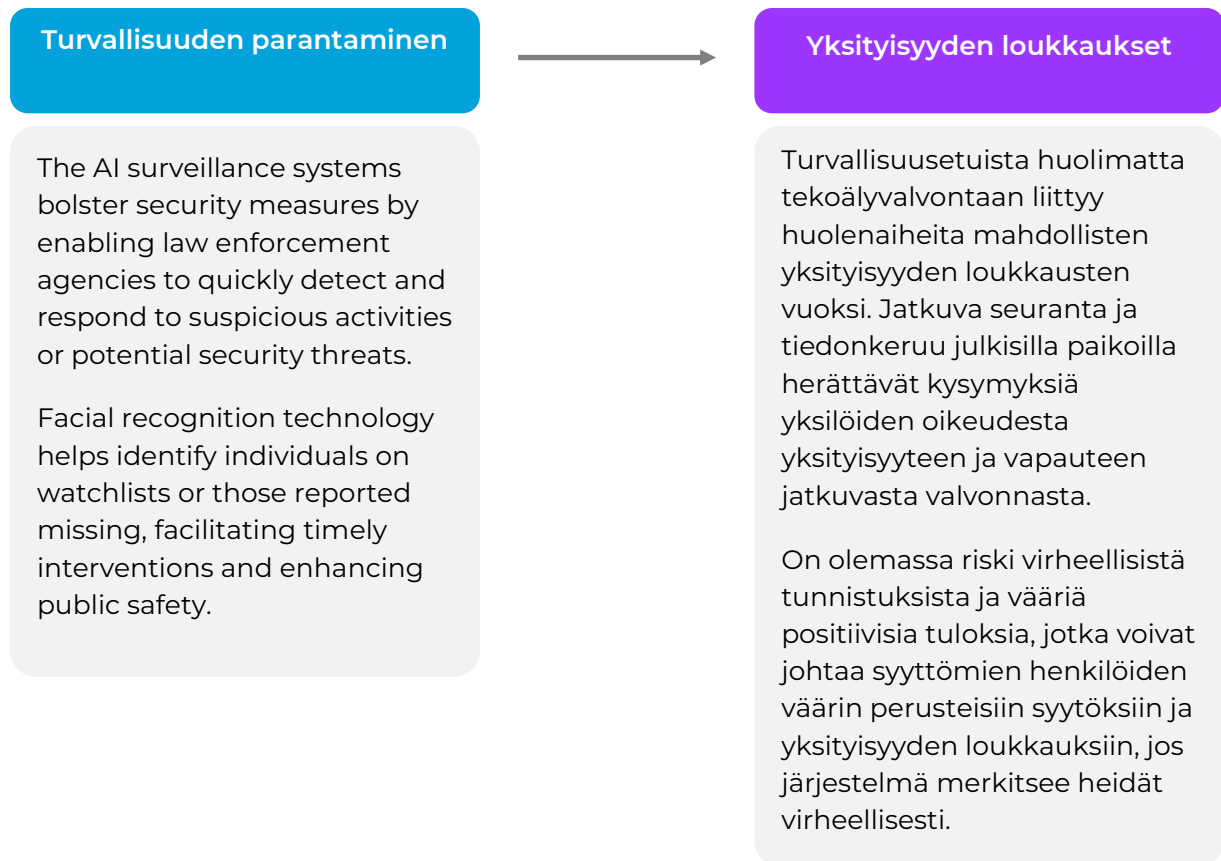
Tilannekuvaus

Paikallisviranomaiset ottavat käyttöön tekoälypohjaiset valvontajärjestelmät vilkkaassa kaupungissa parantaakseen yleistä turvallisuutta ja järjestystä. Näissä järjestelmissä hyödynnetään kehittyneitä teknologioita, kuten kasvojentunnistusta, käyttäytymisanalyysiä ja esineiden tunnistusta, julkisten tilojen valvomiseksi, mahdollisten uhkien tunnistamiseksi ja hätätilanteisiin reagoimiseksi reaaliajassa.



KUVAN LÄHDE | Kuvan on luonut Dall-E.

Turvallisuuden parantaminen vs. yksityisyyden loukkaukset:



Harkitun tekoälyn käyttöönoton merkitys:

Tekoälyn käyttö valvonnassa korostaa harkitun toteutuksen tärkeyttä, jotta turvallisuustarpeet ja yksityisyydensuoja pysyvät tasapainossa:

- **Selkeät politiikat ja sääntely:** Harkittu tekoälyn käyttöönotto edellyttää selkeiden politiikkojen ja sääntöjen luomista valvontateknologioiden käytölle. Tavoitteena on varmistaa läpinäkyvyys, vastuullisuus ja yksityisyydensuojalainsäädännön noudattaminen.
- **Datan eettinen käyttö:** Organisaatioiden on asetettava etusijalle valvontajärjestelmien kautta kerätyn datan eettinen käyttö, kunnioitettava yksilöiden oikeutta yksityisyyteen ja minimoitava riski yksityisyydensuojan loukkauksiin tai henkilötietojen väärinkäyttöön.
- **Algoritmien läpinäkyvyys ja vastuullisuus:** Läpinäkyvien tekoälyalgoritmien ja vastuullisuusmekanismien käyttöönotto varmistaa, että valvontajärjestelmät toimivat oikeudenmukaisesti, tarkasti ja ovat vastuussa toiminnastaan. Tämä vähentää vinoumien tai virheiden riskiä, jotka voisivat johtaa yksityisyyden loukkauksiin.
- **Yleisön osallistaminen ja valvonta:** Kansalaisten osallistaminen tekoälyvalvontaa koskeviin keskusteluihin ja sidosryhmien mukaan ottaminen päätöksentekoon edistää läpinäkyvyyttä, luottamusta ja

vastuullisuutta. Tämä tukee valvontateknologioiden vastuullista käyttöönottoa ja käyttöä.

Yhteenvetona tekoälyn käyttö valvonnassa korostaa harkitun toteutuksen tarvetta, jotta turvallisuuden parantaminen ja yksityisyydensuoja voidaan sovittaa yhteen. Organisaatiot voivat ottaa käyttöön valvontajärjestelmiä, jotka parantavat turvallisuutta ja samalla asettavat etusijalle eettiset periaatteet, läpinäkyvyyden ja vastuullisuuden – kunnioittaen yksilön yksityisyyttä ja edistäten yleistä luottamusta.

10. Yhteenveto

Yksityisyyden ja käytettävyyden tarkastelu tekoälyn yhteydessä korostaa kriittistä tasapainoa, joka on välttämätön luottamuksen rakentamiseksi ja teknologian laajemman käyttöönoton edistämiseksi. Kun tekoäly sulautuu yhä tiiviimmin terveydenhuoltoon, talouteen ja arkipäivän toimintoihin, se tuo mukanaan merkittäviä hyötyjä – kuten tehokkuuden parantamista ja yksilöllisiä palveluita – mutta samalla myös tarpeen vahvoille tietosuojakäytännöille ja eettiselle harkinnalle. Vastuunalainen tekoälyn käyttöönotto edellyttää lainsäädännön, kuten GDPR:n, noudattamista ja eettisten ristiriitojen käsittelemistä.

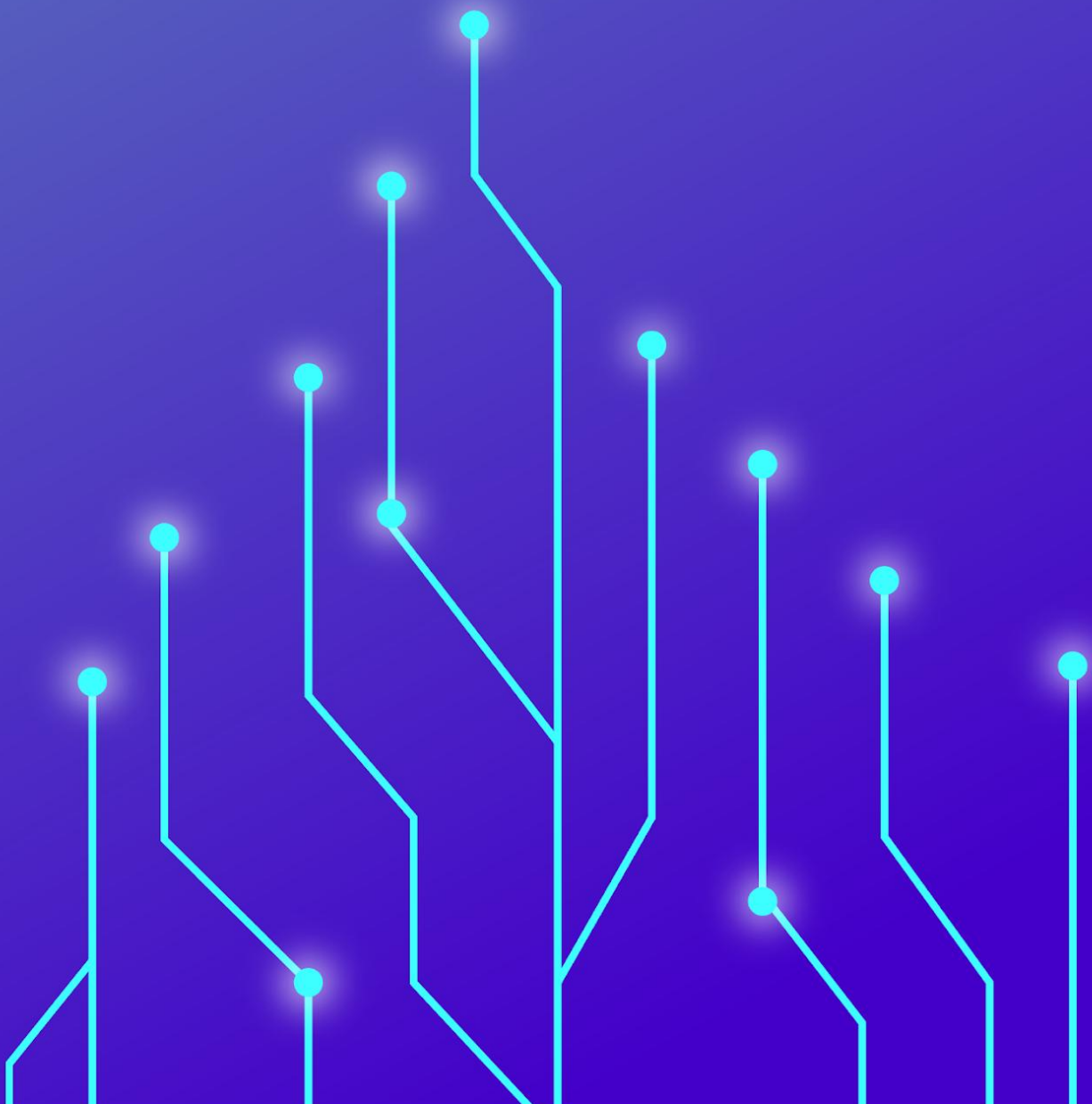
Koska tekoäly muokkaa yhteiskunnallisia normeja ja yksilöiden käyttäytymistä, tasapainon säilyttäminen käytettävyyden ja tietosuojan välillä on entistä tärkeämpää. Kehittäjien, käyttäjien ja päättäjien on tehtävä yhteistyötä sen varmistamiseksi, että tekoälyteknologiat parantavat elämää vaarantamatta perusoikeuksia. Läpinäkyvyyden ja vastuullisuuden korostaminen on avain tekoälyn mahdollisuuksien hyödyntämiseen vastuullisesti, jotta se pysyy hyödyllisenä ja luotettavana teknologiana yhä digitaalisemmassa maailmassa.



Universitat
de les Illes Balears



Osaamiskokonaisuus 3 | Algoritmit ja niiden rajoitukset



Sisällys

1. Introduction	59
2. Understand the fundamental concepts of algorithms, their models, and limitations	62
2.1. Algorithmic Complexity	62
2.2. Algorithm characterization	63
2.3. Algorithm structure	64
2.4. Algorithm Efficiency	67
2.5. Limitations	69
2.6. Limitations of machine learning algorithms	70
3. Recognize the role of algorithms in various sectors and the potential pitfalls arising from their use	73
3.1. Role of algorithms in various sectors	74
3.2. Potential pitfalls of algorithm use	82
4. Comprehend algorithmic complexity, optimization, and the limitations of algorithmic decision-making	85
4.1. Algorithmic Complexity	85
4.2. Time algorithmic complexity	86
4.3. Memory algorithmic complexity	87
4.4. Optimisation	88
4.5. Limitations of algorithmic decision-making	90
4.6. How to identify the potential sources of bias and errors of algorithmic systems? A crucial issue for ensuring the fairness, transparency, and accountability	91
4.7. Addressing and Mitigating Algorithmic Risks and Biases	92
5. References	95

1. Johdanto¹

Tässä kokonaisuudessa perehdytään algoritmien peruskäsitteisiin, niiden malleihin ja rajoituksiin. Osallistujat saavat ymmärryksen siitä, millainen rooli algoritmeilla on eri toimialoilla ja millaisia sudenkuoppia niiden käyttöön voi liittyä. Käsiteltäviä aiheita ovat muun muassa algoritmien kompleksisuus, optimointi sekä algoritmipohjaisen päätöksenteon rajoitukset.

Nykyajan digitaalisessa yhteiskunnassa algoritmeilla on yhä merkittävämpi rooli elämämme eri osa-alueilla – aina verkossa kuluttamastamme sisällöstä automatisoitujen järjestelmien tekemisiin päätöksiin. Algoritmien monimutkaisuuden, optimointistrategioiden ja päätöksenteon rajoitteiden ymmärtäminen on tärkeää kaikille, jotka toimivat näiden teknologioiden parissa.

Algoritmien kompleksisuuden ymmärtäminen tarkoittaa niiden tehokkuuden ja suorituskyvyn tarkastelua erilaisissa laskennallisissa tehtävissä. Se edellyttää käsitystä siitä, kuinka paljon resursseja – kuten aikaa ja muistia – algoritmit vaativat ongelmien ratkaisemiseksi ja miten nämä vaatimukset kasvavat syötteen koon kasvaessa.

Tämä ymmärrys antaa meille valmiudet arvioida, kuinka toteutuskelpoisia ja käyttökelpoisia tietyt algoritmit ovat todellisissa sovelluksissa (Cormen ym., 2009).

Optimointimenetelmät muodostavat toisen keskeisen osa-alueen algoritmien ymmärtämisessä. Näiden menetelmien tavoitteena on parantaa algoritmien tehokkuutta ja suorituskyyä säätämällä niiden parametreja tai uudelleenrakentamalla niiden peruslogiikkaa. Olipa kyse laskenta-ajan vähentämisestä, resurssien hyödyntämisen maksimoimisesta tai tiettyjen tavoitteiden optimoinnista, sopivien optimointistrategioiden hyödyntäminen voi merkittävästi parantaa algoritmien suorituskyyä eri sovellusalueilla (Cormen ym., 2009).

Tekoälypohjaiseen päätöksentekoon liittyvien mahdollisten hyötyjen rinnalla on kuitenkin myös sisäsyntyisiä rajoituksia, jotka vaativat huolellista harkintaa. Algoritmit toimivat ennalta määritellyissä kehyksissä ja ovat riippuvaisia syötteen laadusta sekä taustalla olevien oletusten paikkansapitävyydestä. Vääristymät datassa, virheelliset mallinnusoletukset tai odottamattomat poikkeustilanteet voivat johtaa virheellisiin lopputuloksiin tai ei-toivottuihin seurauksiin, mikä korostaa algoritmipohjaisen päätöksenteon haavoittuvuutta.

Algoritmisen kompleksisuuden, optimointistrategioiden ja päätöksenteon rajoitteiden syvällisempi ymmärtäminen voi vahvistaa näkemystä algoritmien hienovaraisuuksista ja haasteista. Se antaa yksilöille kriittisen ajattelun taitoja,

¹ Note: In the preparation of this document, the authors employed Large Language Models (LLM) to assist with text proofreading, content structuring, and the identification of pertinent examples included in the material. After utilizing this AI tool, the authors meticulously reviewed and refined the content as needed. The authors take full responsibility for the content's integrity and accuracy in the final publication.

joiden avulla he voivat arvioida algoritm ratkaisuja analyttisesti, tunnistaa mahdolliset sudenkuopat ja edistää vastuullisia ja eettisiä algoritmisia käytäntöjä yhä algoritmivetoisemmassa maailmassa.

VINKKEJÄ JA SUOSITUKSIA OPETTAJILLE

Vinkkejä ja suosituksia opettajille algoritmien käsitteiden käsittelemiseksi

1. Selkeytä keskeiset käsitteet

Ennen kuin siirrytään syvällisiin keskusteluihin, varmista, että kaikki osallistujat ymmärtävät peruskäsitteet, kuten "algoritmisen kompleksisuus", "optimointi" ja "algoritmien rajoitteet". Tämä voidaan toteuttaa lyhyellä esityksellä tai jakamalla sanastokooste session alussa.

2. Yhdistä reaali maailman sovelluksiin

Yhdistä abstraktit käsitteet konkreettisiin sovelluksiin eri toimialoilla, kuten rahoituksessa, terveydenhuollossa tai verkkokaupassa. Tämä auttaa osallistujia näkemään algoritmien merkityksen ja vaikutukset arjessa, tehden sisällöstä kiinnostavampaa ja helpommin ymmärrettävää.

3. Hyödynnä visuaalisia apuvälineitä

Algoritmit voivat olla monimutkaisia ja vaikeasti hahmotettavia. Käytä kaavioita, vuokaavioita ja animaatioita havainnollistamaan käsitteitä, kuten algoritmin kulkua, kompleksisuutta ja optimointiprosesseja. Tämä helpottaa osallistujien ymmärrystä ja parantaa muistijälkeä.

4. Rohkaise kysymyksiin

Luo avoin ilmapiiri, jossa osallistujat uskaltavat esittää kysymyksiä. Tämä onnistuu esimerkiksi pitämällä säännöllisiä kysymys- ja vastaustuokioita kunkin pääaiheen jälkeen, varmistaen, että sisältö on ymmärretty ennen etenemistä.

5. Hyödynnä tapaustutkimuksia

Keskustele todellisista tapauksista, joissa algoritmisen päätöksenteko on onnistunut tai kohdannut haasteita. Näiden tapausten analysointi voi herättää keskustelua algoritmien eettisistä näkökohdista, tehokkuudesta ja käytännön toteutuksesta.

6. Kannusta kriittiseen ajatteluun

Rohkaise osallistujia ajattelemaan kriittisesti algoritmimalleihin liittyviä rajoitteita ja mahdollisia vinoumia. Tämä voidaan toteuttaa ryhmäkeskusteluina tai väittelyinä siitä, miten näitä haasteita voidaan lieventää algoritmien suunnittelussa ja toteutuksessa.

Suositteluja virittäytymisharjoituksia

1. Algoritmiroolipeli

Roolipeli, jossa osallistujat simuloivat algoritmin toimintaa joko lajittelemalla

esineitä käsin, seuraamalla yksinkertaista ohjesarjaa tietyn tavoitteen saavuttamiseksi tai näyttelemällä algoritmin päätöksentekoprosessin vaiheita. Harjoitus havainnollistaa askelaskeleelta etenevän prosessoinnin periaatetta algoritmeissa.

2. Algoritmisen kompleksisuuden jäänrikköjä

Aloita yksinkertaisella tehtävällä, kuten kirjojen järjestely värin mukaan, ja lisää siihen vaiheittain vaikeustasoa: ensin järjestely tyyllilajin mukaan, sitten kunkin tyyllilajin sisällä kirjoittajan mukaan. Keskustelkaa, kuinka tehtävän monimutkaisuus kasvaa jokaisen uuden kerroksen myötä ja vertailkaa tätä algoritmisen kompleksisuuden kasvuun ja resurssien kulutukseen.

3. Optimointihaaste

Anna pienryhmille tehtävä, kuten tuolien järjestäminen huoneeseen niin, että mahdollisimman moni mahtuu istumaan, mutta rivejä on mahdollisimman vähän. Anna heidän kokeilla ratkaisuja, keskustella strategioistaan ja esittele lopuksi, miten optimointimenetelmät voisivat parantaa heidän ratkaisuaan.

4. Eettinen väittely

Järjestä väittely algoritmien päätöksenteon eettisistä seurauksista, erityisesti mahdollisista vinoumista ja siitä, mitä tapahtuu, kun luotetaan liikaa automaattisiin järjestelmiin. Tämä virittää osallistujat pohtimaan algoritmien monimutkaisuutta ja vastuullista käyttöä.

5. Riko algoritmi -peli

Anna osallistujille yksinkertainen algoritmi tai sääntöjoukko ja haasta heidät löytämään poikkeustilanteet tai skenaariot, joissa algoritmi epäonnistuu. Tämä harjoitus auttaa ymmärtämään algoritmien rajoituksia ja mahdollisia loogisia virheitä.

Nämä harjoitukset ja strategiat tekevät oppimisprosessista paitsi innostavaa, myös antavat osallistujille syvällisemmän käsityksen algoritmisen päätöksenteon monimutkaisuudesta ja merkityksestä.

2. Ymmärrys algoritmien peruskäsitteistä, niiden malleista ja rajoituksista

2.1. Algoritmien kompleksisuus

Algoritmit ovat nykyään kaikkialla. Ne ohjaavat tietokoneita käsittelemään tietoa, ratkaisemaan ongelmia ja tekemään tärkeitä päätöksiä. Ne voivat ratkaista monenlaisia ongelmia aina tiedon järjestämisestä parhaiden matkareittien löytämiseen tai elokuva-suositusten antamiseen. Suurten tietomassojen ja sosiaalisen median aikakaudella algoritmien toimintaperiaatteiden ymmärtäminen on avain niiden rajoitteiden ja mahdollisten vinoumien hahmottamiseen. Algoritmit eivät pelkästään paranna tietokoneiden suorituskykyä, vaan ne myös vaikuttavat päivittäisiin kokemuksiimme – tehden niistä keskeisiä nykyaikaisessa tietojenkäsittelyssä.

Algoritmi on yksiselitteisten ohjeiden jono jonkin ongelman ratkaisemiseksi. Sen täytyy olla oikea – eli antaa aina oikea ratkaisu – ja päättyvä – sen tulee valmistua lopulta. Se on vaiheittainen menettelytapa, joka kertoo, mitä tehdä tietyn tavoitteen saavuttamiseksi. Voimme verrata algoritmia ruoanlaittoon: noudatamme reseptiä, joka kertoo tarkalleen, mitä aineksia käytetään, kuinka paljon mitäkin, ja mitkä vaiheet suoritetaan. Samalla tavalla algoritmi ohjaa tietokonetta tai ihmistä suorittamaan sarjan toimintoja, jotta saavutetaan haluttu lopputulos – olipa kyseessä numeroiden lajittelu, lyhimmän reitin löytäminen kartalta tai elokuvien suosittelu mieltymystesi perusteella.

Algoritmin käsite on ollut olemassa jo pitkään – aina muinaisten sivilisaatioiden ajoista lähtien. Yksi varhaisimmista esimerkeistä on Eukleideen algoritmi, jonka kehitti antiikin kreikkalainen matemaatikko Eukleides noin vuonna 300 eaa. Se on menetelmä kahden luvun suurimman yhteisen tekijän löytämiseksi.

Historian saatossa monet kulttuurit ja sivilisaatiot kehittivät omia algoritmejaan ratkaistakseen matemaattisia ongelmia, suunnistaakseen maastossa tai suorittaakseen tehtäviä tehokkaammin. Näitä algoritmeja välitettiin usein suullisesti tai kirjoitettujen tekstien kautta.

Itse sana "algoritmi" juontaa juurensa persialaisen matemaatikon Muhammad ibn Musa al-Khwarizmin nimeen, joka eli 800-luvulla islamilaisen kultakauden aikana. Al-Khwarizmi kirjoitti teoksen "Al-Kitab al-Mukhtasar fi Hisab al-Jabr wal-Muqabala" ("Ytimekäs kirja laskennasta täydennyksen ja tasapainottamisen avulla"), jossa hän esitteli systemaattisia menetelmiä matemaattisten yhtälöiden ratkaisemiseksi. Sana "algoritmi" on johdettu hänen latinalaistetusta nimestään Algoritmi.

Sen jälkeen algoritmit ovat kehittyneet merkittävästi ja muodostavat nykyään keskeisen osan monilla aloilla – erityisesti modernin tietotekniikan myötä.

Nykyään algoritmeja käytetään matematiikan lisäksi muun muassa tietojenkäsittelytieteessä, insinööritieteissä, biologiassa, taloustieteessä ja monilla muilla aloilla ratkaisemaan monimutkaisia ongelmia ja automatisoimaan tehtäviä.

2.2. Algoritmin luonnehdinta

Algoritmin luonnehdinnat ovat yrityksiä määritellä tarkasti, mitä algoritmilla tarkoitetaan. Termillä "algoritmi" ei ole yleisesti hyväksyttyä virallista määritelmää. Knuth (1997) esitti kuitenkin viisi ominaisuutta, joita pidetään laajalti algoritmin edellytyksinä:

- **Rajoittuneisuus (Finiteness):** "Algoritmin on aina päätyttävä äärellisen määrän askelia jälkeen ... hyvin rajallisen määrän, kohtuullisen määrän." Rajoittuneisuus takaa sen, että algoritmi päättyy lopulta, eikä jää ikuisen silmukkaan kuluttamaan laskentaresursseja loputtomasti. Tätä voidaan verrata palapelin kokoamiseen: lopulta viimeinenkin pala löytää paikkansa, ja algoritmin suoritus päättyy. Tämä ominaisuus tuo ennakoitavuutta ja hallittavuutta laskentaprosesseihin.
- **Tarkkarajaisuus (Definiteness):** "Jokainen algoritmin askel on määriteltävä tarkasti; suoritettavat toiminnot on kuvattava täsmällisesti ja yksiselitteisesti jokaisessa tilanteessa." Tämä korostaa tarvetta yksiselitteisille ohjeille, jotka eivät jätä tilaa tulkinnalle. Tarkkarajainen algoritmi määrittää selkeästi, mitä toimenpiteitä tulee tehdä missäkin vaiheessa ongelmanratkaisua. Näin vältetään epäselvyydet ja varmistetaan, että algoritmi voidaan toteuttaa oikein – riippumatta käyttäjän taustasta tai asiantuntemuksen tasosta.
- **Syöte (Input):** "...suureet, jotka annetaan algoritmillemme ennen sen käynnistymistä. Nämä syötteet otetaan määrittelyistä objektijoukoista." Syöte viittaa dataan tai informaatioon, jota algoritmi käsittelee tuottaakseen lopputuloksen. Syöte voi olla esimerkiksi numeroita, tekstiä, kuvia tai jopa muita algoritmeja. Syöte on oleellinen, koska se on algoritmin käsittelemä raakamateriaali, jonka avulla se ratkaisee ongelman tai suorittaa tehtävän.
- **Tuloste (Output):** "...suureet, joilla on määriteltä suhde syötteisiin." Tuloste tarkoittaa algoritmin tuottamaa tulosta tai ratkaisua sen jälkeen, kun se on käsitellyt syötteen. Tulosteen muoto voi vaihdella ratkaistavan ongelman mukaan – se voi olla yksittäinen arvo, joukko arvoja, päätös, visuaalinen esitys tai muu informaatio, joka edustaa haluttua lopputulosta.
- **Effectiveness:** "...kaikkien algoritmien suoritettavien operaatioiden on oltava riittävän yksinkertaisia, jotta ne voidaan periaatteessa suorittaa täsmällisesti ja rajallisessa ajassa pelkällä paperilla ja kynällä." Tehokkuus korostaa algoritmien käytännöllisyyttä ja toteutettavuutta. Samalla tavalla kuin resepti, jossa käytetään vaikeasti saatavilla olevia tai epämääräisiä aineksia, menettää tehonsa, algoritmienkin on hyödynnettävä helposti saavutettavia resursseja tehtävän suorittamiseksi. Varmistamalla, että

algoritmi on käytännöllinen ja toteutettavissa, tehokkuus varmistaa, että ratkaisut ovat saavutettavissa käytettävissä olevien resurssien ja ajan puitteissa

Kuten edellä kuvatuista algoritmin ominaisuuksista voidaan huomata, ne voivat vaikuttaa intuitiivisesti selkeiltä, mutta ne eivät ole muodollisesti täsmällisiä. Esimerkiksi ei ole täysin yksiselitteistä, mitä tarkoitetaan "täsmällisesti määriteltynä" tai "yksiselitteisesti ja tarkasti määriteltynä" tai "riittävän yksinkertaisena" ja niin edelleen. Näitä määritelmiä voidaan kuitenkin pitää riittävinä tarkoituksiamme varten.

2.3. Algoritmin rakenne

Yksi algoritmien keskeisistä käsitteistä on abstraktio. Tässä yhteydessä abstraktio tarkoittaa ajatusta siitä, että monimutkaiset yksityiskohdat piilotetaan ja keskitytään vain olennaisiin asioihin, jotka ovat tarpeen ongelman ymmärtämiseksi ja ratkaisemiseksi. Se on kuin zoomaamista ulospäin, jolloin nähdään kokonaiskuva ilman että eksytään yksityiskohtiin. Algoritmien suunnittelussa abstraktio auttaa yksinkertaistamaan ongelmanratkaisua jakamalla sen pienempiin, hallittaviin osiin.

Tämä mahdollistaa sen, että voimme käsitellä jokaista osaa erikseen ilman, että meidän tarvitsee ymmärtää kaikkia yksityiskohtia samanaikaisesti.

Algoritmit koostuvat tyypillisesti useista toisiaan tukevista ohjeista yhden yksittäisen vaiheen sijaan. Kun luodaan monimutkaisempia ohjelmia, suoritetaan sarja näitä ohjeita peräkkäin, toistetaan niitä tai tehdään valintoja tiettyjen ehtojen perusteella.

Algoritmien toiminnan ymmärtäminen edellyttää sen hahmottamista, miten erilaisia ohjeita voidaan yhdistellä. Näiden rakenteiden avulla algoritmin ohjeet voidaan järjestää ja jäsenellä loogisesti. Ne auttavat jakamaan monimutkaiset ongelmat hallittaviin osiin, mikä tekee algoritmista helpommin ymmärrettävän ja toteutettavan.

Algoritmit koostuvat tyypillisesti yksinkertaisista rakenteista, jotka on järjestetty hierarkkisesti. Nämä rakenteet ohjaavat algoritmin ohjeiden kulkua ja määräävät, missä järjestyksessä ohjeet suoritetaan. Tätä kutsutaan usein algoritmin ohjausrakenteeksi.

Ohjausrakenne määrittää olennaisesti sen, missä järjestyksessä algoritmin ohjeet suoritetaan. Se määrittää polun, jota suoritusprosessi seuraa rakenteessa määriteltujen ehtojen perusteella. Ohjausrakenteet ovat keskeinen osa mitä tahansa algoritmia, koska ne varmistavat, että ohjeet suoritetaan oikeassa järjestyksessä ja että haluttu lopputulos saavutetaan.

Yksi kaikkia ohjausrakenteita yhdistävä piirre on se, että niissä on yksi sisään tulopiste ja yksi ulostulopiste. Sisään tulopiste on kohta, jossa suoritusprosessi saapuu ohjausrakenteeseen. Ulostulopiste taas merkitsee kohtaa, jossa suoritusprosessi poistuu ohjausrakenteesta ja siirtyy algoritmin seuraavaan osaan. Tämä ominaisuus varmistaa, että suoritusprosessi etenee ennalta määritellyn polun mukaisesti ohjausrakenteen sisällä, säilyttäen algoritmin eheyden ja oikeellisuuden.

On olemassa kolme ohjeiden koostumismallia:

• **Peräkkäinen (Sequential):** Tässä rakenteessa ohjeet suoritetaan yksi kerrallaan, lineaarisessa järjestyksessä. Se muistuttaa hyvin paljon reseptin seuraamista askel askeleelta. Peräkkäiset algoritmit ovat yleisiä arjen tehtävissä, kuten voileivän tekemisessä tai lukujen summan laskemisessa. Ne varmistavat, että toiminnot tapahtuvat tietyssä järjestyksessä ilman haarautumista tai päätöksentekoa.

Esimerkki | Algoritmi kakun valmistamiseksi

1. Lämmitä uuni.
2. Voitele ja jauhota kakkuvuoka.
3. Vaahdota kulhossa voi ja sokeri.
4. Lisää kananmunat yksitellen, vatkaton jokaisen jälkeen.
5. Sekoita joukkoon vaniljauute.
6. Sekoita toisessa kulhossa jauhota, leivinjauhe ja suola.
7. Lisää kuivat aineet vähitellen taikinaan, vuorotellen maidon kanssa.
8. Kaada taikina valmisteltuun kakkuvuokaan.
9. Laita kakkuvuoka esilämmitettyyn uuniin.
10. Paista 25–30 minuuttia tai kunnes keskelle pistetty tikkuun ei tartu taikinaa.
11. Ota kakku uunista ja anna sen jäähtyä vuoassa 10 minuuttia.
12. Tarjoile kakku.

• **Ehdollinen tai vaihtoehtoinen (Conditional or Alternative):** Tässä rakenteessa ohjeet sisältävät päätöksentekopisteitä. Tiettyjen ehtojen perusteella ohjelma etenee eri polkuja pitkin. Tätä voi verrata tilanteeseen, jossa valitaan useiden vaihtoehtojen välillä. Ehdollisten lausekkeiden (kuten "jos", "muu" ja "vaihtaa") avulla voimme käsitellä erilaisia skenaarioita.

Esimerkki | Algoritmi, joka simuloi päätöksentekoa siitä, otetaanko mukaan sateenvarjo vai takki sään mukaan

1. Tarkista sääennuste sateen varalta.
2. Jos ennuste lupaa sadetta, niin:
 - a. Ota mukaan sateenvarjo.
3. Muussa tapauksessa, jos ennuste ei lupaa sadetta ja lämpötila on alle 15 astetta:
 - a. Ota mukaan takki.
 - b. Lähde ulos.

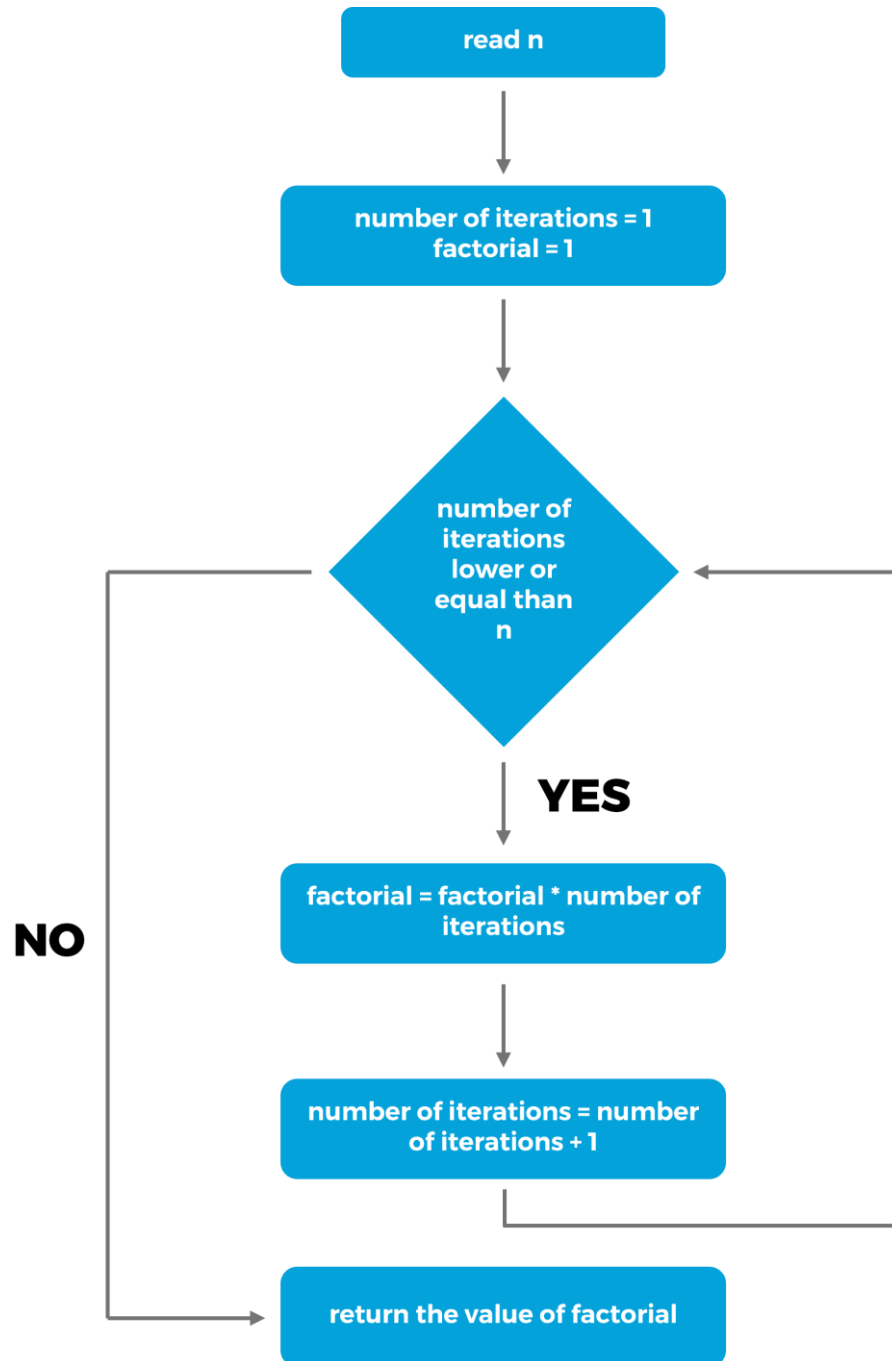
Huomaa, että ohjeiden 3 ja 5 sisennys on merkityksellinen algoritmin määrittelyssä. Nämä ohjeet suoritetaan vain, jos niitä edeltävä ehto täyttyy.

· **Toistuva tai iteratiivinen (Repetitive or Iterative):** Toisto on näiden algoritmien ydin. Ne suorittavat joukon ohjeita toistuvasti, kunnes jokin tietty ehto täyttyy. Silmukat (kuten "for" ja "while") kuuluvat tähän kategoriaan. Ne ovat keskeisiä esimerkiksi suurten tietomassojen käsittelyssä, tapahtumien simuloinnissa tai optimointiongelmien ratkaisemisessa.

Esimerkki | Ohjelma, joka laskee luvun kertoman silmukkaa käyttäen

1. Kertoma (faktoriaali) on matemaattinen operaatio, jossa jokin luku n kerrotaan jokaisella sitä pienemmällä kokonaisluvulla alaspäin lukien.
2. Lue n , eli luku, jonka kertoma halutaan laskea.
3. Aseta kertoman (faktoriaali) alkuarvoksi 1.
4. Toista silmukassa luvut 1:stä n :ään.
5. Kerro kertoma silmukan nykyisellä arvolla.
6. Näytä kertoman lopullinen arvo.

Seuraavassa kuvassa on esimerkki silmukasta.



Source | Own elaboration

Voimme rakentaa yhä monimutkaisempia algoritmeja yhdistelemällä näitä kolmea tekniikkaa. Tarkastellaanpa esimerkkiä algoritmista, joka hyödyntää kaikkia kolmea ohjeiden yhdistelytapaa löytääkseen kokonaislukulistasta suurimman arvon:

10 5 33 12 0 14 55 13 9 11

Lista sisältää 10 kokonaislukua. Luku 10 on listan ensimmäinen alkio.

- Nykyinen alkio on listan ensimmäinen alkio.
- Suurin arvo on nykyinen alkio.
- Niin kauan kuin kaikkia listan alkioita ei ole käsitelty, toista:
- Jos nykyinen alkio on suurempi kuin suurin arvo, niin:
- Suurin arvo on nykyinen alkio.
- Nykyinen alkio on listan seuraava alkio.
- Palauta suurin arvo.

2.4. Algoritmin tehokkuus

Tietojenkäsittelytieteessä algoritmit esitetään ohjelmina. Ohjelma voidaan määritellä algoritmin kuvaukseksi koneen käytettävissä olevien rajallisten käskyjen joukossa. Tietokoneen ohjelmakokonaisuutta kutsutaan ohjelmistoksi. Jotta tietokone saadaan suorittamaan jokin tehtävä, meidän on ensin suunniteltava algoritmi ongelman ratkaisemiseksi ja sen jälkeen ohjelmoitava se tietokoneelle. Tavoitteena on muuntaa käsitteellinen algoritmi selkeäksi joukoksi ohjeita jollakin ohjelmointikielellä, valitun ohjelmointitavan (eli ohjelmointiparadigman) mukaisesti.

Kun ratkaistavat ongelmat muuttuvat monimutkaisemmiksi, myös niiden ratkaisemiseen tarvittavat algoritmit monimutkaistuvat. Tässä vaiheessa tulee tärkeäksi arvioida algoritmin aiheuttamaa kustannusta. Algoritmin tehokkuudella tarkoitetaan sitä, kuinka tehokkaasti algoritmi käyttää tietokoneen resursseja. Tämä ominaisuus on oleellinen algoritmin suorituskyvyn ymmärtämisessä, ja se kattaa resurssien, kuten ajan ja muistin, kulutuksen analysoinnin. Maksimaalinen tehokkuus saavutetaan, kun resurssien käyttö on mahdollisimman vähäistä. Algoritmien tehokkuuden vertailu voi kuitenkin olla haastavaa, koska eri resurssityyppejä ei voida suoraan verrata keskenään.

Voimme pitää algoritmia tehokkaana, kun sen resurssien kulutus – eli laskentakustannus – pysyy hyväksyttävän rajan sisällä tai sen alapuolella. Tämä raja määritellään sen perusteella, kykeneekö algoritmi toimimaan kohtuullisessa ajassa tai tilassa tavallisella tietokoneella suhteessa syötteen kokoon. Käytännössä

"hyväksyttävä" tarkoittaa, että algoritmi pystyy toimimaan todellisen maailman olosuhteissa rajoitusten puitteissa.



Universitat
de les Illes Balears



Algoritmin tehokkuutta mitataan kahdella tavalla:

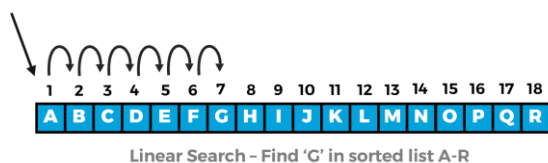
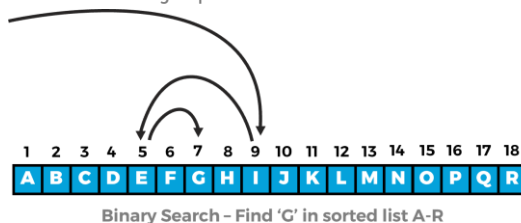
- **Aikavaativuus (Time Complexity):** Tämä kuvaa laskennallista monimutkaisuutta siinä, kuinka kauan algoritmin suorittaminen kestää syötteen koon funktiona.
- **Tilavaativuus (Space Complexity):** Tämä tarkoittaa sitä muistimäärää, jonka algoritmi tarvitsee suorituksensa loppuun viemiseksi.

Tavoitteena on minimoida sekä aika- että muistivaatimukset. Näiden kahden välillä on kuitenkin usein kompromissi: esimerkiksi algoritmi, joka toimii nopeammin, saattaa vaatia enemmän muistia, kun taas vähemmän muistia käyttävä algoritmi saattaa olla hitaampi. Algoritmin valinta riippuu tehtävän erityisvaatimuksista. Tehokkaat algoritmit löytävät tasapainon näiden tekijöiden välillä ja pyrkivät optimaaliseen suorituskäyttöön käytännön tilanteissa.

Tarkastellaan yksinkertaista esimerkkiä kahdesta eri algoritmista, joita voidaan käyttää alkion etsimiseen listasta. Molemmilla on sama tilavaativuus, mutta eri aikavaativuus. Tämän jälkeen analysoimme, milloin kumpakin kannattaa käyttää:

Lineaarinen haku on algoritmi, jota käytetään arvon etsimiseen listasta. Se muistuttaa hyvin paljon aiemmin käsittelemäämme algoritmia suurimman arvon löytämiseksi listasta. Algoritmi käy läpi jokaisen listan alkion alusta alkaen, kunnes se löytää etsittävän arvon tai saavuttaa listan lopun. Jokaisella silmukan kierroksella nykyistä alkioa verrataan etsittävään arvoon. Jos vastaavuus löytyy, algoritmi päättyy ja palauttaa arvon sijainnin (indeksin) listassa.

Binäärihaku on tehokas algoritmi, jota käytetään etsittävän arvon paikantamiseen lajitellusta listasta jakamalla hakuväli toistuvasti kahtia. Haku alkaa tutkimalla listan keskimmäistä alkioa ja vertaamalla sitä etsittävään arvoon. Jos keskimäinen alkio vastaa etsittävää arvoa, haku onnistuu. Mikäli etsittävä arvo on keskimmäistä pienempi, hakua jatketaan listan alemmassa puoliskossa; jos etsittävä on suurempi, siirrytään ylempään puoliskoon. Tätä puolittamista jatketaan, kunnes arvo löytyy tai hakuväli tyhjenee. Binäärihaku perustuu "jaa ja hallitse" -periaatteeseen ja pienentää tehokkaasti hakualuetta jokaisella



iteraatiolla

Source | Board Infinity

Näiden kahden esimerkin avulla ymmärrämme helposti, että samaa tehtävää voidaan suorittaa useilla eri algoritmeilla. Jokaisella algoritmilla on omat ehtonsa, vahvuutensa ja heikkoutensa. Lineaarinen haku on helppo toteuttaa ja sopii hyvin pienille listoille ja lajittelemattomalle datalle. Kuitenkin sen tehokkuus heikkenee listan koon kasvaessa, sillä jos etsittävä alkio on lähellä listan loppua, algoritmin on käytävä läpi suurin osa alkioista. Binäärihaku on erittäin tehokas ja sopii erityisen hyvin suurten ja lajiteltujen tietoaaineistojen etsintään. Sen rajoitteena on kuitenkin se, että data täytyy olla valmiiksi lajiteltu, mikä voi olla edellytys algoritmin käyttämiselle.

2.5. Rajoitukset

Viime vuosien aikana olemme nähneet merkittävän kehitysharppauksen koneoppimisen ja tekoälyn saralla – teknologioiden, jotka nojaavat vahvasti algoritmeihin käsitelläkseen valtavia tietomääriä ja tuottaakseen niistä oivalluksia. Tämä kehitys on edistänyt merkittävästi monia aloja, kuten kuva- ja videontunnistusta ja -generointia, robotiikkaa, lääketieteellistä analyysiä, päätöksentekoa, äänijäljittelyä ja luonnollisen kielen käsittelyä. Nämä edistysaskleet osoittavat algoritmien rajattoman potentiaalin ja vievät meitä kohti ennennäkemättömän innovaation aikakautta. On kuitenkin tärkeää ymmärtää, että näistä merkittävistä edistysaskelista huolimatta algoritmeilla on omat rajoituksensa. On olemassa ongelmia, joita ei voida ratkaista algoritmilla, ja ongelmia, joiden ratkaisemiseen meillä ei vielä ole valmiuksia.

Aloitamme kuvailemalla ongelmat, joita algoritmi ei voi ratkaista. Nämä voidaan jakaa kahteen kategoriaan: Ratkaisemattomat (undecidable) ongelmat ovat sellaisia, joita ei teoreettisesti voida ratkaista millään algoritmilla. Vaikeasti ratkaistavat (intractable) ongelmat ovat ongelmia, joille ei ole olemassa kohtuullisessa ajassa toimivaa ratkaisua.

Klassinen esimerkki ratkaisemattomasta ongelmasta on pysähtymisongelma (halting problem), joka on päätösongelma, johon vastataan kyllä tai ei – mutta jota ei voida ratkaista algoritmisesti. Siinä esitetään haaste: voidaanko määrittää, pysähtyykö jokin tietty tietokoneohjelma lopulta vai jääkö se pyörimään loputtomasti? Ongelma voidaan muotoilla seuraavasti:

"Onko mahdollista kirjoittaa ohjelma, joka pystyy kertomaan – ilman ohjelman suorittamista – pysähtyykö mikä tahansa annettu ohjelma tietyllä syötteellä?"

Vastaus tähän kysymykseen on ei – tällaista universaalia ohjelmaa ei voida rakentaa.

Toisaalta on olemassa vaikeasti ratkaistavia ongelmia (intractable problems). Osa näistä ongelmista voidaan ratkaista kohtuullisessa ajassa vain silloin, kun syöte on pieni – mutta algoritmi ei skaalaudu isompiin syötteisiin. Joitakin vaikeasti ratkaistavia ongelmia kuitenkin ratkaistaan päivittäin, vaikka ne ovat

teoreettisesti vaikeita. Tällaisia ovat esimerkiksi reititysongelmat, aikataulutus ja resurssien hallinta.

Näissä tapauksissa käytetään usein arvioituja ratkaisuja, jotka voidaan tarkistaa nopeasti. Tätä lähestymistapaa kutsutaan heuristiikaksi – se hyödyntää tietoa ja kokemusta älykkäiden arvausten tekemiseen. Heuristinen lähestymistapa ei välttämättä pyri löytämään täydellistä tai ihanneratkaisua, vaan riittävän hyvän ratkaisun, joka toimii käytännössä.

Yksi tunnetuimmista vaikeasti ratkaistavista ongelmista on kauppamatkustajan ongelma (Travelling Salesman Problem, TSP). Tässä ongelmassa tehtävänä on löytää lyhin mahdollinen reitti, jonka avulla kauppamatkustaja voi vierailla jokaisessa kartan kaupungissa täsmälleen kerran, ja palata lopuksi lähtökaupunkiin. Tavoitteena on minimoida kokonaismatkan pituus, mutta ongelma muodostuu laskennallisesti erittäin haastavaksi, koska mahdollisten reittien määrä kasvaa eksponentiaalisesti kaupunkien määrän kasvaessa.

Tehdäänpieni laskuesimerkki havainnollistamaan, kuinka nopeasti tämä kasvu tapahtuu. Kaupunkien välisten reittien (tai mahdollisten järjestysten) määrä saadaan laskemalla (kaupunkien määrä – 1)! (eli kertoma). Tutkittavien reittien määrä on tästä puolet, koska jokainen tie kuljetaan vain kerran, eikä suunnalla ole merkitystä.

Alla olevasta taulukosta nähdään, kuinka nopeasti mahdollisten reittien määrä kasvaa²:

Number of cities	Number of connections between cities	Number of possible routes
5	$4! = 24$	12
10	$9! = 362.880$	181.440
15	$14! = 87.178.291.200$	43.589.145.600

2.6. Koneoppimisalgoritmien rajoitukset

Koneoppimisen (ja syväoppimisen) aikakaudella meillä on käytössämme valtava määrä algoritmeja, jotka on suunniteltu ratkaisemaan monenlaisia tehtäviä, esimerkiksi: päätöspuut, satunnaismetsät (Random Forest), tukivektorikoneet (Support Vector Machines) ja neuroverkot (kaikissa eri muodoissaan)... Näitä kaikkia voidaan mukauttaa eri ongelmiin koulutusprosessin avulla, jossa algoritmillemme näytetään dataa, ja se mukauttaa itseään sen perusteella. Tällä tavalla voimme saada aikaan tarkoituksenmukaisia ratkaisuja. On kuitenkin tärkeää huomioida, että myös näillä algoritmeilla on omat rajoitteensa.

- **Datan laatu ja määrä:** Koneoppimisalgoritmit ovat vahvasti riippuvaisia saatavilla olevan datan laadusta ja määrästä. Rajoittunut tai vinoutunut

² Example taken from <https://runestone.academy/ns/books/published/mobilecsp/Unit5-Algorithms-Procedural-Abstraction/Limits-of-Algorithms.html>

data voi johtaa epätarkkoihin tai puolueellisiin mallien ennusteisiin. Lisäksi riittämätön määrä dataa voi estää algoritmia yleistämästä hyvin uusiin, aiemmin näkemättömiin tilanteisiin.

- **Ylisovittaminen ja alisovittaminen:** Koneoppimismallit voivat kärsiä ylisovittamisesta (overfitting), jolloin ne oppivat ulkoa koulutusdatan kohinaa tai epäolennaisia kuvioita, mikä heikentää suorituskkyä uudella datalla. Toisaalta alisovittaminen (underfitting) tapahtuu, kun malli on liian yksinkertainen eikä pysty mallintamaan datan rakenteita kunnolla, mikä johtaa heikkoon suorituskkyyn.
- **Tulkittavuus:** Monet koneoppimismallit – erityisesti monimutkaiset, kuten neuroverkot – ovat heikosti tulkittavia, mikä vaikeuttaa niiden ennusteiden ymmärtämistä ja luottamusta niihin. Tämä voi olla ongelmallista erityisesti aloilla, joilla tulkittavuus on kriittistä, kuten terveydenhuollossa ja rahoitusallalla.
- **Laskentaresurssit:** Monimutkaisten koneoppimismallien kouluttaminen vaatii usein merkittäviä laskentaresursseja, mukaan lukien tehokasta laitteistoa ja suurteholaskentaan soveltuvia tietojenkäsittelyratkaisuja. Rajoitetut resurssit voivat rajoittaa koneoppimisratkaisujen skaalautuvuutta ja käytännön toteutettavuutta.
- **Toimialakohtainen osaaminen:** Tehokkaiden koneoppimisratkaisujen kehittäminen edellyttää usein toimialakohtaista asiantuntemusta, jotta data voidaan esikäsittää oikein, oleelliset piirteet tunnistaa ja mallin tuottamat tulokset tulkita oikein.
- **Jatkuva oppiminen ja sopeutuminen:** Koneoppimisalgoritmit voivat kohdata vaikeuksia sopeutuessaan muuttuviin olosuhteisiin tai ajan myötä kehittyviin datan jakaumiin. Mallien suorituskkyyn säilyttämiseksi tarvitaan jatkuvaa seurantaa ja uudelleen koulutusta, jotta ne pysyvät luotettavina ja ajan tasalla.

VINKKEJÄ JA SUOSITUKSIA OPETTAJILLE

Vinkkejä ja suosituksia opettajille algoritmien monimutkaisuuden käsitteiden käsittelyyn

1. Rakenna vahva perusta

Aloita algoritmien perusteiden ja niiden historiallisen kehityksen yleiskatsauksella, jotta osallistujat saavat kontekstin. Tämä auttaa luomaan kertomuksen, joka tekee abstraktista käsitteestä konkreettisemmän ja helpommin lähestyttävän.

2. Käytä vertauksia ja metaforia

Vertaile algoritmeja arkisiin toimintoihin, kuten ruoanlaittoon tai kaupungissa suunnistamiseen. Tämä auttaa poistamaan mystiikkaa käsitteen ympäriltä ja

tekee siitä helpommin ymmärrettävän osallistujille, joilla ei ole vahvaa teknistä taustaa.

3. Interaktiiviset demonstraatiot

Näytä algoritmien toimintaa käytännössä esimerkiksi live-koodauksen tai interaktiivisten verkkotyökalujen avulla, joissa osallistujat voivat syöttää esimerkkejä ja nähdä, miten algoritmi niitä käsittelee. Tämä voi olla erittäin innostavaa ja opettavaista.

4. Edistä vuorovaikutteisuutta opetustilanteessa

Kannusta kysymyksiin ja keskusteluun, jotta osallistujat voivat ilmaista ymmärrystään tai epäselvyyksiään käsiteltävistä aiheista. Tämä auttaa myös seuraamaan osallistujien sitoutumista ja ymmärryksen tasoa reaaliajassa.

5. Oppiminen vaiheittain

Pilko monimutkaiset aiheet pienempiin ja hallittavampiin osiin. Aloita peruskäsitteistä ja lisää vähitellen monimutkaisuutta. Tämä "rakennuspalikkamalli" auttaa välttämään kognitiivista ylikuormitusta.

Alkulämmittelytehtävät

1. Yksinkertaisen algoritmin laatiminen

Pyydä osallistujia kirjoittamaan yksinkertainen algoritmi arkipäiväisestä tehtävästä, kuten teen keittämisestä tai töihin valmistautumisesta. Tämä aktiviteetti saa heidät ajattelemaan loogisia, vaihteellaisia toimintaketjuja, jotka ovat algoritmisuunnittelun perusta.

2. Algoritmien historian tarkastelu

Keskustelkaa ryhmässä Eukleideen algoritmista, tutkien sen vaiheet ja historiallinen käyttö ongelmien ratkaisussa. Tämä voi syventää osallistujien ymmärrystä siitä, kuinka perustavanlaatuiset algoritmikonseptit ovat syntyneet jo kauan sitten.

3. Ohjausrakenteiden roolileikki

Järjestä roolileikki, jossa kukin osallistuja esittää osaa algoritmin ohjausrakenteesta, kuten ehtoa tai silmukkaa. Tämä auttaa heitä hahmottamaan ja ymmärtämään, miten eri osat algoritmista toimivat yhdessä.

4. Algoritmin virheiden etsintä -haaste

Anna osallistujille yksinkertainen algoritmi, jossa on tarkoituksella virheitä, ja pyydä heitä "debuggaamaan" se. Tämä on hauska ja osallistava tapa syventää ymmärrystä algoritmien rakenteesta ja toiminnasta.

Arviointivinkit ja -suositukset

1. Jatkuva palaute

Anna välitöntä ja jatkuvaa palautetta kurssin aikana. Tämä auttaa osallistujia korjaamaan väärinymmärryksiä ajoissa ja vahvistaa oppimista kurssin edetessä.

2. Käytännön tehtävät

Suunnittele tehtäviä, joissa osallistujat voivat soveltaa oppimaansa käytännön tilanteissa. Esimerkiksi he voivat optimoida olemassa olevan algoritmin tai tunnistaa tehottomuuksia annetusta algoritmista.

3. Vertaisarviointi

Sisällytä vertaisarviointikertoja, joissa osallistujat arvioivat toistensa työtä. Tämä tarjoaa useita palautekanavia ja edistää yhteistyöhön perustuvaa oppimista ja kriittistä ajattelua.

4. Arviointikriteerit (rubriikit)

Laadi selkeät, kriteeriperustaiset arviointitaulukot tehtävien arviointiin. Tämä takaa arvioinnin yhdenmukaisuuden ja tekee osallistujien odotuksista selkeitä.

5. Itsearviointi

Kannusta osallistujia arvioimaan omaa työtään annettujen kriteerien perusteella. Tämä edistää itsereflektiota ja syvempää oppimista.

3. Algoritmien roolin tunnistaminen eri toimialoilla sekä niiden käytöstä mahdollisesti aiheutuvat sudenkuopat

Algoritmien roolin syvälinen ymmärtäminen eri toimialoilla sekä samanaikainen tietoisuus mahdollisista sudenkuopista on keskeistä, kun pyritään muodostamaan kattava ja kokonaisvaltainen käsitys algoritmien käyttöönottoon liittyvistä vaikutuksista ja vastuista. Tämä ajatus korostaa perusteellisen ymmärryksen tarvetta siitä, miten algoritmit voivat vaikuttaa yhteiskunnan, talouden ja hallinnon eri osa-alueisiin. Algoritmien monipuolinen käyttö — aina toiminnan tehostamisesta käyttäjäkokemusten personointiin — osoittaa niiden merkittävän vaikutuksen nykypäivän elämään (Mourtzis ym., 2022).

Tämä vaikutus tuo kuitenkin mukanaan myös merkittäviä riskejä, kuten mahdollisuuden puolueellisuuteen, yksityisyydensuojan loukkauksiin sekä olemassa olevien yhteiskunnallisten eriarvoisuuksien vahvistumiseen (Patel, 2024).

Algoritmien käyttöönottoa tulee ohjata eettisillä periaatteilla ja sääntelykehyksillä näiden riskien lieventämiseksi (Tsamados ym., 2021). On ratkaisevan tärkeää kehittää vahvat vastuullisuuden rakenteet, jotka varmistavat, etteivät algoritmit ylläpidä tai pahenna haitallisia käytäntöjä tai eriarvoisuuksia.

Tähän kuuluu perusteellinen testaus- ja validointiprosessi, jossa algoritmeja arvioidaan oikeudenmukaisuuden ja tarkkuuden näkökulmasta – erityisesti herkissä sovellusalueissa, kuten terveydenhuollossa, rikosoikeudessa ja rahoituspalveluissa. Sidosryhmien – mukaan lukien teknologia-asiantuntijat, päättäjät ja kansalaiset – tulee osallistua jatkuvaan vuoropuisteluun algoritmien päätösten vaikutuksista varmistaakseen, että nämä työkalut palvelevat koko yhteiskunnan etua (Donovan ym., 2018).

Lisäksi algoritmien toiminnan ja päätöksenteon läpinäkyvyys on olennaista. Yleisölle on tarjottava varmuus siitä, että algoritmiset päätökset tehdään eettisiä periaatteita noudattaen ja että ne ovat avoimia tarkastelulle. Tämä läpinäkyvyys on ratkaisevaa paitsi luottamuksen rakentamiseksi, myös siksi, että asiantuntijat ja valvontaviranomaiset voivat suorittaa tarkastuksia ja arviointeja tehokkaasti (Felzmann ym., 2019).

Tarvitaan myös koulutuksellisia aloitteita, jotka parantavat algoritmista lukutaitoa niin kansalaisten kuin päättäjien keskuudessa (Reisdorf & Blank, 2021). Tällaiset toimet mahdollistavat sen, että yhä useampi voi osallistua keskusteluun ja tehdä perusteltuja päätöksiä näiden teknologioiden käytöstä ja sääntelystä.

Algoritmisen oikeudenmukaisuuden ja etiikan tutkimuksen ja kehityksen on jatkuttava. Teknologian kehittyessä ilmenee uusia haasteita, jotka vaativat innovatiivisia ratkaisuja niin teknisiin kuin eettisiin kysymyksiin.

Yhteistyö akateemisen maailman, teollisuuden ja hallinnon välillä voi edistää parhaiden käytäntöjen jakamista ja tukea oikeudenmukaisempien ja vastuullisempien algoritmien kehittämistä.

Keskusteluun algoritmeista on myös sisällytettävä niiden positiivinen potentiaali. Kun algoritmit suunnitellaan ja toteutetaan vastuullisesti, niillä on valta parantaa saavutettavuutta, tehokkuutta ja oikeudenmukaisuutta. Esimerkiksi koulutuksessa algoritmit voivat mahdollistaa yksilöllisiä oppimiskokemuksia, jotka mukautuvat opiskelijan omiin tarpeisiin – tämä voi auttaa kaventamaan oppimiserot. Terveystieteiden huollossa ennustavat algoritmit voivat parantaa diagnostiikkaa ja potilastuloksia tunnistamalla riskejä ja suositellessaan yksilöllisiä hoitosuunnitelmia.

Yhteenvetona: algoritmien käyttöönotto eri toimialoilla tarjoaa sekä mahdollisuuksia että haasteita (Dwivedi ym., 2021). Edistämällä eettistä toimintakulttuuria, läpinäkyvyyttä ja osallisuutta, yhteiskunta voi hyödyntää algoritmien tarjoamat edut ja samalla minimoida niihin liittyvät riskit. Kuten Williamson (2018) toteaa, näiden dynamiikkojen monipuolinen ymmärrys on olennaista, jotta algoritmien sovelluksia voidaan kehittää tavalla, joka hyödyttää oikeudenmukaisesti kaikkia yhteiskunnan jäseniä.

3.1. Algoritmien rooli eri toimialoilla

Kuten Parker ym. (2016) ovat todenneet: "algoritmit ovat läpäisseet jokaisen toimialan, mullistaen prosesseja ja päätöksentekoa." Tässä osiossa tarkastelemme joitakin keskeisiä aloja, joilla algoritmien vaikutus on erityisen merkittävä:

a. Rahoitusala: Tekoälyn käyttö rahoitusalaalla on kehittynyt merkittävästi yksinkertaisista alkuvaiheista nykytilanteeseen, jossa siitä on tullut alan korvaamaton työkalu. Tekoälyn juuret rahoituksessa ulottuvat 1980-luvulle, jolloin perusohjelmia hyödynnettiin yksinkertaisten tehtävien, kuten tiedonsyötön ja kirjanpidon, automatisointiin.

Todellinen murros tapahtui kuitenkin 1990-luvun lopulla ja 2000-luvun alussa, kun käyttöön otettiin kehittyneempiä koneoppimisalgoritmeja ja datan määrä alkoi kasvaa räjähdysmäisesti (Aksoy & Gurol, 2021). Rahoituslaitokset havaitsivat nopeasti tekoälyn mahdollisuudet tarkempien riskianalyyysien tuottamisessa, nopeammassa maksuliikenteen käsittelyssä ja asiakaspalvelun parantamisessa. Kehitystä kiihdyttivät myös laskentatehon kasvu ja kehittyneiden data-analytiikkamenetelmien käyttöönotto. 2010-luvulle mennessä tekoäly oli alkanut muovata kaupankäyntiä, vakuutusliiketoimintaa, petosten tunnistamista ja asiakassuhteiden hallintaa syvällisellä tavalla (Davenport & Mittal, 2023).

Tässä muutamia esimerkkejä tekoälyn käytöstä rahoitusalla:

- **Algoritminen kaupankäynti:** Yksi tunnetuimmista esimerkeistä algoritmisessa kaupankäynnissä on Renaissance Technologies – hedge-rahasto, joka tunnetaan Medallion Fund -rahastostaan. Tämä rahasto käyttää monimutkaisia matemaattisia malleja analysoimaan markkinoita ja toteuttamaan kauppvoja erittäin suurella nopeudella. Kehittyneiden algoritmien ansiosta Renaissance on onnistunut saavuttamaan erinomaisia tuottoja, selvästi markkinoiden keskiarvoa parempia (Qin, 2012).
- **Riskienhallinta:** JPMorgan Chase hyödyntää algoritmeja luottoriskien hallinnassa. Heidän järjestelmänsä arvioi lainojen takaisinmaksun riskiä useiden tekijöiden perusteella, kuten taloudelliset olosuhteet, toimialan kehitys ja asiakkaan luottohistoria. Tämä algoritminen lähestymistapa mahdollistaa räätälöityjen lainatarjousten tekemisen sekä luottotappioiden minimoimisen.
- **Petosten tunnistus:** PayPal hyödyntää koneoppimisalgoritmeja petollisten maksutapahtumien tunnistamiseen. Nämä algoritmit analysoivat tuhansia maksutapahtuman ominaisuuksia reaaliajassa, kuten summan, käytetyn laitteen ja käyttäjän maksuhistorian. Tunnistamalla poikkeamia normaalista käyttäytymisestä PayPalin järjestelmät pystyvät tunnistamaan mahdolliset petokset erittäin tarkasti, mikä auttaa estämään taloudellisia menetyksiä ja suojaamaan käyttäjiä.
- **Asiakaspalvelu:** Bank of America käyttää tekoälypohjaista virtuaaliassistenttia nimeltä Erica. Tämä algoritminen työkalu auttaa asiakkaita hallinnoimaan tilejään, seuraamaan kulutustaan ja tekemään maksuja. Erica voi myös antaa reaaliaikaisia ilmoituksia tilimuutoksista ja ehdottaa säästötapoja käyttäjän kulutustottumuksiin perustuen, joita algoritmi analysoi.

b. Terveysthuolto: Tekoälyn kehitys terveydenhuollossa on ollut huomattavaa – varhaisista kokeellisista sovelluksista siihen, että siitä on tullut olennainen osa nykyaikaista lääketieteellistä käytäntöä. Tämä muutos on pitkälti seurausta koneoppimisen, big datan analytiikan ja terveydenhuollon tietojen digitalisaation kehityksestä. Syvennyessämme aiheeseen tarkastelemme tekoälyn eri käyttökohteita terveydenhuollossa, kuten diagnostiikan tukemista, yksilöllistä

hoitoa, sairauksien ennustamista, potilashallintaa ja robottikirurgiaa (Bohr & Memarzadeh, 2020).Tässä muutamia merkittäviä esimerkkejä:

- **Diagnostiikan tukeminen:** Tekoälyn kyky analysoida suuria tietomääriä on mullistanut terveydenhuollon diagnostiikkaprosessit. Kun tekoäly yhdistetään kuvantamisteknologioihin, voivat lääketieteen ammattilaiset tunnistaa sairauksia tarkemmin ja nopeammin kuin perinteisin menetelmin. IBM Watson Health on esimerkki tekoälyn voimasta diagnostiikan tarkkuuden parantamisessa. Watsonin kehittyneet tekoälyalgoritmit pystyvät analysoimaan sekä rakenteellista että rakenteetonta dataa potilastiedoista ja kliinisistä raporteista. Esimerkiksi onkologiassa Watsonia on käytetty löytämään hoitovaihtoehtoja syöpäpotilaille vertaamalla miljoonia onkologisia potilasmuistiinpanoja potilaan omiin tietoihin ja lääketieteelliseen kirjallisuuteen
- **Yksilöllistetty hoito:** Tekoäly mahdollistaa yksilöllisen lääketieteen hyödyntämisen ottamalla huomioon potilaan geneettisen profiilin, ympäristötekijät ja elämäntavat hoitojen räätälöimiseksi. Tämä lähestymistapa parantaa merkittävästi hoitotuloksia, kun potilaille voidaan tarjota juuri heille todennäköisimmin toimivia hoitomuotoja. Tempus käyttää tekoälyä kerätäkseen ja analysoidakseen valtavia määriä genomi- ja potilastietoa auttaakseen lääkäreitä laatimaan yksilöllisiä hoitosuunnitelmia syöpäpotilaille. Alusta hyödyntää koneoppimisalgoritmeja molekyyli- ja hoitotiedon ymmärtämiseksi ja ennustaa, mitkä hoidot ovat todennäköisesti tehokkaimpia kullekin potilaalle.
- **Sairauksien ennakointi ja hallinta:** Tekoälymalleja käytetään yhä enemmän ennustamaan sairauksien kehittymistä, etenemistä ja mahdollisia komplikaatioita. Tämä ennakoiva lähestymistapa mahdollistaa aikaisemmat toimenpiteet, mikä voi pelastaa henkiä ja vähentää terveydenhuollon kustannuksia. Googlen DeepMind on kehittänyt tekoälyjärjestelmiä, jotka pystyvät ennustamaan akuutin munuaisvaurion jopa 48 tuntia ennen sen puhkeamista erittäin tarkasti. Malli analysoi laajasti terveystietoa reaaliajassa, mikä mahdollistaa ajoissa tehtävät toimenpiteet ja parantaa merkittävästi potilaan ennustetta.
- **Potilashallinta ja seuranta:** Tekoäly tehostaa potilashallintaa seuraamalla jatkuvasti terveysdataa esimerkiksi puettavan teknologian ja IoT-laitteiden avulla. Nämä välineet voivat varoittaa terveydenhuollon ammattilaisia potilaan tilan muutoksista, mikä mahdollistaa välittömän reagoinnin ja hoidon jatkuvan mukauttamisen. Virta Health hyödyntää tekoälypohjaisia työkaluja kroonisten sairauksien, kuten tyypin 2 diabeteksen, hallinnassa. Sen alusta seuraa potilaan terveystietoja reaaliajassa ja mukauttaa hoitosuunnitelmia dynaamisesti, mikä auttaa pitämään verensokeritasot optimaalisina ja vähentää lääkkeiden tarvetta.
- **Robottikirurgia:** Tekoälyn avustama robottikirurgia tarjoaa korkeaa tarkkuutta, vähemmän kudosaivourioita ja nopeampaa toipumista. Tekoälyalgoritmit tarjoavat kirurgeille reaaliaikaista tietoa ja voivat jopa automatisoida tiettyjä toimenpiteen osia potilasturvallisuuden ja

lopputuloksen parantamiseksi. Intuitive Surgicalin kehittämä da Vinci - robottikirurginen järjestelmä hyödyntää tekoälyä kirurgisen tarkkuuden parantamiseen. Järjestelmä tarjoaa kirurgeille erittäin suurennettuja, kolmiulotteisia HD-kuvia leikkausalueesta ja muuntaa kirurgin kädenliikkeet tarkemmiksi, pienemmiksi liikkeiksi, joita suorittavat miniatyriset instrumentit potilaan sisällä.

Tekoälyn kehittyessä sen potentiaali muuttaa terveydenhuoltoa kasvaa eksponentiaalisesti. Nämä esimerkit korostavat tekoälyn kykyä parantaa diagnoosien tarkkuutta, räätälöidä hoidot yksilöllisesti, ennustaa sairauksien kulkua, hallita kroonisia sairauksia tehokkaasti sekä suorittaa leikkauksia ennennäkemättömällä tarkkuudella. Tekoälyn integroiminen terveydenhuoltoon ei ainoastaan optimoi potilaiden hoitotuloksia, vaan myös tehostaa terveydenhuollon ammattilaisten työtä, avaten tien entistä innovatiivisemmille lähestymistavoille lääketieteessä ja kirurgiassa. Tekoälyteknologioiden jatkuva kehittäminen ja käyttöönotto tarjoavat merkittäviä mahdollisuuksia vastata nykyterveydenhuollon suurimpiin haasteisiin.

c. Verkkokauppa: Tekoäly parantaa merkittävästi sekä asiakkaiden että yritysten kokemuksia – erityisesti vähittäiskaupassa ja verkkokaupassa. Laajan kuluttajadatan keräämisen ja sen yhdistämisen avulla koneoppimisalgoritmeihin jälleenmyyjät voivat kehittää kehittyneitä personoinnin, suosittelun ja automaation toimintoja. Näistä tekoälyyn perustuvista ominaisuuksista on tullut vakiotoimintoja ostosalustoilla, ja ne lisäävät asiakasosallistumista sekä tehostavat liiketoimintaa, mikä parantaa yritysten kannattavuutta ja resurssien käyttöä. Tässä muutamia esimerkkejä tekoälyn käytöstä verkkokaupassa ja vähittäiskaupassa:

- **Mainonnan strateginen tehostaminen**

Smartly.io: hyödyntää tekoälyä sosiaalisen median mainoskampanjoiden automaattiseen hallintaan ja optimointiin, mikä vähentää manuaalista työtä ja parantaa mainosten suorituskykyä ja sitoutumista eri alustoilla.

eBay: käyttää tekoälyä personoitujen ostosuositusten ja asiakasneuvonnan tarjoamiseen, mikä lisää ostajien tyytyväisyyttä ja sitoutuneisuutta.

- **Autokaupan optimointi**

Cox Automotive: hyödyntää tekoälyä Esntial-työkalunsa kautta auton ostoprosessin digitalisointiin ja sujuvoittamiseen, sisältäen maksuarvion ja riskinarvion laskennan, parantaen asiakkaan digitaalisia ostokokemuksia.

- **Tehokkaat toimitusjärjestelmät**

Gopuff: integroi tekoälyn toimitusreittien optimointiin, varmistaen nopean ja tehokkaan päivittäistuotteiden toimituksen mikrologistiikkakeskustensa kautta.

- **Innovatiivinen tuotekehitys**

Mondelez International: hyödyntää tekoälyä tutkimus- ja kehitystoiminnassaan kiihdyttääkseen innovaatiota ja

parantaakseen kustannustehokkuutta välipalatuotteiden valmistuksessa.

- **Personoidut ostokokemukset**

Mirakl: käyttää tekoälyä tuotesuosittelun räätälöintiin, parantaen asiakasosallistumista ja myyntiä digitaalisilla markkinapaikoilla.

Rue Gilt Groupe: soveltaa tekoälyä tuotesuosittelun tarkentamiseen, parantaen ostokokemusta ylellisyystuotteisiin keskittyvillä verkkosivuillaan.

- **Kehittynyt varastohallinta ja toimitusketjun optimointi**

IBM Watson: tukee vähittäiskaupan yrityksiä toiminnan tehostamisessa ja asiakas kohtaamisten personoinnissa reaaliaikaisen data-analytiikan avulla.

Anaplan: hyödyntää ennakoivaa analytiikkaa liiketoiminnan tulosten ennustamiseen ja vähittäiskaupan strategioiden optimointiin, vaikuttaen muun muassa varastohallintaan ja asiakashankintaan.

- **Väärentämisen tunnistus ja sisällön moderointi**

3PM Solutions: käyttää tekoälyä väärennetyjen tuotteiden tunnistamiseen ja myyjien arvioiden tarkkuuden varmistamiseen suurilla verkkokauppa-alustoilla.

- **Keskustelevat tekoälyratkaisut ja asiakaskokemus**

Valyant AI: kehittää tekoälypohjaisia keskustelualustoja pikaruokaravintoloille, parantaen asiakastilauksia puheohjatun teknologian avulla.

d. Liikenne: Ei ole epäilystäkään siitä, että tekoäly muuttaa liikennealaa parantamalla tehokkuutta, turvallisuutta ja käyttäjäkokemuksia. Tässä osiossa tarkastellaan, miten tekoälyä integroidaan eri liikennejärjestelmiin – julkisesta liikenteestä ja logistiikasta henkilökohtaiseen liikkumiseen ja liikenteen hallintaan.

- **Autonomiset ajoneuvot**

Itseajavat autot: Tekoäly ohjaa autonomisten ajoneuvojen ydintoimintoja mahdollistaen niiden suunnistuksen, esteiden väistämisen ja reaaliaikaisen päätöksenteon. Yritykset kuten Tesla ja Waymo ovat tämän alan edelläkävijöitä, ja ne ovat merkittävästi edistäneet liikenneturvallisuutta ja ajoneuvojen tehokkuutta.

Toimitusdronet: Yritykset kuten Amazon kokeilevat tekoälyllä ohjattuja droneja, jotka pystyvät toimittamaan paketteja itsenäisesti, minkä odotetaan lyhentävän toimitusaikoja ja vähentävän viimeisen kilometrin logistiikkakustannuksia.

- **Liikenteen hallinta ja älykaupungit**

Älykkäät liikennejärjestelmät: Tekoälyä käytetään liikennemallien analysointiin ja liikennevalojen ajoituksen optimointiin, mikä vähentää ruuhkia ja parantaa liikenneturvallisuutta. Esimerkiksi Barcelona ja

Singapore hyödyntävät tekoälyjärjestelmiä varmistaakseen sujuvan liikennevirran vilkkaissa risteyksissä.

Ennakoiva kunnossapito: Tekoäly auttaa ennustamaan, milloin joukkoliikenteen ajoneuvot ja infrastruktuuri (kuten sillat ja tiet) tarvitsevat huoltoa, mikä ehkäisee toimintahäiriöitä ja pidentää julkisten rakenteiden elinkaarta.

- **Joukkoliikenteen optimointi**

Reittisuunnittelu: Tekoälyalgoritmit optimoivat bussi- ja junavuorot hyödyntämällä reaaliaikaista dataa ja historiallisia liikennetrendejä, mikä maksimoi operatiivisen tehokkuuden ja minimoi matkustajien odotusajat.

Kapasiteetin hallinta: Ruuhka-aikoina tekoälyjärjestelmät voivat ennustaa matkustajamääriä ja säätää liikennöintiä sen mukaan, mikä parantaa työmatkaliikenteen sujuvuutta ja mukavuutta.

- **Rahti ja logistiikka**

Automaattiset varastot: Yritykset kuten Amazon käyttävät tekoälyohjattuja robotteja tehostamaan varastotoimintoja, mikä nopeuttaa ja tarkentaa tuotteiden keräys- ja pakkausprosesseja.

Optimoitu reititys: Tekoälytyökalut analysoivat lukuisia muuttujia, kuten sääolosuhteet, liikennetilanteen ja toimitusaikataulut, ja ehdottavat tehokkaimpia reittejä kuljetuksille, säästäen aikaa ja polttoainekustannuksia.

- **Parannetut turvallisuusominaisuudet**

Törmäyksenestojärjestelmät: Tekoälyllä toimivat sensorit ja ajoneuvojen sisäiset järjestelmät voivat tunnistaa mahdolliset vaaratilanteet ja reagoida välittömästi onnettomuuksien estämiseksi.

Kuljettajien valvontajärjestelmät: Nämä järjestelmät käyttävät tekoälyä kuljettajan vireystilan ja terveydentilan seuraamiseen, ehkäisten onnettomuuksia, jotka johtuvat väsymyksestä tai äkillisistä sairaskohtauksista.

- **Asiakaspalvelu ja tuki**

Tekoälychatbotit: Liikennealan yritykset hyödyntävät tekoälychatbotteja tarjoamaan reaaliaikaista asiakaspalvelua matkustajille, käsitellen esimerkiksi aikatauluja, lippuhintoja ja häiriöitä koskevia kysymyksiä ilman ihmistä.

Henkilökohtaiset matkasuosituks: Tekoälyalgoritmit analysoivat käyttäjien mieltymyksiä ja aiempaa käyttäytymistä tarjotakseen yksilöllisiä matkasuosituksia, mikä parantaa asiakaskokemusta.

e. Koulutus: Tekoäly muuttaa koulutuksen kenttää mahdollistamalla yksilölliset oppimiskokemukset, automatisoimalla hallinnollisia tehtäviä ja tukemalla elämyksellisiä oppimisympäristöjä. Tässä osiossa tarkastellaan tekoälyn monipuolisia sovelluksia koulutuksessa ja sitä, miten se tukee opettajia, parantaa oppimista ja tehostaa koulutuksen hallintoa.

- **Yksilöllinen oppiminen**

Mukautuvat oppimisalustat: Tekoälyjärjestelmät, kuten DreamBox Learning ja Knewton, tarjoavat personoituja oppimiskokemuksia mukautumalla kunkin oppijan tahtiin ja oppimistyyliin. Nämä alustat arvioivat oppilaan osaamistasoa ja muokkaavat sisältöä automaattisesti hänen yksilöllisten tarpeidensa mukaisesti.

- **Tekoälyopettajat ja -assistentit:** Tekoälypohjaiset tutorointijärjestelmät, kuten Carnegie Learning, tarjoavat oppilaille reaaliaikaista palautetta ja tukea, auttaen heitä ymmärtämään monimutkaisia aiheita yksilöllisen ohjauksen ja harjoittelun avulla.

- **Oppimateriaalien kehittäminen**

Sisällön automaattinen tuottaminen: Tekoälytyökalut voivat auttaa opetussisällön luomisessa ja räätälöinnissä. Esimerkiksi Quillionz käyttää tekoälyä luodakseen kokeita ja yhteenvetoja oppikirjoista ja artikkeleista, rikastuttaen opiskelumateriaaleja ilman suurta manuaalista työmäärää.

Kieltenoppimissovellukset: Sovellukset kuten Duolingo hyödyntävät tekoälyä räätälöidäkseen kielenopetusta käyttäjän taitotason ja edistymisen perusteella, tehden kielen oppimisesta helpommin lähestyttävää ja tehokkaampaa.

- **Arviointi ja palaute**

Automaattiset arviointijärjestelmät: Tekoäly automatisoi standardoitujen kokeiden ja jopa esseiden arvioinnin, vähentäen opettajien hallinnollista kuormaa ja vapauttaen aikaa varsinaiseen opetustyöhön.

Ennakoiva analytiikka: Tekoälyjärjestelmät analysoivat opiskelijadataa ennustaakseen akateemisia riskejä ja oppimistuloksia, mahdollistaen opettajille varhaisen puuttumisen riskiryhmään kuuluvien opiskelijoiden tukemiseksi ja heidän oppimistulostensa parantamiseksi.

- **Hallinnollinen automaatio**

Ilmoittautumis- ja hakuprosessit: Tekoäly tehostaa hallinnollisia tehtäviä, kuten opiskelijahakemusten ja ilmoittautumisten käsittelyä, automatisoimalla tiedonkäsittelyä ja päätöksentekoa. Tämä vähentää paperityötä ja parantaa tarkkuutta.

Resurssien hallinta: Tekoälytyökalut auttavat koulun resurssien, aikataulujen ja läsnäolotietojen hallinnassa, optimoiden toimintaa ja vähentäen yleiskustannuksia.

- **Parannettu saavutettavuus**

Avustavat teknologiat: Tekoälypohjaiset työkalut, kuten puheentunnistus ja tekstinlukijat, tekevät oppimateriaalista helpommin saavutettavaa vammaisille opiskelijoille, varmistaen osallisuuden ja yhdenvertaiset mahdollisuudet kaikille oppijoille.

Visuaaliset ja auditiiviset oppimisen apuvälineet: Tekoälyn ohjaamat sovellukset, kuten Microsoftin Immersive Reader, auttavat esimerkiksi lukivaikeuksista kärsiviä opiskelijoita tarjoamalla lukemisen tukea, osoittaen kuinka teknologia voi tasata oppimisen eroja.

- **Elämykselliset oppimisympäristöt**

Virtuaalitodellisuus (VR) ja lisätty todellisuus (AR): Tekoälyn integrointi VR- ja AR-teknologioihin mahdollistaa elämyksellisiä oppimiskokemuksia, jotka simuloivat todellisia tilanteita. Tämä tekee esimerkiksi luonnontieteiden ja historian kaltaisista monimutkaisista aiheista kiinnostavampia ja helpommin ymmärrettäviä.

Opetuspohjaiset pelit ja simulaatiot: Tekoälyä hyödyntävät opetuspelit mukautuvat oppilaan vastauksiin, tarjoten dynaamisen oppimisympäristön, joka edistää aktiivista osallistumista ja syventää oppimista.

f. Turvallisuus- ja puolustussektori: Tekoälyllä on yhä keskeisempi rooli turvallisuuden ja sotilasoperaatioiden kehittämisessä. Tässä osiossa tarkastellaan, miten tekoälyteknologioita hyödynnetään puolustusstrategioissa, valvontajärjestelmissä ja turvallisuusprotokollissa, parantaen tehokkuutta ja vaikuttavuutta sekä vastaten monimutkaisiin kansallisiin ja globaaleihin turvallisuushaasteisiin.

- **Tehostettu valvonta ja seuranta**

Automaattiset valvontajärjestelmät: Tekoälypohjaisia järjestelmiä käytetään herkkien alueiden valvontaan, joissa ne analysoivat dataa reaaliajassa ja tunnistavat poikkeavaa toimintaa tai uhkia. Nämä järjestelmät osaavat erottaa normaalin ja epäilyttävän käyttäytymisen toisistaan, vähentäen merkittävästi inhimillisiä virheitä ja parantaen reagointinopeutta.

Drone-valvonta: Tekoälyohjatut dronet toimivat ilmatiedustelussa, tarjoten kattavaa valvontaa vaikeapääsyisillä alueilla tai konfliktivyöhykkeillä ilman, että ihmishenkiä altistetaan vaaraan.

- **Kyberturvallisuus ja puolustus**

Uhkatunnistus ja reagointi: Tekoälyjärjestelmät analysoivat verkkoliikenteen malleja tunnistakseen mahdollisia kyberuhkia, kuten haittaohjelmia tai kiristysohjelmia, huomattavasti nopeammin kuin ihmisoperaattorit. Kun uhka havaitaan, tekoäly voi myös automaattisesti käynnistää vastatoimia tai ehdottaa lieventämisstrategioita.

Tietojen salaus: Tekoäly parantaa salaustekniikoita, tehden luvattomien osapuolten entistä vaikeammaksi purkaa arkaluontoista tietoa, ja siten turvaa tiedonsiirron esimerkiksi sotilas- ja turvallisuusverkoissa.

- **Autonomiset puolustusjärjestelmät**

Robottitaisteluyksiköt: Tekoälyä käytetään ohjaamaan autonomiaa tai puoliautonomiaa robottiyksiköitä, jotka voivat suorittaa monenlaisia

sotilaallisia tehtäviä tiedustelusta taistelutoimintaan, vähentäen tarvetta lähettää ihmisiä vaarallisille operaatioalueille.

Ohjuspuolustusjärjestelmät: Tekoälyalgoritmit auttavat ennustamaan ja torjumaan lähestyviä uhkia suurella tarkkuudella, kuten ohjuksia tai miehittämättömiä lennokkeja (UAV), varmistaen ennakoivan puolustusasenteen.

- **Tiedusteluanalyysi ja päätöksenteon tuki**

Datan yhdistäminen ja analysointi: Tekoälyjärjestelmät yhdistävät ja analysoivat valtavia määriä tiedustelutietoa eri lähteistä, tarjoten kokonaisvaltaisen näkymän, joka auttaa sotilassuunnittelijoita tekemään nopeita ja tarkkoja päätöksiä.

Simulaatiot ja koulutus: Tekoälyohjatut simulaatiot ja virtuaalitodellisuusympäristöt (VR) tarjoavat realistista koulutusta turvallisuushenkilöstölle ja sotilaille, valmistellen heitä monenlaisiin tilanteisiin ilman elävän taistelun riskejä.

- **Huolto ja logistiikka**

Ennakoiva huolto: Tekoälytyökalut ennustavat, milloin sotilaskalusto tarvitsee huoltoa, mikä auttaa ehkäisemään vikoja ja pidentämään arvokkaiden varusteiden käyttöikää.

Toimitusketjun optimointi: Tekoäly optimoi logistiikan ja toimitusketjun hallinnan, varmistaen, että resurssit jaetaan ja toimitetaan tehokkaasti oikeaan paikkaan, erityisesti monimutkaisissa sotilasoperaatioissa.

g. Media ja vapaa-aika: Tekoälystä on tullut muutosta ajava voima media- ja viihdeteollisuudessa, tehostaen sisällöntuotantoa, personoiden käyttökokemuksia ja optimoiden toimintoja. Tässä osiossa tarkastellaan, kuinka tekoälyteknologiat integroituvat mediatuotannon, jakelun ja vapaa-ajan eri osa-alueisiin.

- **Sisällöntuotanto ja hallinta**

Automaattinen journalismi: Tekoälytyökalut, kuten luonnollisen kielen tuottamiseen (NLG) perustuvat ohjelmistot, luovat uutisartikkeleita ja raportteja, erityisesti dataan perustuvista aiheista kuten urheilutuloksista ja talouspäivityksistä. Tämä mahdollistaa nopeamman julkaisun ja vapauttaa ihmistoimittajia syvällisiin tutkimuksiin.

Video- ja musiikkituotanto: Tekoälyalgoritmit auttavat videon editoinnissa (esim. värinkorjauksessa) ja musiikin miksaamisessa, mikä tehostaa tuotantoprosesseja ja mahdollistaa luovemman kokeilun.

- **Personointi ja suosittelujärjestelmät**

Median suositukset: suoratoistopalvelut kuten Netflix ja Spotify käyttävät tekoälyä analysoidakseen katselu- ja kuuntelutottumuksia tarjotakseen käyttäjille personoituja sisältösuosituksia, joiden avulla pyritään lisäämään käyttäjien sitoutumista ja parantamaan tyytyväisyyttä.

Kohdennettu mainonta: Tekoälyn ohjaama käyttäjädatan analysointi auttaa mediayhtiöitä luomaan ja toimittamaan kohdennettuja mainoksia, parantaen markkinointikampanjoiden tehokkuutta ja vaikuttavuutta.

- **Yleisön sitouttaminen ja interaktiivinen media**

Chatbotit ja virtuaaliassistentit: Tekoälypohjaiset chatbotit tarjoavat reaaliaikaista vuorovaikutusta käyttäjien kanssa, tarjoten asiakastukea, sisältösuosituksia ja interaktiivisia kokemuksia esimerkiksi peleissä ja virtuaalitodellisuusympäristöissä (VR).

Lisätty todellisuus (AR) ja virtuaalitodellisuus (VR): Tekoäly on keskeinen AR- ja VR-kokemusten kehittämisessä, parantaen virtuaaliympäristöjen realismia ja vuorovaikutteisuutta peleissä, virtuaalikierröksissä ja opetussisällöissä.

- **Toiminnallinen tehokkuus**

Ohjelmallinen mainonta: Tekoäly automaattisesti ostaa ja myy mainostilaa, optimoiden mainosten sijoittelun ja hinnoittelun reaaliajassa maksimoidakseen mainostajien ja mediatoimijoiden sijoitetun pääoman tuoton.

Sisällön jakelu: Tekoälytyökalut auttavat optimoimaan sisällön jakelustrategioita, analysoimalla kulutustottumuksia ja määrittämällä parhaat kanavat ja ajankohdat sisällön julkaisemiseksi, tavoittavuuden ja sitoutumisen maksimoimiseksi.

- **Luovan työn tukeminen**

Käsikirjoittaminen ja juonikehitys: Tekoälytyökalut auttavat käsikirjoittamisessa ehdottamalla juonenkäänteitä, dialogia ja hahmojen vuorovaikutuksia lajityyppien konventioiden ja yleisön mieltymysten pohjalta.

Taide ja graafinen suunnittelu: Tekoälypohjaiset työkalut tukevat taiteilijoita ja suunnittelijoita ehdottamalla parannuksia, tuottamalla ideoita ja automatisoimalla toistuvia tehtäviä, mikä mahdollistaa luovemman työskentelyn ja kokeilun.

- **Tapahtumien hallinta ja suunnittelu**

Väkijoukon hallinta: Tekoälyä käytetään suurten tapahtumien suunnittelussa ja hallinnassa, analysoiden osallistujadataa optimoidakseen tilajärjestelyt, aikataulutuksen ja turvallisuusratkaisut.

Personoidut kokemukset: Huvi- ja tapahtumapuistoissa tekoälyjärjestelmät tarjoavat henkilökohtaisia aikatauluja ja kokemuksia kävijän mieltymysten ja aiemman käyttäytymisen perusteella, mikä lisää asiakastytyytyväisyyttä ja -uskollisuutta.

3.2. Algoritmien käytön mahdolliset sudenkuopat

Huolimatta monista eduistaan algoritmit voivat aiheuttaa merkittäviä ongelmia, kuten harhaa ja syrjintää, yksityisyydensuojaan liittyviä haasteita, läpinäkyvyyden ja vastuullisuuden puutetta sekä liiallista automaation varaan luottamista (O'Neil, 2016). Kukin näistä sudenkuopista ei ainoastaan aiheuta eettisiä ja toiminnallisia riskejä, vaan voi myös heikentää yleisön luottamusta tekoälyteknologioihin. Tässä osiossa tarkastellaan näitä osa-alueita tarkemmin ja tuodaan esiin esimerkkejä eri toimialoilta tekoälyn mahdollisista vaaroista ja väärinkäytöksistä.

a. Harha ja syrjintä: Tekoälyjärjestelmät voivat vahvistaa ja toistaa ennakkoluuloja, jos ne on koulutettu vinoutuneilla tai edustamattomilla aineistoilla, jotka sisältävät ennakkoluuloisia virheitä. Tämä on erityisen merkittävä ongelma esimerkiksi rahoituksessa ja terveydenhuollossa, joissa algoritmipohjaiset päätökset voivat johtaa epäoikeudenmukaiseen kohteluun.

- **Rahoitusala:** Luottoluokituksessa käytettävät tekoälyjärjestelmät voivat heijastaa historiallisissa aineistoissa esiintyviä rotu- tai sukupuolivirheitä. Esimerkiksi jos malli koulutetaan aiemman lainapäätösaineiston perusteella ja tuo aineisto sisältää systemaattista syrjintää jotakin väestöryhmää kohtaan, malli saattaa jatkaa samanlaista syrjintää kyseistä ryhmää kohtaan.
- **Terveydenhuolto:** Tekoälyyn perustuvat diagnostiikkatyökalut ovat joissain tapauksissa antaneet ristiriitaisia hoitosuosituksia rodullisten tai sukupuolten välisten erojen perusteella. Tunnettu esimerkki on tekoälyjärjestelmä, joka ei onnistunut diagnosoimaan tiettyjä ihosairauksia tummemman ihon omaavilla henkilöillä, koska sen opetusaineisto sisälsi pääosin vaaleaihoisia potilaita.

b. Yksityisyyden suojaan liittyvät huolenaiheet: Tekoälyjärjestelmien kouluttamiseen ja käyttöön tarvitaan laajoja tietomääriä, mikä herättää merkittäviä yksityisyyden suojaan liittyviä huolia, erityisesti siitä, miten tietoa kerätään, säilytetään ja käytetään. Tämä näkyy erityisesti verkkokaupan ja koulutuksen aloilla, joissa henkilötietoja hyödynnetään laajasti asiakas- ja opiskelijakokemuksen parantamiseksi.

- **Verkkokauppa:** Jälleenmyyjät, jotka käyttävät tekoälyä ostokokemusten personointiin, saattavat kerätä laajoja määriä henkilökohtaisia tietoja, kuten ostohistoriaa ja reaaliaikaista sijaintia. Tämä herättää huolta tietomurtojen mahdollisuudesta ja henkilötietojen luvattomasta käytöstä.
- **Koulutus:** Koulutussektorilla tekoälyjärjestelmät, joita käytetään opiskelijoiden edistymisen seuraamiseen, voivat kerätä herkkää tietoa, kuten oppimisvaikeuksia, terveystietoja ja käyttäytymismalleja. Tämä nostaa esiin kysymyksiä siitä, kuka pääsee käsiksi näihin tietoihin ja kuinka niitä voidaan käyttää opetuskontekstin ulkopuolella.

c. Läpinäkyvyyden ja vastuullisuuden puute: Tekoälyjärjestelmät toimivat usein niin sanottuina "musta laatikko" -ratkaisuina, joiden päätöksentekoprosessit ovat epäselviä tai täysin piilossa käyttäjiltä ja viranomaisilta. Tämä läpinäkyvyyden puute voi johtaa vakaviin vastuukysymyksiin, erityisesti kun tekoälyn tekemät päätökset vaikuttavat merkittävästi ihmisten elämään.

- **Turvallisuus ja puolustus:** Sotilaallisissa sovelluksissa tekoälyn ohjaama päätöksenteko autonomisissa asejärjestelmissä voi johtaa elämän ja kuoleman tilanteisiin. Päätösten taustalla olevien prosessien läpinäkyvättömyys monimutkaistaa eettistä arviointia ja nostaa merkittäviä vastuullisuuteen liittyviä huolia.
- **Terveydenhuolto:** Tekoälyjärjestelmä, jota käytetään potilasdiagnooseissa, ei välttämättä tarjoa selvitystä päätöksenteon perusteista, jolloin lääkärin on vaikea ymmärtää, miten diagnoosiin on päädytty. Tämä heikentää luottamusta tekoälyn arvioihin ja monimutkaistaa vastuunjakoa väärän diagnoosin tapauksessa.

d. Liiallinen luottamus automaatioon: Voimakas riippuvuus tekoälystä voi johtaa ihmisten asiantuntemuksen heikkenemiseen, kun taidot rapistuvat tekoälyn ottaessa päätöksenteon haltuunsa. Tämä liiallinen luottamus automaatioon voi olla erityisen ongelmallista korkean riskin ympäristöissä, kuten liikenteessä ja rahoitusallalla.

- **Liikenne:** Lentoliikenteessä on tapahtunut tapauksia, joissa liiallinen luottamus automaattijärjestelmiin on osaltaan vaikuttanut onnettomuuksiin. Lentäjät saattavat luottaa liikaa autopilottiin ja puuttua heiltä riittävä manuaalilentokokemus järjestelmän vikatilanteessa.
- **Rahoitus:** Osakemarkkinoilla on koettu useita niin sanottuja "flash crash" -romahduksia, joihin on osittain vaikuttanut automaattisten kaupankäyntialgoritmien toiminta. Liiallinen luottamus näihin järjestelmiin ilman riittävää inhimillistä valvontaa voi johtaa äkillisiin markkinoiden laskuihin, jotka perustuvat algoritmien virheellisiin päätöksiin.

Vinkkejä ja suosituksia opettajille

Vinkkejä ja suosituksia opettajille algoritmien roolin hyödyntämiseen eri aloilla

1. Käytä tapaustutkimuksia

Aloita oppitunnit tarkastelemalla tosielämän tapaustutkimuksia, joissa algoritmeilla on ollut keskeinen rooli eri toimialoilla. Tämä tarjoaa käytännön ymmärryksen siitä, miten abstrakteja käsitteitä sovelletaan eri aloilla, kuten rahoituksessa, terveydenhuollossa ja verkkokaupassa.

2. Korosta eettisiä näkökulmia

Painota algoritmien käytön eettisiä vaikutuksia keskustelemalla mahdollisista vinoumista, tietosuojaan liittyvistä huolista ja vastuullisuudesta. Tämä kehittää osallistujien kriittistä ajattelua ja eettistä tietoisuutta.

3. Kannusta vuorovaikutteisiin keskusteluihin

Järjestä ryhmäkeskusteluja algoritmien käyttöönoton vaikutuksista. Esitä kysymyksiä, kuten: "Mitä voi mennä pieleen tämän algoritmin kanssa todellisessa tilanteessa?" tai "Miten voisimme vähentää mahdollisia riskejä?"

4. Hyödynnä monipuolisia materiaaleja

Sisällytä opetukseen videoita, infografiikkaa ja podcasteja, jotka havainnollistavat algoritmien vaikutuksia. Monimuotoisuus tukee erilaisia oppimistyyliä ja tekee sisällöstä kiinnostavampaa.

5. Simuloi algoritmien vaikutuksia

Käytä simulaatioita tai ohjelmistoja, joiden avulla osallistujat näkevät, miten algoritmien muutokset vaikuttavat mallinnetuissa ympäristöissä. Tämä käytännönläheinen lähestymistapa auttaa ymmärtämään asioita syvällisemmin.

Alkulämmittelytehtävät

1. Algoritmiroolien kartoitus

Anna osallistujille roolit, joissa kukin esittää algoritmin osaa jollakin toimialalla. Heidän tulee selittää roolinsa merkitys ja siihen liittyvät mahdolliset haasteet. Tämä syventää ymmärrystä algoritmien toiminnasta ja eettisistä vaikutuksista.

2. Eettinen väittely algoritmeista

Järjestä väittely kiistanalaisesta algoritmien käytöstä, kuten kasvojentunnistuksesta tai rikollisuuden ennakoinnista. Tämä rohkaisee osallistujia tarkastelemaan eri näkökulmia ja pohtimaan eettisiä haasteita.

3. Skenaarioanalyysi

Esittele pienryhmille skenaarioita, joissa algoritmeja käytetään eri aloilla. Pyydä heitä tunnistamaan mahdolliset hyödyt, riskit ja eettiset kysymykset, joita päättäjien tulisi harkita.

Arviointivinkit ja -suositukset

1. Tapaustutkimusanalyysi

Anna osallistujille tehtäväksi analysoida tapaustutkimus, jossa algoritmeja käytetään tietyllä alalla. Arvioi heidän kykyään tunnistaa sekä hyödyt että eettiset näkökohdat.

2. Ryhmäesitykset

Pyydä osallistujia esittelemään algoritmien roolia ja riskejä eri toimialoilla. Arvioi heidän ymmärrystään analyysin syvyyden ja eettisten näkökulmien käsittelyn perusteella.

3. Reflektiiviset esseet

Pyydä osallistujia kirjoittamaan essee siitä, miten algoritmien ymmärtäminen voi vaikuttaa heidän ammatilliseen vastuuseensa ja eettisiin näkemyksiinsä. Arvioi heidän kykyään yhdistää kurssin sisältöä omiin tai kuvitteellisiin ammatillisiin tilanteisiin.

4. Testit ja visat

Käytä visoja arvioidaksesi osallistujien ymmärrystä keskeisistä käsitteistä, kuten algoritmista vinoumista ja läpinäkyvyydestä. Sisällytä tilanteisiin perustuvia kysymyksiä, jotka vaativat kriittistä ajattelua.

5. Vertaispalaute

Sisällytä arviointiin vertaisarviointi. Tämä tuo esiin useita näkökulmia osallistujan ymmärryksestä ja kannustaa oppimaan tarkastelemalla materiaalia arvioijan näkökulmasta.

Näillä menetelmillä kouluttajat voivat tehokkaasti opettaa algoritmien merkitystä eri aloilla varmistaen, että osallistujat ymmärtävät sekä tekniset näkökulmat että laajemmat eettiset vaikutukset.

4. Algoritmien kompleksisuuden, optimoinnin sekä algoritmipohjaisen päätöksenteon rajoitusten ymmärtäminen

4.1. Algoritmien kompleksisuus

Tietojenkäsittelytieteessä algoritmit voivat joko mahdollistaa tai estää järjestelmien ja sovellusten menestyksen, koska niiden on toimittava tehokkaasti. Algoritmien kompleksisuus toimii kuin opastava valo kehittäjille ja insinööreille – se auttaa arvioimaan, kuinka paljon laskentatehoa eri algoritmit vaativat, ja tätä arviointia varten käytetään usein Big O -notaatiota. Kyse ei ole vain suoritusnopeudesta, vaan algoritmien kompleksisuuden ymmärtäminen auttaa myös hahmottamaan, kuinka hyvin algoritmi toimii järjestelmän kasvaessa ja kuinka resurssien käyttöä voidaan hallita.

Algoritmien kompleksisuus tarkoittaa käytännössä sitä, kuinka nopeasti ja kuinka paljon muistia algoritmi tarvitsee tehtävän suorittamiseen. Aikakompleksisuus kertoo, kuinka kauan algoritmilta kestää ratkaista ongelma ongelman koon kasvaessa. Tilakompleksisuus puolestaan kuvaa, kuinka paljon muistia algoritmi tarvitsee suoritusvaiheen aikana. Nämä mittarit auttavat meitä arvioimaan, kuinka tehokas algoritmi on ja soveltuuko se suurten tehtävien käsittelyyn ilman merkittävää hidastumista.

Tietojenkäsittelytieteessä käytetään Big O -notaatiota kuvaamaan matemaattisesti, kuinka nopeasti tai hitaasti algoritmit toimivat. Se auttaa meitä ymmärtämään, kuinka paljon aikaa tai muistia algoritmi tarvitsee, kun sille

syötetään enemmän tietoa. Jos esimerkiksi sanotaan, että algoritmi on $O(n)$, se tarkoittaa, että kun syötteen koko (" n ") kasvaa, algoritmin suorittamisaika kasvaa lineaarisesti. Mutta jos algoritmi on $O(n^2)$, sen suorittamisaika kasvaa huomattavasti nopeammin, koska kasvu on neliöllistä. Big O on siis nopea tapa hahmottaa, kuinka tehokas algoritmi on, kun se käsittelee suuria tietomääriä.

Algoritmien kompleksisuuden analysoinnin tärkeys näkyy useilla alueilla:

- **Suorituskyvyn ennustaminen:** Kun analysoimme algoritmien kompleksisuutta, pystymme ennakoimaan, kuinka ne toimivat eri kokoisilla syötteillä. Tämä ennustettavuus on ratkaisevaa, kun valitaan sopivin algoritmi tiettyyn ongelmaan ja optimoidaan järjestelmän suorituskykyä.
- **Skaalautuvuuden arviointi:** Ympäristöissä, joissa syötteen koko voi vaihdella, algoritmien kompleksisuuden ymmärtäminen auttaa meitä arvioimaan, kuinka hyvin algoritmi skaalautuu. Hyvin skaalautuvat algoritmit toimivat luotettavasti myös silloin, kun tietomäärät kasvavat merkittävästi varmistaen järjestelmien skaalautuvuuden ja reagoitokyvyn.
- **Resurssien hallinta:** Tehokas resurssien käyttö on ratkaisevan tärkeää ympäristöissä, joissa resurssit ovat rajallisia, kuten sulautetuissa järjestelmissä ja pilvipalveluissa. Algoritmisen kompleksisuuden analysointi auttaa tekemään päätöksiä resurssien jakamisesta, varmistaen resurssien optimaalisen käytön ja hävikin minimoinnin.
- **Algoritmien suunnittelu:** Algoritmisen kompleksisuuden ymmärtäminen vaikuttaa suunnitteluprosessiin, ohjaten kehittäjiä kohti algoritmeja, jotka tasapainottavat tehokkuuden ja toiminnallisuuden. Valitsemalla algoritmeja, joilla on suotuisat kompleksisuusominaisuudet, kehittäjät voivat luoda järjestelmiä, jotka toimivat optimaalisesti ilman, että toiminnallisuudesta tarvitsee tinkiä.

4.2. Aikakompleksisuus

Otetaan esimerkiksi laajasti tunnettu järjestämisongelma. Algoritmit, kuten kuplajärjestys (bubble sort), joiden algoritmisen kompleksisuus on merkitty $O(n^2)$, suoriutuvat huonosti suurten tietomassojen käsittelyssä verrattuna tehokkaampiin vaihtoehtoihin, kuten yhdistämisjärjestykseen (merge sort) tai pikalajitteluun (quicksort), joiden algoritmisen kompleksisuus on $O(n \log n)$. Tämän eron ymmärtäminen on ratkaisevaa, kun valitaan sopivia algoritmeja. Se parantaa järjestelmien suorituskykyä ja siten myös käyttäjäkokemusta (Karp, 2010).

Katsotaanpa selkeyttävä esimerkki: meidän täytyy lajitella tietomassa, jossa on miljardi tietuetta ($n = 1\,000\,000\,000$), mikä voisi vastata suuren sosiaalisen verkoston käyttäjälistan lajittelua.

- **Kuplajärjestys:** Kuplajärjestyksen algoritmisen kompleksisuuden ollessa $O(n^2)$, suoritusajan kasvu olisi valtava tämän kokoisella tietomassalla. Arviona suoritusajasta: Suoritusajan määrä olisi verrannollinen noin 10^{18} :aan. Jos jokaisen tietueen käsittely vie 1 nanosekunnin, kuplajärjestyksellä koko massan käsittely kestäisi noin 10^{18} nanosekuntia, mikä vastaa 11574,07 päivää eli noin 31,69 vuotta. Tästä voidaan päätellä, miksi algoritmin valinnalla on valtava merkitys suurten tietomassojen käsittelyssä. Tämä tehoton prosessi johtaisi todennäköisesti merkittäviin viiveisiin datan lajittelussa, mikä heikentäisi käyttäjäkokemusta huomattavasti.
- **Pikalajittelu:** Sen sijaan pikalajittelu, jonka aikakompleksisuus on $O(n \log n)$, osoittautuu huomattavasti tehokkaammaksi. Kun lajitellaan miljardi tietuetta, suoritusajan määrä olisi verrannollinen noin $1\,000\,000\,000 \times \log_2(1\,000\,000\,000)$. Jos jokaisen tietueen lajittelu kestää 1 nanosekunnin, pikalajittelulla koko tietomassan käsittely kestäisi noin 29 897 352 853,99 nanosekuntia, mikä vastaa noin 29,90 sekuntia eli noin 0,00034603 päivää. Näin ollen käyttäjät saisivat nopeamman vasteajan ja kokonaisuudessaan sulavamman käyttäjäkokemuksen.

Tehokkaampien lajittelualgoritmien, kuten pikalajittelun, käyttöönotto havainnollistaa, kuinka optimointi voi muuttaa laskennallisten prosessien tehokkuutta (Bentley, 1999).

Miksi lajitteluesimerkki? Datan järjestäminen on ratkaisevan tärkeää, jotta algoritmit toimisivat tehokkaasti — aivan kuten hyödyt, kuten nopeampi käsittely ja sujuvammat käyttäjäkokemukset, löytyvät tehokkaan datan hallinnan kautta.

Kuvittele etsiväsi tiettyä esinettä sotkuisesta huoneesta – se vie aikaa ja on työlästä. Samoin lajittelematon data algoritmeissa edellyttää lineaarista hakua, jossa jokainen alkio tarkastetaan yksitellen, kunnes haluttu löytyy. Tämän prosessin aikakompleksisuus on $O(n)$ (missä n edustaa aineiston kokoa), ja se muistuttaa sanan etsimistä kirjasta sivu sivulta – toimiva mutta tehoton menetelmä etenkin suurissa aineistoissa.

Nyt kuvittele sama huone siistinä ja esineet järjestettyinä. Tarvittavan esineen löytäminen on helppoa, koska tiedät, mistä etsiä. Samoin lajiteltu data mahdollistaa algoritmien käyttävän binäärihakua, joka on huomattavasti tehokkaampi menetelmä. Binäärihaku puolittaa hakuvälin toistuvasti, kunnes haluttu alkio löytyy, ja sen aikakompleksisuus on $O(\log n)$. Tämä on kuin sanan etsimistä sanakirjasta sivuja puolittaen jokaisella käännöllä.

Yhteenvetona: datan lajittelu saa algoritmit toimimaan fiksummin, ei kovemmin. Se muuttaa työlään lineaarisen haun salamannopeaksi binäärihauksi, lyhentäen käsittelyaikaa ja parantaen suorituskykyä kokonaisuudessaan.

Seuraavan kerran kohdatessasi lajittelematonta dataa, muista lajittelun merkitys ja sen kyky vapauttaa algoritmien todellinen tehokkuus.

4.3. Algoritmien muistikompleksisuus

Kun perehdytään algoritmien kompleksisuuteen muistin käytön näkökulmasta, kohdataan kriittinen osa-alue algoritmianalyysissa, joka täydentää aikakompleksisuuden tarkastelua. Muisti, rajallinen resurssi tietojenkäsittely-ympäristöissä, vaikuttaa merkittävästi algoritmien suorituskykyyn ja järjestelmän tehokkuuteen. Algoritmien muistikompleksisuuden tarkastelu voi kuulostaa monimutkaiselta, mutta puretaanpa se yksinkertaisempaan muotoon.

Kun puhumme algoritmien kompleksisuudesta ja muistista, tarkastelemme oikeastaan sitä, kuinka paljon tilaa algoritmi tarvitsee ongelman ratkaisemiseen. Voit ajatella sitä kuin reppusi pakkaamista eri tehtäviä varten – jotkut tehtävät vaativat enemmän tilaa, toiset vähemmän.

Kuvittele, että lajittelet nipun kortteja, joissa on numeroita. Yksi tapa tehdä tämä on käyttää merge sort -nimistä lajittelualgoritmia. Merge sort jakaa kortit pienempiin ryhmiin, lajittelee jokaisen ryhmän ja yhdistää ne sitten takaisin yhdeksi nipuksi. Vaikka merge sort on erittäin nopea lajittelussa, se tarvitsee myös ylimääräistä tilaa toimiakseen. Tämä ylimääräinen tila on kuin lisäpöytiä, joille asetat kortteja järjestelyn aikana.

Haluatko, että jatkan myös seuraavien kappaleiden suomentamista, jos sinulla on niitä valmiina?

Toinen esimerkki on, kun lasketaan Fibonacci-lukuja, joita esiintyy luonnossa (kuten kukan terälehtien määrässä tai simpukankuoren spiraalissa). Yksi tapa laskea näitä lukuja on käyttää menetelmää nimeltä dynaaminen ohjelmointi. Tämä menetelmä auttaa löytämään luvut nopeammin, mutta se vaatii myös ylimääräistä muistia muistaakseen jo lasketut luvut. Se on kuin muistikirja, johon kirjoitat ylös luvut sitä mukaa kun lasket niitä.

Puhutaanpa nyt siitä, miten tallennamme tietoa tietokoneohjelmissa. Meillä on käytössämme erilaisia työkaluja, kuten taulukot ja listat, jotka auttavat pitämään asiat järjestyksessä. Jokainen näistä työkaluista vie tietyn määrän tilaa tietokoneen muistista.

Esimerkiksi, kuvittele että teet listaa ystäväsi nimistä. Jos käytät yksinkertaista listaa, kuten kirjoittamalla nimet paperille, se ei vie paljoa tilaa. Mutta jos käytät jotain monimutkaisempaa, kuten taulukkoa, jossa on paljon sarakkeita, se vie enemmän tilaa, koska tallennat lisätietoja jokaisesta ystävästä.

Kun analysoimme algoritmien muistikompleksisuutta, pohdimme siis sitä, kuinka paljon tilaa algoritmimme ja tietorakenteemme tarvitsevat hoitaakseen tehtävänsä tehokkaasti.

Kun ymmärrämme tämän, voimme tehdä parempia valintoja ohjelmoinnissa varmistaaksemme, että ohjelmat toimivat sujuvasti eivätkä käytä liikaa muistia. Se on kuin pakkaisit laukkusi fiksusti niin, että saat kaiken tarvitsemasi mukaan ilman, että laukku käy liian raskaaksi.

Lyhyesti sanottuna: muistin ja algoritmisen kompleksisuuden huomioon ottaminen auttaa meitä kirjoittamaan parempia ohjelmia, jotka toimivat tehokkaasti eivätkä kuluta liikaa tietokoneen resursseja. Se on kuin olisit tietokoneesi fiksu järjestelijä!

Algoritminen kompleksisuus on keskeinen käsite nykyaikaisessa tietojenkäsittelyssä. Se tarjoaa jäsenneilyn tavan arvioida ja parantaa algoritmien tehokkuutta. Käyttämällä Big O -notaatiota voimme ymmärtää algoritmien perusominaisuuksia ja sitä, miten ne suoriutuvat eri kokoisilla aineistoilla.

Algoritmisen kompleksisuuden ymmärtäminen auttaa kehittäjiä ja insinöörejä luomaan järjestelmiä, jotka kestävät tietojenkäsittelyn jatkuvasti muuttuvia vaatimuksia. Kun teknologia kehittyy, algoritmisen kompleksisuuden tuntemus pysyy elintärkeänä innovaation edistämisessä ja tietokoneiden kyvykkyyden parantamisessa.

4.4. Optimointi

Optimointi toimii innovaatioiden ja tehokkuuden majakkana, ohjaten algoritmeja suoriutumaan ennennäkemättömällä nopeudella ja tarkkuudella. Ytimeltään optimointi on taidetta jalostaa algoritmeja siten, että kompleksisuus minimoidaan ja suorituskyky maksimoidaan, vapauttaen niiden koko potentiaalin ratkaisemaan monimutkaisia ongelmia hallitusti.

Kuvittele algoritmit digitaalisen maailmamme moottoreiksi, jotka pyörittävät kaikkea hakukoneista logistiikkajärjestelmiin. Kuten hyvin viritetty moottori, myös optimoitu algoritmi toimii sulavasti, kulkien nopeasti läpi valtavien aineistojen ja monimutkaisten laskentojen. Mutta miten optimointi oikein saavuttaa tämän?

Optimointi kattaa joukon strategioita, joiden tavoitteena on virtaviivaistaa algoritmeja. Siihen kuuluu pullonkaulojen tunnistaminen, tarpeettomien vaiheiden poistaminen ja prosessien hienosäätö, jotta saavutetaan huippuluokan tehokkuus. Optimoinnin perimmäinen tavoite on löytää optimaalinen tasapaino resurssien käytön ja lopputuloksen laadun välillä. Yksi optimoinnin keskeisistä päämääristä on kompleksisuuden minimoiminen. Algoritmit kohtaavat usein haasteita joutuessaan ratkaisemaan monimutkaisia ongelmia, jotka vaativat suuria laskentaresursseja.

Optimoimalla algoritmeja pyritään yksinkertaistamaan niiden rakennetta ja vähentämään laskennallista kuormitusta, tehden niistä hallittavampia ja skaalautuvampia.

Lisäksi optimoinnin tavoitteena on parantaa suorituskkyä hyödyntämällä innovatiivisia lähestymistapoja. Tämä tarkoittaa esimerkiksi edistysellisten tekniikoiden, kuten rinnakkaislaskennan, heurististen algoritmien ja koneoppimisen, hyödyntämistä tehokkuuden lisäämiseksi. Näiden edistysaskeleiden omaksuminen antaa algoritmeille mahdollisuuden tuottaa nopeampia tuloksia käyttäen samalla vähemmän resursseja.

Ajatellaanpa esimerkiksi internetin hakualgoritmien optimointia. Huolellisen hienosäädön ja algoritmien parantelun avulla hakukoneet pystyvät seulomaan valtavia määriä tietoa ja tarjoamaan olennaisia hakutuloksia muutamissa millisekunneissa. Tämä optimointi parantaa paitsi käyttäjäkokemusta, myös vähentää palvelinten kuormitusta ja energiankulutusta.

Kun keskitytään rinnakkaislaskentaan, tietojenkäsittelyn maailmassa grafiikkaprosessorit (GPU:t) ovat nousseet korvaamattomiksi komponenteiksi, erityisesti tekoälyn (AI) ja koneoppimisen yhteydessä. Vaikka niitä käytetään laajasti, monille GPU:iden toimintaperiaatteet ja niiden keskeinen rooli tekoälyteknologioiden kehityksessä ovat vielä vieraita.

GPU:n ydinajatus on, että se on erikoistunut elektroninen piiri, joka on suunniteltu nopeasti käsittelemään ja muuttamaan muistia nopeuttaakseen kuvien luontia kehysmuistiin, josta ne siirretään näyttölaitteelle. Alun perin grafiikan renderöintiin kehitetyt GPU:t ovat kehittyneet erittäin rinnakkaistetuiksi prosessoreiksi, jotka pystyvät suorittamaan tuhansia laskentatehtäviä samanaikaisesti.

Toisin kuin perinteiset keskusyksiköt (CPU:t), jotka on optimoitu peräkkäisten tehtävien käsittelyyn, GPU:t on suunniteltu rinnakkaiskäsittelyyn. Ne koostuvat lukuisista pienemmistä käsittely-yksiköistä eli "ytimistä", joista kukin kykenee suorittamaan tehtäviä itsenäisesti. Tämä rinnakkaisarkkitehtuuri mahdollistaa sen, että GPU:t voivat käsitellä laskennallisesti vaativia tehtäviä erittäin tehokkaasti.

GPU:t on suunniteltu erityisesti tehtäviin, joissa samaa operaatiota suoritetaan usealle tietoelementille samanaikaisesti. Tämä rinnakkaisuus tekee niistä erittäin tehokkaita grafiikan renderöinnissä sekä kuva- ja videonkäsittelyssä, joissa jokaiselle pikselille tehdään lukuisia laskelmia. Lisäksi niitä hyödynnetään simulaatioissa ja mallinnuksessa, kryptovaluutan louhinnassa sekä rinnakkaistetuissa matriisioperaatioissa, muiden esimerkkien ohella.

Rinnakkaisuus on erityisen hyödyllistä tekoälysovelluksissa, joissa tehtävät vaativat valtavien tietomäärien käsittelyä samanaikaisesti.

Tehtävät kuten syvien neuroverkkojen kouluttaminen, kuvantunnistus, luonnollisen kielen käsittely ja data-analyysi hyötyvät valtavasti GPU:iden rinnakkaiskäsittelykyvystä.

Tekoäly perustuu vahvasti monimutkaisiin matemaattisiin operaatioihin, kuten matriisikertolaskuihin ja konvoluutioihin, jotka ovat olennaisia koneoppimismallien kouluttamisessa ja käyttöönotossa. Näissä operaatioissa käsitellään suuria matriiseja ja suoritetaan lukuisia laskutoimituksia samanaikaisesti, mikä tekee niistä ihanteellisia rinnakkaiseen suorittamiseen GPU:illa.

Tässä vaiheessa meidän on otettava esiin tekoälyalgoritmien energiankulutus. Koska nämä järjestelmät ovat yleistyneet eri toimialoilla, on tärkeää keskittyä energiatehokkuuteen ympäristövaikutusten ja käyttökustannusten vähentämiseksi. Optimointitekniikat ja algoritmien kompleksisuus ovat keskeisessä roolissa energiansäästön saavuttamisessa tekoälykehyksissä. Optimointi tarkoittaa algoritmien hienosäätöä siten, että saavutetaan maksimaalinen suorituskyky mahdollisimman vähäisillä laskentaresursseilla. Energiankulutusta tekoälykehyksissä syntyy esimerkiksi datan käsittelystä ja mallien kouluttamisesta. Matala algoritmien kompleksisuus on ratkaisevaa energiankulutuksen vähentämisessä, koska se vaatii vähemmän laskennallisia operaatioita.

4.5. Algoritmisen päätöksenteon rajoitukset

Algoritmien päätöksenteko on noussut välineeksi prosessien automatisointiin ja optimointiin monilla eri aloilla. Vaikka algoritmien päätöksenteko on lupaavaa, siihen liittyy luontaisia rajoitteita, jotka vaativat tarkkaa arviointia. Rahoituksesta terveydenhuoltoon, koulutuksesta rikosoikeuteen – algoritmeja on otettu käyttöön luvaten ennennäkemätöntä tehokkuutta, tarkkuutta ja objektiivisuutta. Silti niiden vaikutuspiirissä on monimutkainen verkosto rajoitteita, jotka estävät algoritmisen päätöksenteon rajoittamattoman tehokkuuden (Yeung, 2018).

Vinouma ja syrjintä: Yksi algoritmisen päätöksenteon merkittävimmistä ja haitallisimmista rajoitteista on taipumus vinoomaan ja syrjintään. Vaikka tavoitteena olisi puolueettomuus, algoritmit perivät usein vinoumia käyttämästään koulutusdatasta. Historiallisten epäoikeudenmukaisuuksien ja yhteiskunnallisten vinoumien esiintyminen datassa voi ylläpitää ja pahentaa eriarvoisuutta, mikä johtaa syrjiviin lopputuloksiin. Lisäksi algoritmien menetelmien läpinäkyvättömyys vaikeuttaa vinoumien tunnistamista ja torjumista, mikä lisää huolta päätöksenteon oikeudenmukaisuudesta ja tasavaroisuudesta. Tämä voi olla myönteinen mahdollisuus havaita vinoumia ja syrjintää, joita ihmishavainnoijat eivät aiemmin ole tunnistaneet. Suurten tietoaisteistojen analysointi kehittyneillä algoritmeilla mahdollistaa hienovaraisten mallien ja erojen tunnistamisen, minkä avulla voidaan ryhtyä korjaaviin toimiin systeemisten vinoumien oikaisemiseksi. Tiettyjen algoritmien läpinäkyvyys ja

jäljitettävyyden mahdollistavat päätöksentekoprosessin tarkan tarkastelun, mikä antaa sidosryhmille valmiudet puuttua vinoumien perussyihin.

Kontekstuaalisen ymmärryksen puute: Algoritminen päätöksenteko toimii kontekstuaalisesti naiivissa ympäristössä, nojaten yksinomaan kvantitatiivisiin mittareihin ja ennalta määriteltäviin sääntöihin. Tämä rajoite korostuu erityisesti monimutkaisissa ja dynaamisissa ympäristöissä, joissa kontekstin hienovaraisuudet ovat keskeisiä päätöksenteossa. Ihmisen harkintakyky, jota tukee kontekstuaalinen ymmärrys ja asiantuntemus, ylittää usein algoritmisten järjestelmien kyvyt havaita vivahteikkaita yksityiskohtia ja tehdä perusteltuja päätöksiä. Tämän seurauksena algoritmiset järjestelmät voivat osoittaa jäykkyyttä ja riittämättömyyttä muuttuvien kontekstien huomioimisessa, mikä heikentää niiden päätösten tehokkuutta ja soveltuvuutta.

Ennalta arvaamattomat seuraukset ja eettiset dilemmat: Algoritmien deterministinen luonne altistaa ne ennalta arvaamattomille seurauksille ja eettisille pulmatilanteille, jotka johtuvat algoritmisen päätöksenteon prosessien luontaisesta yksinkertaistamisesta. Odottamattomia tuloksia, ketjureaktioita ja tahattomia vaikutuksia voi ilmetä, koska algoritmit eivät kykene huomioimaan todellisten tilanteiden monimutkaisuutta ja hienovaraisia piirteitä. Lisäksi eettiset kysymykset, kuten yksityisyyden loukkaaminen, autonomian heikentyminen ja moraalinen vastuu, korostavat tarvetta huolelliselle harkinnalle ja valvonnalle algoritmisten järjestelmien käyttöönotossa.

4.6. Kuinka tunnistaa algoritmisten järjestelmien mahdolliset vinoumien ja virheiden lähteet? Keskeinen kysymys oikeudenmukaisuuden, läpinäkyvyyden ja vastuullisuuden varmistamiseksi

Yksi algoritmisen päätöksenteon vinoumien pääasiallisista lähteistä on algoritmien kouluttamiseen ja testaamiseen käytetty data. Datassa esiintyvät vinoumat voivat siirtyä koko päätöksentekoprosessiin, mikä johtaa vääristyneisiin lopputuloksiin. Mahdollisten vinoumien tunnistamiseksi datassa on tarpeen suorittaa perusteellinen analyysi datan keräämisprosessista arvioiden mahdollisia systemaattisia vinoumia tai tiettyjen ryhmien aliedustusta, tarkastella aineiston demografisia ominaisuuksia epätasapainojen tai erojen havaitsemiseksi, jotka voivat johtaa vinoutuneisiin tuloksiin, sekä arvioida datan esikäsittelyvaiheita, kuten datan puhdistusta ja piirteiden valintaa, tunnistaa mahdolliset tahattomat muunnokset tai vääristymät datassa.

Mahdollisten vinoumien ja virheiden tunnistamiseksi algoritmien suunnittelussa ja toteutuksessa tulisi tarkastella algoritmimalleja ja -menetelmiä arvioiden algoritmin taustaoletuksia ja rajoitteita sen määrittämiseksi, vastaavatko ne eettisiä ja yhteiskunnallisia normeja, tutkia koulutusprosessia mahdollisten ylioppimisen (overfitting) tai alioppimisen (underfitting) lähteiden tunnistamiseksi, sillä nämä voivat johtaa epätarkkoihin tai epäoikeudenmukaisiin

ennusteisiin, ja tutkia algoritmin käyttämien piirteiden ja muuttujien valintaa, jotta voidaan varmistaa, että ne ovat olennaisia, syrjimättömiä ja edustavat väestöä. Ylioppiminen (overfitting) ja alioppiminen (underfitting) ovat kaksi yleistä ongelmaa koneoppimismallien koulutuksessa. Ylioppimisessa malli oppii koulutusdatan satunnaisia vaihteluita ja kohinaa todellisten kaavojen sijaan. Tämän seurauksena malli suoriutuu hyvin koulutusdatassa, mutta ei kykene yleistämään uusiin, ennestään näkymättömiin datoihin. Alioppimisessa taas malli on liian yksinkertainen hahmottaakseen datan taustalla olevia rakenteita, mikä johtaa heikkoon suorituskyykyyn sekä koulutusdatassa että uudessa datassa.

Arvioinnin ja validoinnin vaiheessa mahdollisten vinoumien ja virheiden tunnistamiseksi tulisi käyttää testausmenetelmiä, joilla mitataan algoritmin suorituskyykyä eri väestöryhmissä mahdollisten eriarvoisuuksien ja epäoikeudenmukaisuuksien havaitsemiseksi. Lisäksi tulisi suorittaa herkkyyksianalyysiä algoritmin vakauden arvioimiseksi syötedatan tai malliparametrien vaihteluihin nähden. On myös tärkeää kerätä palautetta sidosryhmiltä – mukaan lukien vaikutuksille altistuvat yhteisöt ja alakohtaiset asiantuntijat – mahdollisten kehitysvaiheissa huomaamatta jääneiden vinoumien tai eettisten huolten tunnistamiseksi. Lisäksi jatkuva seuranta ja auditointi ovat välttämättömiä vinoumien, virheiden ja odottamattomien seurausten havaitsemiseksi ja lieventämiseksi.

4.7. Algoritmisten riskien ja vinoumien käsitleminen ja lieventäminen

Kun algoritmisiin riskeihin ja vinoumiin on saatu tunnistettua, ne on käsiteltävä monialaisella lähestymistavalla, joka kattaa tekniset, menettelytavat huomioivat ja eettiset näkökulmat. Teknisiin toimenpiteisiin voi kuulua algoritmien hienosäätö käyttämällä esimerkiksi vinoumia vähentäviä menetelmiä, koulutusdatan monipuolistamista sekä oikeudenmukaisuusmittareiden sisällyttämistä mallien arviointiin.

Menettelylliset toimenpiteet tarkoittavat vankkojen validointiprotokollien, läpinäkyvyyttä edistävien käytäntöjen ja vastuullisuusmekanismien käyttöönottoa, jotta algoritmipohjaiset päätökset voidaan asianmukaisesti tarkastaa ja validoida. Eettiset toimenpiteet liittyvät eettisen tietoisuuden ja vastuullisuuden kulttuurin vahvistamiseen kehittäjien, päättäjien ja loppukäyttäjien keskuudessa. Niissä korostetaan algoritmisen päätöksenteon eettisiä vaikutuksia sekä edistetään eettisten ohjeistusten ja standardien noudattamista.

Algoritmisten riskien ja vinoumien lieventäminen edellyttää jatkuvaa seuranta, arviointia ja mukautumista muuttuviin haasteisiin ja toimintaympäristöihin. Algoritmien suorituskyykyyn ja vaikutusten jatkuva seuranta mahdollistaa uusien vinoumien ja riskien havaitsemisen ja niihin puuttumisen ajoissa.

Arviointimekanismit, kuten vinouma-auditoinnit, herkkyyssanalyysit ja vaikutustenarvioinnit, tarjoavat tietoa riskienhallintastrategioiden tehokkuudesta ja tukevat jatkuvaa kehittämistä.

Lisäksi monitieteisen yhteistyön ja vuorovaikutuksen edistäminen mahdollistaa oivallusten ja parhaiden käytäntöjen jakamisen eri alojen välillä, mikä vahvistaa yhteisiä pyrkimyksiä minimoida algoritmisiin riskeihin ja vinoumiin liittyviä ongelmia

Vinkkejä ja suosituksia opettajille

Vinkkejä ja suosituksia opettajille algoritmisen kompleksisuuden käsittelyyn

1. Käsitteellinen aloitus

Aloita selittämällä algoritminen kompleksisuus yksinkertaisesti arkipäiväisten esimerkkien avulla, kuten tapahtuman järjestäminen tai matkan suunnittelu. Näin havainnollistat tehokkuuden ja resurssienhallinnan käsitteitä.

2. Visuaaliset havainnollistukset

Käytä visuaalisia apuvälineitä, kuten kaavioita ja graafeja, jotka näyttävät, miten eri algoritmit toimivat eri olosuhteissa. Tämä auttaa osallistujia näkemään, miten kompleksisuus vaikuttaa suorituskyykyyn.

3. Käytännön harjoituksia

Sisällytä käytännön tehtäviä, joissa osallistujat voivat kokeilla eri algoritmeja tiettyjen ongelmien ratkaisemiseksi. Tämä voi sisältää yksinkertaisia koodausharjoituksia tai ohjelmistoja, jotka simuloivat algoritmien suorituskyykyä.

4. Yhdistä tosielämän sovelluksiin

Kytke keskustelu todellisiin tilanteisiin, joissa algoritminen kompleksisuus on kriittinen, kuten suurten teknologiayritysten datankäsittely tai GPS-reittien laskenta. Tämä korostaa aiheen merkitystä.

5. Kannusta yhteistyöhön

Järjestä ryhmäkeskusteluja ja ongelmanratkaisutilanteita, joissa osallistujat pohtivat, mitä algoritmeja käyttää eri tilanteissa. Tämä syventää ymmärrystä yhteistyön kautta.

Alkulämmittelytehtävät

1. Lajittelutehtävä

Järjestä käytännön tehtävä, jossa osallistujat lajittelevat esineitä (esim. kirjoja tai

kortteja) eri tavoin matkiakseen lajittelualgoritmeja. Keskustelkaa, miten kukin menetelmä liittyy algoritmin aika- ja tilakompleksisuuteen.

2. Algoritmipeli

Luo korttipeli, jossa osallistujat yhdistävät eri algoritmit niiden Big O -merkintöihin ja käyttötapauksiin. Tämä interaktiivinen lähestymistapa auttaa vahvistamaan teoreettista ymmärrystä pelillisesti.

3. Tehokkuusaivorihi

Pyydä osallistujia listaamaan arjen tehtäviä, joita voisi optimoida algoritmien avulla (esim. sähköpostien lajittelu tai tiedostojen järjestäminen). Keskustelkaa, miten algoritmien kompleksisuuden ymmärtäminen voisi parantaa näitä prosesseja.

Arviointivinkit ja -suositukset

1. Käsitetarkistukset

Käytä lyhyitä tietovisoja säännöllisesti arvioidaksesi ymmärrystä keskeisistä käsitteistä, kuten Big O -merkinnästä ja aika- ja tilakompleksisuuden eroista. Tämä varmistaa, että perusasiat ovat hallussa ennen monimutkaisempien aiheiden käsittelyä.

2. Käytännön koodauskokeet

Jos mahdollista, sisällytä koodauskokeita, joissa osallistujat kirjoittavat tai tunnistavat tehokkaimmat algoritmit annettuihin ongelmiin. Näin mitataan teorian soveltamista käytännössä.

3. Projektipohjainen arviointi

Anna osallistujille pieni projekti, jossa heidän tulee analysoida jonkin järjestelmän tai ohjelmiston suorituskykyä ja ehdottaa parannuksia algoritmisen kompleksisuuden perusteella.

4. Ryhmäesitykset

Jaa osallistujat ryhmiin valmistelevaan esityksiä siitä, miten algoritmisen kompleksisuus vaikuttaa eri aloihin, kuten rahoitukseen, terveydenhuoltoon tai teknologiaan. Arvioi heidän ymmärrystään ja kykyään esittää monimutkaista tietoa selkeästi.

5. Reflektiiviset esseet

Pyydä osallistujia kirjoittamaan pohdiskeleva essee siitä, miten algoritmisen kompleksisuuden ymmärtäminen voi vaikuttaa heidän työhönsä tai opintoihinsa. Näin arvioit heidän kykyään sisäistää ja soveltaa oppimaansa.

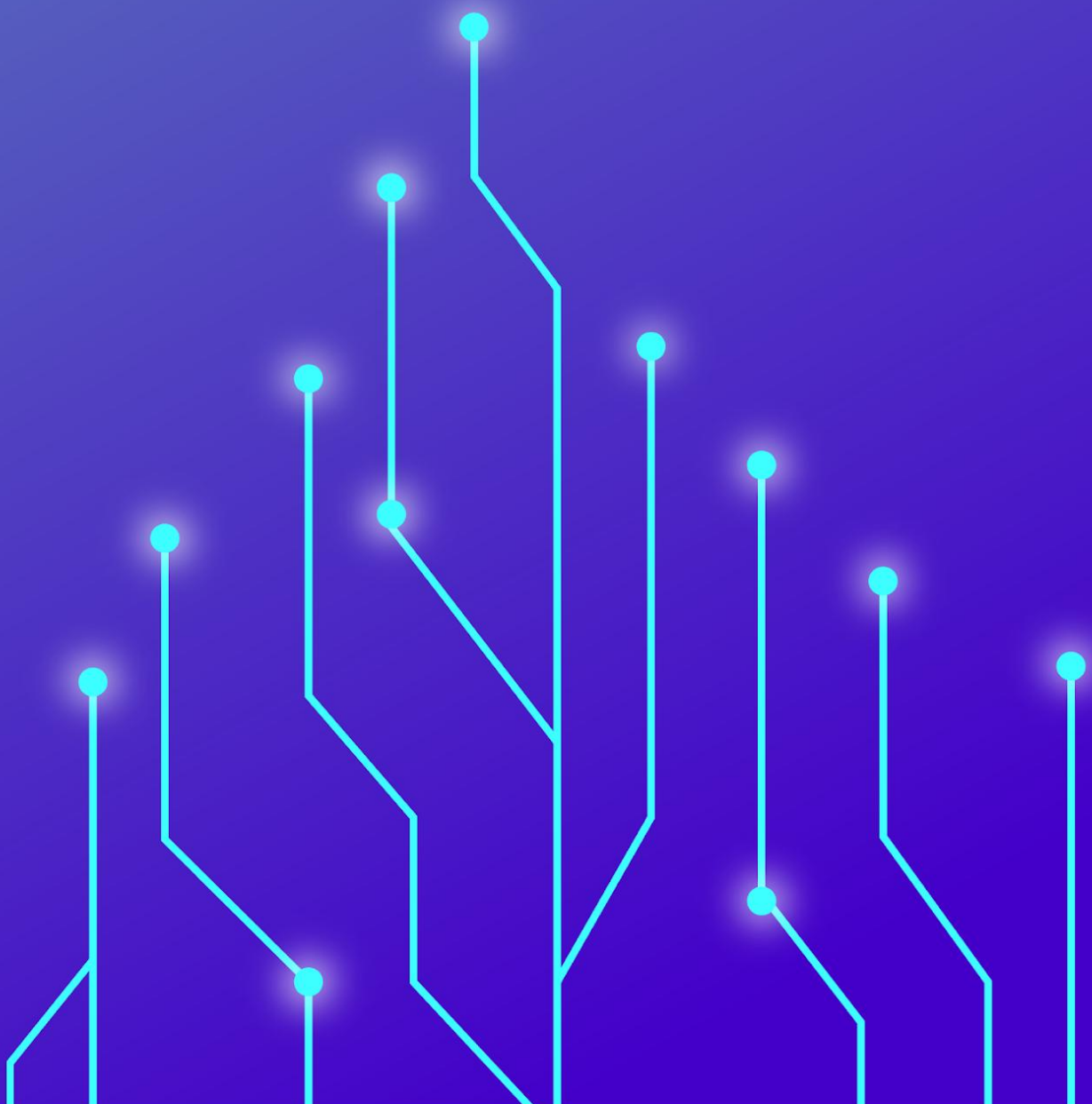
Näillä menetelmillä opettajat voivat tehokkaasti opettaa algoritmista kompleksisuutta, varmistaa osallistujien sitoutumisen ja tukea oppimista käytännön sovellusten kautta.

5. Viittaukset

- Aksoy, T., & Gurol, B. (2021). Artificial intelligence in computer-aided auditing techniques and technologies (CAATs) and an application proposal for auditors. In *Auditing Ecosystem and Strategic Accounting in the Digital Era: Global Approaches and New Opportunities* (pp. 361-384). Cham: Springer International Publishing.
- Bentley, P. (1999). *Evolutionary design by computers*. Morgan Kaufmann.
- Bohr, A., & Memarzadeh, K. (2020). The rise of artificial intelligence in healthcare applications. In *Artificial Intelligence in healthcare* (pp. 25-60). Academic Press.
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms*. MIT Press.
- Davenport, T. H., & Mittal, N. (2023). *All-in on AI: How smart companies win big with artificial intelligence*. Harvard Business Press.
- Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & governance*, 12(4), 505-523. <https://doi.org/10.1111/rego.12158>
- Donovan, J., Caplan, R., Matthews, J., & Hanson, L. (2018). *Algorithmic accountability: A primer*. <https://apo.org.au/node/142131>
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>
- Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1), 2053951719860542. <https://doi.org/10.1177/2053951719860542>
- Karp, R.M. (2010). Reducibility Among Combinatorial Problems. In: Jünger, M., et al. *50 Years of Integer Programming 1958-2008*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-68279-0_8
- Knuth, D. E. (1997). *The Art of Computer Programming: Fundamental Algorithms, volume 1*. Addison-Wesley Professional.
- Mourtzis, D., Angelopoulos, J., & Panopoulos, N. (2022). A Literature Review of the Challenges and Opportunities of the Transition from Industry 4.0 to Society 5.0. *Energies*, 15(17), 6276. <https://doi.org/10.3390/en15176276>

- O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.
- Parker, G. G., Van Alstyne, M. W., & Choudary, S. P. (2016). *Platform revolution: How networked markets are transforming the economy and how to make them work for you*. WW Norton & Company.
- Patel, K. (2024). Ethical reflections on data-centric AI: balancing benefits and risks. *International Journal of Artificial Intelligence Research and Development*, 2(1), 1-17. <https://orcid.org/0009-0005-9197-2765>
- Qin, X. (2012). Making use of the big data: next generation of algorithm trading. In *Artificial Intelligence and Computational Intelligence: 4th International Conference, AICI 2012, Chengdu, China, October 26-28, 2012*. Proceedings 4 (pp. 34-41). Springer Berlin Heidelberg.
- Reisdorf, B. C., & Blank, G. (2021). Algorithmic literacy and platform trust. In *Handbook of digital inequality* (pp. 341-357). Edward Elgar Publishing.
- Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The Ethics of Algorithms: Key Problems and Solutions. In: Floridi, L. (eds) *Ethics, Governance, and Policies in Artificial Intelligence*. Philosophical Studies Series, vol 144. Springer, Cham. https://doi.org/10.1007/978-3-030-81907-1_8

Kurssi 4 | Datan oikeudenmukaisuus ja vinoumat



Sisältö

1. Johdanto	99
2. Oikeudenmukaisuus & vinoumat - esimerkkinä makeiset	100
3. Kokonaiskuvan ymmärtäminen: Tekoälyn oikeudenmukaisuus ja vinoumat	101
4. Vinoumat tekoälyjärjestelmissä	102
4.1. Vinoumien esiintyminen tekoälyjärjestelmissä	102
4.2. Tekoälyn vinoumien seuraukset	103
4.3. Tekoälyn vinoumien käsittely	103
4.4. Yleiskatsaus vinoumien tyypeistä	103
5. Menetelmät tekoälyn vinoumien tunnistamiseen ja mittaamiseen	104
6. Strategiat tekoälyn vinoumien käsittelemiseksi ja lieventämiseksi	105
7. Oikeudenmukaisuuden mittareiden ymmärtäminen	106
8. Oikeudenmukaisuus datan keräämisessä ja esikäsittelyssä	107
8.1. Esikäsittelyn tekniikat	108
9. Nousevat suuntaukset tekoälyn oikeudenmukaisuudessa	110
10. Johtopäätös	111

1. Johdanto

Tekoälyjärjestelmät eivät ole vain abstrakteja käsitteitä, vaan ne ovat yhä kiinteämpi osa arkeamme. Nämä järjestelmät ohjaavat päätöksiä terveydenhuollossa, koulutuksessa, rahoituksessa ja turvallisuudessa. Tekoälyn vaikutusvallan kasvaessa on vaarana, että nämä järjestelmät eivät ainoastaan heijasta olemassa olevia yhteiskunnallisia ennakkoluuloja, vaan myös vahvistavat tai luovat niitä uudelleen. Tekoälyn oikeudenmukaisuuden ja vinoumien ymmärtäminen ei ole pelkästään akateeminen harjoitus, vaan olennaisen tärkeää oikeudenmukaisten ja tasapuolisten järjestelmien kehittämisessä, jotka vaikuttavat jokapäiväiseen elämäämme.

Tällä kurssilla käsitellään tekoälyn vinoumia ja oikeudenmukaisuutta, tarjoten syvemmän ymmärryksen vinoumiin liittyvistä ongelmista sekä keinoja niiden ratkaisemiseen. Kurssi tarjoaa tietoa, miten tekoälyalgoritmeihin voi tahattomasti syntyä vinoumia, millaisia vaikutuksia niillä voi olla ja millaisia strategioita niiden lieventämiseksi voidaan käyttää – sekä miten juuri sinä voit olla osa ratkaisua. Kurssilla perehdytään oikeudenmukaisten tekoälyjärjestelmien kehittämiseen liittyviin eettisiin, yhteiskunnallisiin ja teknisiin haasteisiin, korostaen samalla sinun rooliasi tekoälyn vastuullisessa ja eettisessä kehittämisessä. Tämän kurssin suoritettuaasi osaat:

Tunnistaa ja ymmärtää vinoumien tyyppejä:

Tunnistat erityyppisiä vinoumia tekoälyjärjestelmissä, kuten datan vinoumat, algoritmin vinoumat ja tulosten vinoumat. Lisäksi kykenet tunnistamaan, mitkä vaikuttavat näiden vinoumien ilmenemiseen.

Mitata ja määrittää vinoumia:

Tunnistat käytössäsi olevat työkalut ja menetelmät tekoälyn vinoumien arviointiin ja mittaamiseen eri tekniikoin.

Laatia strategia vinouman vähentämiseksi:

Ymmärrät miten hyödynnät eri strategioita tekoälyn vinoumien vähentämiseksi, mukaan lukien erilaiset tiedonkeruukäytännöt, algoritmiset mukautukset ja käyttöönoton seuranta.

Arvioida tekoälyn oikeudenmukaisuutta:

Pystyt arvioimaan tekoälyn oikeudenmukaisuutta eri ulottuvuuksilla varmistaen, etteivät järjestelmät ylläpidä tai pahenna sosiaalista eriarvoisuutta.

Varautua tulevaisuuden haasteisiin:

Kykenet ennakoimaan ja vastaamaan tekoälyn oikeudenmukaisuuden haasteisiin teknologian kehittyessä, säilyttäen mukautuvuuden ja ennakoivan lähestymistavan.

Vinkkejä ja suosituksia opettajille

Ehdotus: Aloita lyhyellä tehtävällä, jossa osallistujat kertovat käsityksistään tai henkilökohtaisista kokemuksistaan tekoälyn käytöstä arjessa. Tämä voi auttaa osallistujia yhdistämään tekoälyn abstraktin käsitteen ja sen käytännön vaikutukset.

2. Oikeudenmukaisuus & vinoumat - esimerkkinä makeiset



IMAGE SOURCE | Generated by DALL-E

Kuvittele itsesi luokkahuoneeseen ja katselet opettajan pöydällä olevaa kulhollista lempikarkkejasi. Jokainen oppilas saa valita palan, mutta valintaa määrittää epäoikeudenmukaisuus: punapaitaiset oppilaat saavat kaksi palaa, kun taas kaikki muut saavat yhden. Se ei tunnu reilulta. Tätä kutsutaan vinoumaksi (bias) - kun joku (tai jokin, kuten tekoäly) antaa epäreilua etua tietyille ihmisryhmälle ilman hyvää syytä.

Kuvittele nyt, että opettaja sen sijaan pyytää kaikkia ratkaisemaan palapelin, ja se, joka saa sen oikein, saa paidan väristä riippumatta ylimääräisen karkkipalan. Tämä tuntuu reilulta, koska kaikilla on sama mahdollisuus - kyse on palapelistä, ei paidasta. Tämä on oikeudenmukaista (fairness), jossa kaikilla on yhtäläiset mahdollisuudet, eikä ketään suosita mielivaltaisista syistä.

Kuten tässä skenaariossa, tekoälyjärjestelmien pitäisi antaa kaikille oikeudenmukaiset mahdollisuudet, olipa kyse sitten elokuvien suosittelusta, lainojen hyväksymisestä tai musiikin soittamisesta. Kun tekoäly on reilua, kaikki ratkaisevat palapelin ja saavat saman määrän karkkia.

Kun tekoäly on vinoutunut, se on kuin jotkut ihmiset saisivat enemmän karkkia vain paidan värin vuoksi. Ja me haluamme pyrkiä oikeudenmukaisuuteen!

Vinkkejä ja suosituksia opettajille

Ehdotus: Käytä tätä esimerkkiä vuorovaikutteisesti antamalla osallistujien käydä skenaario lävitse joko oikeiden karkkien tai vastaavasti jonkin virtuaalisen toiminnon avulla. Tämä käytännönläheinen lähestymistapa voi auttaa vahvistamaan oikeudenmukaisuuden ja vinoumien käsitteitä.

3. Kokonaiskuvan ymmärtäminen: Tekoälyn oikeudenmukaisuus ja vinoumat

Tietojen oikeudenmukaisuudella (fairness) tekoälyssä tarkoitetaan periaatetta, jonka mukaan tekoälyjärjestelmien olisi kohdeltava kaikkia yksilöitä tasapuolisesti ilman epäoikeudenmukaisia tai ennakkoluuloisia tuloksia, jotka perustuvat luontaisiin tai hankittuihin ominaisuuksiin. Tekoälyn oikeudenmukaisuus varmistaa, että näiden järjestelmien tekemät päätökset eivät suosi tai aseta mitään tiettyä ihmisryhmää muita heikompaan asemaan. Tämä on erityisen tärkeää ihmisten elämään vaikuttavassa päätöksenteossa, kuten palkkaamisessa, lainojen hyväksymisessä, lainvalvonnassa ja terveydenhuollon diagnostiikassa.

Tekoälyssä esiintyvällä vinoumilla (bias) taas tarkoitetaan puolestaan järjestelmällisiä virheitä järjestelmän tiedoissa tai algoritmeissa, jotka johtavat epäoikeudenmukaisiin tuloksiin tiettyjen ihmisryhmien kannalta. Ennakkoluulot voivat olla selviä, kuten järjestelmään tarkoituksellisesti sisäänrakennettuja, tai useammin implisiittisiä, jotka johtuvat epäedustavasta tai epätäydellisestä data-aineistosta tai virheellisestä algoritmisuunnittelusta, joka tahattomasti suosii tai haittaa tiettyjä väestöryhmiä.



IMAGE SOURCE | Generated by DALL-E

Tekoälyn oikeudenmukaisuuden ja vinoumien käsittelyn merkitystä ei voida liikaa korostaa. Vinoutuneet tekoälyjärjestelmät voivat ylläpitää ja vahvistaa olemassa olevaa sosiaalista eriarvoisuutta, mikä johtaa vaikeasti katkaistaviin epäoikeudenmukaisuuden kierteisiin. Lisäksi vinoumat heikentävät luottamusta teknologiaan, mikä voi hidastaa sen käyttöönottoa ja innovointia. Oikeudenmukaisuuden varmistaminen tekoälyjärjestelmissä ei ole vain tekninen välttämättömyys, vaan myös moraalinen velvollisuus, jotta yhteiskunnan luottamus säilyy ja sen mahdolliset haitat voidaan ennaltaehkäistä.

Vinkkejä ja suosituksia opettajille

Ehdotus: käytä tapaustutkimuksia tai uutisartikkeleita, joissa tuodaan esiin tekoälyjärjestelmien vinoumien tuottamia ongelmia. Näin voidaan tuoda esiin tekoälyn oikeudenmukaisuuden merkitys.

Rekrytointi-tekoäly suosi toista sukupuolta työnhaussa

Vuonna 2018 paljastui, että Amazonin rekrytinnissa käyttämä tekoälyjärjestelmä syrji naishakijoita. Järjestelmän vinouma suosi mieshakijoita, sillä tekoäly oli hyödyntämän historiadatan perusteella, miesehdokkaat olivat olleet suositeltavampia viimeisen kymmenen vuoden ajalta. Todellisuudessa, tämä ajanjakso koostui pääasiassa mieshenkilöiden työhakemuksista, mikä johtui teknologiateollisuuden epätasaisesta sukupuolijakautumisesta. Järjestelmä luokitteli alaspäin ansioluettelot, joissa esiintyi sana "naisten", kuten "naisten shakkikerhon kapteeni" tai "naisten college".

4. Vinumat tekoälyjärjestelmissä

Tekoälyn vinoumilla tarkoitetaan systemaattisia virheitä, jotka johtavat epäoikeudenmukaisiin tuloksiin tiettyjen ryhmien kannalta ja tyypillisesti yhden väestöryhmän suosimiseen muiden kustannuksella. Nämä vinumat voivat johtua eri lähteistä, kuten tekoälyjärjestelmien kouluttamiseen käytetyistä tiedoista, näitä tietoja käsittelevistä algoritmeista tai näiden järjestelmien suunnittelijoiden ja käyttöönnottajien tulkinnallisista vinoumista.

4.1. Vinoumien esiintyminen tekoälyjärjestelmissä

- **Datan vinumat:** Tämä vinouma syntyy, kun tekoälyalgoritmien kouluttamiseen käytetyt datakokonaisuudet eivät edusta laajempaa väestöä. Tämä voi johtua historiallisesta epätasa-arvosta tai yksinkertaisesti siitä, että tiedonkeruumenetelmät keräävät yhdestä väestöryhmästä perusteellisemmin tietoa kuin muista.
- **Algoritmin vinumat:** Toisinaan algoritmit saatetaan suunnitella siten, että ne luonnostaan suosivat tiettyjä tuloksia. Tämä voi johtua siitä, miten algoritmit tulkitsevat tietoja, tai tekoälyn kehittämisen aikana asetetuista tavoitteista.
- **Palautteet:** Vinouma voi myös säilyä ja vahvistua käyttäjän palautteen kautta, joissa alkuperäiset vinoutuneet päätökset johtavat vastaavan datan keräämiseen, joka taas vahvistaa näitä vinoumia ja luo syrjinnän kierteen.

4.2. Tekoälyn vinoumien seuraukset

- **Sosiaalinen vaikutus:** Tekoälyn vinumat voivat vahvistaa olemassa olevaa sosiaalista eriarvoisuutta ja stereotypioita sekä johtaa yhteiskuntaan, jossa

teknologia asettaa järjestelmällisesti tiettyjä ryhmiä epäedulliseen asemaan. Luottamuksen heikkenemisellä tekoälyteknologioihin voidaan nähdä laajempia vaikutuksia tekoälyn käyttöönottoon ja tehokkuuteen eri aloilla.

- **Yksilötason vaikutus:** Yksilötasolla, vinoumat voivat johtaa epäoikeudenmukaiseen kohteluun kriittisillä osa-alueilla, kuten palkkauksessa, lainvalvonnassa, terveydenhuollossa ja pankkipalveluissa. Yksilöt voivat kohdata epäoikeudenmukaisia tuloksia, jotka johtuvat virheellisistä päätöksentekoprosesseista, joiden takana ovat vinoutuneet tekoälyjärjestelmät.

4.3. Tekoälyn vinoumien käsittely

On olennaisen tärkeää ottaa käyttöön kattavat strategiat tekoälyn kehittämisen kaikissa vaiheissa ennakkoluulojen lieventämiseksi. Näitä ovat muun muassa:

- Tietolähteiden monipuolistaminen, jolla varmistetaan ihmisryhmien laaja edustus järjestelmien hyödyntämissä tietoaaineistoissa.
- Algoritmien tarkastusten kehittäminen ja toteuttaminen vinoumien tunnistamiseksi ja korjaamiseksi.
- Jatkuvan seurannan -järjestelmä, jotta havaitaan ja käsitellään järjestelmässä esiintyvät vinoumat oikeudenmukaisiksi.

4.4. Yleiskatsaus vinoumien tyypeistä

Tekoälyjärjestelmissä ilmenevien erityyppisten vinoumien ymmärtäminen on olennaista kehittäjille, poliittisille päättäjille ja käyttäjille, jotka pyrkivät luomaan oikeudenmukaisia ja tasapuolisia tekoälyjärjestelmiä.. Tässä kappaleessa vinoumien eri tyyppejä, ja kuhunkin niistä on liitetty tosielämän esimerkkejä, jotka havainnollistavat niiden vaikutusta arjessa.

1. Valintavinouma: Valintavinoumaa esiintyy, kun tekoälyjärjestelmän kouluttamiseen käytetty data ei edusta kohdeväestöstä tai -skenaariota, mikä johtaa vinoutuneisiin lopputuloksiin.

Esimerkiksi tekoälymalli, joka on koulutettu tunnistamaan kasvonpiirteitä käyttämällä pääasiassa nuoremmista henkilöistä koostuvaa aineistoa, järjestelmä voi epäonnistua tunnistamaan tarkasti ikääntyneiden aikuisten piirteitä, koska mallin koulutuksessa ei ollut edustavaa otosta kaikista ikäryhmistä.

2. Vahvistusvinouma: Vahvistusvinouma ilmenee tekoälymallissa, kun sen käyttämä data tai sen tulkinta tahattomasti vahvistaa ennakkoluuloja ja jättää huomioimatta ristiriitaisen tiedon.

Esimerkiksi rekrytoinnissa käytettävä tekoäly, joka on koulutettu pääasiassa yhden sukupuolen ansioluetteloiden avulla, saattaa edelleen suosia kyseisen sukupuolen ehdokkaita, mikä vahvistaa nykyistä epätasapainoa työpaikkojen valinnassa.

3. Otantavinouma: Otantavinouma on eräänlainen Valintavinouma, jossa tekoälyn kouluttamista varten kerätyt otostiedot eivät edusta tarkasti laajempaa perusjoukkoa.

Esimerkiksi jos puheentunnistusjärjestelmä koulutetaan pääasiassa amerikkalaisella aksentilla puhuvilta henkilöiltä saadun datan perusteella, sillä saattaa olla vaikeuksia ymmärtää muiden alueiden aksentteja, kuten brittiläistä tai tai australialaista aksenttia, koska se on koulutettu epäedustavalla otoksella.

4. Algoritminen vinouma: Algoritminen vinouma syntyy, kun tekoälyjärjestelmiä ohjaavat algoritmit tuottavat vinoutuneita tuloksia, usein niiden suunnittelussa olevien sisäisten virheiden tai luomisvaiheessa asetettujen tavoitteiden vuoksi.

Esimerkiksi luoton myöntämiseen hyödynnettävä tekoälymalli voi hylätä lainahakemuksia perustuen hakijan asuinalueeseen, jos mallin hyödyntämä historiallinen data heijastaa aikaisempia luottopäätöksiä vinoutuneesti.

5. Raportointivinouma: Raportointivinouma syntyy, kun tekoälyjärjestelmään syötetty data on vinoutunutta, koska tiettyjen tulosten raportointi on yleisempää kuin muiden.

Esimerkiksi lääkkeiden haittavaikutuksia seuraava tekoälyjärjestelmä, joka saa pääasiassa raportteja vakavista tapauksista, voi virheellisesti päätellä, että suurin osa potilaista kokee äärimmäisiä reaktioita, jättäen huomiotta lievemmät, mutta yleisemmät haittavaikutukset.

6. Mittausvinouma: Mittausvinouma tarkoittaa tiedonkeruuvirheitä, jotka systemaattisesti vääristävät tietoja

Esimerkiksi ympäristönseurannan tekoäly, joka käyttää vähemmän tehokkaita antureita kylmemmillä alueilla, saa vääristyneen kuvan saasteiden tasoista ja saattaa virheellisesti aliraportoida päästöjä talvikuukausina.

5. Menetelmät tekoälyn vinoumien tunnistamiseen ja mittaamiseen

Tekoälyyn liittyvien ennakkoluulojen ymmärtäminen ja lieventäminen on ratkaisevan tärkeää oikeudenmukaisten ja tehokkaiden järjestelmien kehittämisessä. Tässä kappaleessa esitellään yksityiskohtaisesti menetelmät näiden ennakkoluulojen tunnistamiseksi ja mittaamiseksi, sekä tarjotaan tekoälyn kehittäjille ja tietojenkäsittelytieteilijöille selkeä väylä, jonka avulla he voivat varmistaa, että heidän tekoälyjärjestelmänsä toimivat ilman epäoikeudenmukaisia vinoumia.

- **Data auditointi** on kattava tarkistusprosessi, jossa tekoälymallien kouluttamiseen käytetyt aineistot tutkitaan mahdollisten vääristymien tai poikkeavuuksien varalta. Datan tarkastuksen suorittaminen edellyttää seuraavaa:
 - **Arvioi datan edustavuus:** Arvioi, heijastaako aineisto tarkasti väestön monimuotoisuutta, jota malli tulee palvelemaan. Tämä sisältää demografisen edustavuuden ja eri skenaarioiden kattavuuden tarkistamisen.
 - **Tunnista puuttuva data:** Etsi samankaltaisuuksia puuttuvasta datasta – aukot voivat aiheuttaa vinoumia, jos tietyt alaryhmät ovat ali-edustettuina.
 - **Analysoi merkintöjen johdonmukaisuus:** Varmista, että tietojen merkintäprosessi on johdonmukainen ja puolueeton. Merkintäerot voivat johtaa merkittäviin vinoumiin mallin tuloksissa.
 - **Tilastollinen analyysi:** Käytä tilastollisia työkaluja poikkeavuuksien tai vinoutuneiden datanjakaumien havaitsemiseksi, jotka saattavat viitata vinoutuneeseen dataan.
- **Eroavuusanalyysi** sisältää mallin suorituskyvyn vertaamisen eri demografisten ryhmien tai muiden relevanttien alaryhmien välillä erojen tunnistamiseksi, jotka voivat viitata vinoumaan. Eroavuusanalyysin vaiheet ovat:
 - **Määritä relevantit alaryhmät:** Valitse ryhmät, jotka perustuvat mallin sovelluksen kannalta merkityksellisiin ominaisuuksiin, kuten ikä, sukupuoli, etnisyys tai muut tunnusmerkit.
 - **Suorituskykymittareiden mittaaminen:** Laske suorituskykymittarit, kuten tarkkuus, täsmällisyys, palautus ja F1-tulos, kullekin alaryhmälle.
 - **Vertaa tuloksia:** Analysoi suorituskykymittareiden eroja ryhmien välillä. Merkittävät erot voivat viitata vinouman olemassaoloon.
 - **Juurisyyanalyysi:** Jos eroja havaitaan, tutki niiden taustalla olevia syitä, jotka voivat liittyä datan keräämiseen, mallin rakenteeseen tai ominaisuuksien valintaan.
- **Herkkyysanalyysi** testaa tekoälymallin luotettavuutta tekemällä pieniä muutoksia syöttötietoihin ja tarkkailemalla, miten tulokset muuttuvat. Tämä on erityisen hyödyllistä vinoumien tunnistamisessa:
 - **Muuta syöttötietoja:** Tee pieniä muutoksia syöttötietoihin, jotka heijastavat uskottavia vaihteluita todellisissa tilanteissa.
 - **Seuraa tulosten vaihtelua:** Tarkkaile, johtavatko nämä muutokset suhteettomiin muutoksiin mallin tuloksissa.

- **Arvioi vaikutus eri ryhmiin:** Keskity siihen, miten nämä muutokset vaikuttavat mallin tuloksiin eri demografisille ryhmille
- **Toteuta säätöjä:** Käytä havaintoja hyväksi hienosäätämällä datan esikäsittelyä, ominaisuuksien suunnittelua tai mallia vähentääksesi ei-toivottuja herkkyyksiä.

6. Strategiat tekoälyn vinoumien käsittelemiseksi ja lieventämiseksi

Vinoumien lieventäminen tekoälyjärjestelmissä vaatii monivaiheista lähestymistapaa suunnittelu-, kehitys- ja käyttöönotovaiheissa. Seuraavat strategiat ovat olennaisia, jotta voidaan **vähentää vinoumien riskiä**.

- **Monipuolinen edustus:** Data-aineiston sekä taustalla työskentelevän kehitystiimin monimuotoisuuden varmistaminen on ratkaisevan tärkeää oikeudenmukaisten tekoälyjärjestelmien kehittämisessä.
 - **Datan monimuotoisuus:** Kerää dataa monenlaisista lähteistä, jotta se kattaa laajan osan väestöstä. Tämä sisältää eri alueet, sosio-ekonomiset taustat ja muut relevantit ominaisuudet
 - **Tiimin monimuotoisuus:** Tekoälymallin takana työskentelevät monimuotoiset tiimit tuovat mukanaan erilaisia näkökulmia, joiden avulla voidaan tunnistaa ja lieventää mahdollisia ennakkoluuloja, joita homogeenisempi ryhmä ei välttämättä tunnistaisi.
- **Vinoumat huomioivat algoritmit:** Ennakoiva lähestymistapa vinoumien lieventämiseen on kehittää algoritmeja, jotka ovat luonnostaan tietoisia vinoumista ja pystyvät mukautumaan niihin.
 - **Oikeudenmukaisuusmittarien sisällyttäminen:** Sisällytä algoritmin tavoitefunktioon oikeudenmukaisuusnäkökulmat, kuten virheiden jakautumisen minimointi eri ryhmien välillä.
 - **Vastakkainasettelu:** Toteuta vastakkaisia koulutustekniikoita, joissa malleja koulutetaan samanaikaisesti toimimaan oikein ja minimoimaan vinoumia.
- **Säännöllinen seuranta ja arviointi** Tekoälyjärjestelmien jatkuva seuranta ja säännöllinen arviointi ovat keskeisiä, jotta voidaan varmistaa, että niiden pysyminen ajan mittaan puolueettomina ja mukautuminen uusiin tietoihin tai olosuhteisiin.

- **Seurantakehysten luominen:** Luo järjestelmiä, jotka seuraavat jatkuvasti tekoälymallien suorituskykyä ja keskittyvät ensisijaisesti oikeudenmukaisuuden mittareihin.
- **Säännölliset tarkistukset:** Tarkastele ja kalibroi malleja säännöllisesti uudelleen jatkuvan seurannan tulosten perusteella ja mukauta ne tietomallien tai yhteiskunnallisten normien muutoksiin.
- **Sidosryhmien palaute:** Osallista sidosryhmät, myös loppukäyttäjät, antamaan palautetta tekoälyn suorituskyvystä ja koetusta oikeudenmukaisuudesta, mikä voi antaa tietoa, jota ei saada selville määrällisissä arvioinneissa.

7. Oikeudenmukaisuuden mittareiden ymmärtäminen

Oikeudenmukaisuusmittarit ovat ratkaisevan tärkeitä arvioitaessa, miten hyvin tekoälyjärjestelmät kohtelevat eri taustoista tulevia yksilöitä tai ryhmiä. Niiden avulla voidaan mitata ja varmistaa, että tekoälyjärjestelmät eivät ylläpidä tai syvennä olemassa olevia yhteiskunnallisia vinoumia. Alla on lueteltu joitakin **kriittisiä oikeudenmukaisuusmittareita**, jotka olisi otettava huomioon järjestelmien arvioinnissa:

- **Tilastollinen yhdenvertaisuus (väestöllinen yhdenvertaisuus):** Tällä mittarilla arvioidaan, onko myönteisten tulosten osuus sama eri väestöryhmissä (esim. sukupuoli, rotu). Se on ratkaisevan tärkeä sovelluksissa, joissa tavoitteena on varmistaa tulosten yhdenvertaisuus riippumatta ryhmäidentiteettiin liittyvistä ominaisuuksista.
- **Tasapuoliset todennäköisyydet:** Tämä mittari edellyttää, että eri ryhmille tehdyt päätökset ovat yhtä tarkkoja – eli sekä oikeiden positiivisten (herkkyys) että väärin positiivisten (erottelukyky) osuuksien tulee olla samat kaikissa ryhmissä. Sitä käytetään erityisesti tilanteissa, joissa virheellisellä päätöksellä on samanlainen vaikutus kaikkiin yksilöihin, kuten rikosoikeudessa tai rekrytoinnissa.
- **Ehdollinen väestöllinen eriarvoisuus:** Tämä mittari arvioi, kuinka ennustetut tulokset eroavat toisistaan, kun ne on sidottu relevantteihin ominaisuuksiin (esim. ikä, koulutus tai työkokemus). Se mittaa eroja, joita ei voida selittää merkityksellisillä tekijöillä, ja se soveltuu tilanteisiin, joissa ryhmien välillä odotetaan olevan joitakin eroja kyseisten ominaisuuksien vuoksi.

Oikeudenmukaisuus tietoisuuden kautta: Tämä lähestymistapa sisältää herkkien ominaisuuksien (kuten rotu tai sukupuoli) tietoisuuden sisällyttämisen suoraan päätöksentekoprosessiin oikeudenmukaisuuden varmistamiseksi. Tämä mittari puolustaa algoritmeja, jotka kompensoivat

historiallisia epäoikeudenmukaisuuksia tai yhteiskunnallisia vinoumia, jotka voivat vaikuttaa oikeudenmukaisuuteen

Yksilön oikeudenmukaisuus: Tämä mittari edellyttää, että samankaltaiset yksilöt saavat samanlaisia ennusteita riippumatta taustastaan. Käsite on erityisen tärkeä personoiduissa palveluissa, kuten työsuosituksissa, joissa yhdenmukainen kohtelu samankaltaisten käyttäjäprofiilien välillä on olennaista.

Vaihtoehtoinen oikeudenmukaisuus: Tämän mittarin mukaan päätös on oikeudenmukainen, jos se olisi tehty myös vaihtoehtoisessa maailmassa, jossa yksilö olisi täysin samanlainen, mutta kuuluisi eri väestöryhmään. . Tämä mittari on hyödyllinen yksilötason oikeudenmukaisuuden arvioinnissa ja auttaa ymmärtämään tiettyjen ominaisuuksien vaikutusta päätöksiin.

Sopivien oikeudenmukaisuusmittareiden valinta riippuu tekoälysovelluksen erityisestä kontekstista ja tavoitteista, kuten:

- 1. Sovelluskohtaiset vaatimukset:** Oikeudenmukaisuusmittarin valinnan tulisi vastata sovellusalueen tavoitteita ja lakisääteisiä vaatimuksia. Esimerkiksi Tasapuoliset todennäköisyydet voivat olla keskeisiä rikosoikeudellisissa tapauksissa, kun taas Tilastollinen yhdenvertaisuus saattaa sopia paremmin markkinointiin tai mainontaan.
- 2. Sääntely- ja eettiset näkökulmat:** Joillakin aloilla voi olla säädöksiä, jotka edellyttävät tiettyjen oikeudenmukaisuusmittareiden käyttöä. Lisäksi eettiset tekijät, kuten tarve korjata aiempia epäoikeudenmukaisuuksia, voivat ohjata mittarien valintaa.
- 3. Vaikutusten arviointi:** On tärkeää ottaa huomioon tekoälyjärjestelmän mahdolliset yhteiskunnalliset vaikutukset ja valita mittarit, jotka vähentävät haitallisia vinoumia. Esimerkiksi Vaihtoehtoinen oikeudenmukaisuus voi olla erityisen tärkeä terveydenhuollossa, jossa yksilöllinen hoito on välttämätöntä.
- 4. Tekninen toteutettavuus:** Tekoälyjärjestelmän monimutkaisuuden ja saatavilla olevan datan mukaan jotkin mittarit voivat olla vaikeampia toteuttaa kuin toiset. On tärkeää huomioida tekniset rajoitteet, jotta valitut oikeudenmukaisuusmittarit voidaan soveltaa tarkasti ja johdonmukaisesti.

8. Oikeudenmukaisuus datan keräämisessä ja esikäsittelyssä

Varmistamalla, että datan keruukäytännöt ovat monipuolisia ja puolueettomia, voidaan rakentaa oikeudenmukaisia tekoälyjärjestelmiä.

- **Edustava otanta:** Varmista, että datan keruu kattaa kaikki väestöryhmät, joita tekoälyjärjestelmä palvelee. Tämä voi tarkoittaa esimerkiksi ositettua otantaa, jolla varmistetaan aliedustettujen ryhmien mukanaolo.
- **Jatkuva datan keruu:** Päivitä ja laajenna datakokoelmia säännöllisesti, jotta ne heijastavat väestön uusia ja muuttuvia trendejä. Tilastolliset aineistot voivat vanhentua nopeasti, mikä voi johtaa vinoumiin.

8.1. Esikäsittelytekniikat

Oikeudenmukaisuuden varmistaminen datan keruussa ja esikäsittelyssä on olennaista, jotta vinoumat eivät pääse vaikuttamaan tekoälyjärjestelmiin. Alla on käytäntöjä, jotka voivat auttaa saavuttamaan tasapainoisemman ja oikeudenmukaisemman tekoälyjärjestelmän kiinnittämällä huomiota tekoälyn kehityksen alkuvaiheisiin.

- **Puuttuvan datan käsittely:** Analysoi puuttuvan datan toistuvat ominaisuudet, sillä ne voivat aiheuttaa vinoumia. Tällöin puuttuvat tiedon korvaamisessa (imputatio) tulee olla huolellinen, jotta vältetään lisävinoumien syntyminen.
- **Ominaisuuksien valinta:** Arvioi kriittisesti, mitkä ominaisuudet otetaan mukaan mallin koulutusprosessiin. Herkkien ominaisuuksien kanssa korreloivia ominaisuuksia olisi käsiteltävä varovasti, jotta vältetään välillinen syrjintä.
- **Normalisointi ja standardointi:** Sovelletaan normalisointia tai standardointia sen varmistamiseksi, että malli ei luonnostaan painota tiettyjä ominaisuuksia pelkästään niiden lukuarvojen vuoksi.
- **Monimuotoiset datalähteet:** Jotta datan keruu kattaa laajan väestöryhmien kirjon, on keskeistä toteuttaa varmistus tekoälymallien vinoumien estämiseksi. Tavoitteena on kerätä dataa, joka heijastaa todellisen maailman monimuotoisuutta ja vaihtelua, jossa tekoälyjärjestelmä toimii. Tämä tarkoittaa erilaisten väestöryhmien huomioimista ja useiden keruupisteiden hyödyntämistä. Esimerkiksi terveydenhuollon tekoälyä kehittäessä dataa tulisi kerätä sairaaloista eri alueilta, mukaan lukien maaseutu- ja kaupunkialueet, jotta saadaan kattava kuva terveystieteiden ryhmistä ja ympäristön vaikutuksista sairauksiin.
- **Inklusiivinen otanta:** Inklusiivisen otannan tavoitteena on varmistaa, että kaikki väestöryhmät ovat edustettuina koulutusdatassa. Tavoitteena on estää vähemmistön tai aliedustettujen ryhmien poissulkeminen, sillä niiden puuttuminen voi johtaa vinoutuneisiin tekoälyn ennusteisiin. Tällöin käytetään esimerkiksi ositettua otantaa, jossa väestö jaetaan alaryhmiin (ositteisiin), jotka heijastavat tärkeitä piirteitä (kuten etninen tausta, ikä, sukupuoli), ja näistä ositteista valitaan satunnaisesti otoksia, jotta varmistetaan datan monimuotoisuus.

- **Datan puhdistus:** Datan puhdistaminen tarkoittaa epätarkkuuksien ja epäjohdonmukaisuuksien poistamista, jotka voivat vääristää tekoälymallin tuloksia. Tavoitteena on varmistaa, että tietojen laatu on korkea ja että niissä ei ole virheitä, jotka voivat vääristää mallin päätöksiä. Tämä prosessi sisältää poikkeavien arvojen korjaamisen tai poistamisen, puuttuvien arvojen täyttämisen sopivilla imputointistrategioilla, jotka eivät aiheuta vinoumaa, sekä datan yhdenmukaisuuden ja oikean muotoilun varmistamisen.
- **Jatkuva seuranta:** Datan ja tekoälymallin suorituskyvyn seuranta on keskeistä mahdollisten vinoumien havaitsemiseksi ja korjaamiseksi ajan myötä. Tavoitteena on toteuttaa jatkuvaa valvontaa sen varmistamiseksi, että tekoälyjärjestelmä toimii oikeudenmukaisesti myös uuden datan kertyessä ja olosuhteiden muuttuessa. Automaattisten seurantajärjestelmien käyttöönotto, jotka arvioivat tekoälyn tuotoksia oikeudenmukaisuuden mittareilla ja hälyttävät kehittäjiä vinoumien ilmetessä, on tärkeää. Järjestelmän tulisi säännöllisesti arvioida tekoälylle syötettävää dataa datan laadun tai edustavuuden muutosten havaitsemiseksi.
- **Oikeudenmukaisten ominaisuuksien valinta:** Oikeudenmukaisten ominaisuuksien valinta tarkoittaa sellaisten muuttujien valintaa ja muokkaamista, jotka minimoivat vinoumat säilyttäen samalla datan hyödynnettävyyden. Tavoitteena on varmistaa, etteivät mallin syötteet aseta mitään ryhmää epäedulliseen asemaan. Tekniikoihin kuuluu ominaisuuksien todennäköisyyksien analysointi herkkien ominaisuuksien kanssa sekä sellaisten ominaisuuksien muokkaaminen tai poistaminen, jotka voisivat johtaa vinoutuneisiin lopputuloksiin. Ulottuvuuden vähentämistekniikoita voidaan myös käyttää tunnistamaan ja poistamaan tarpeettomia tai vinoumia sisältäviä piirteitä.
- **Läpinäkyvyys ja dokumentointi:** Läpinäkyvyys varmistetaan yksityiskohtaisella dokumentaatiolla datan keruusta, esikäsittelystä ja piirrevalinnasta. Tavoitteena on tarjota selkeät ja auditoitavissa olevat tiedot, jotka täyttävät oikeudenmukaisuuden vaatimukset. Yksityiskohtaisten lokien pitäminen datankäsittelypäätöksistä, puhdistusmenetelmistä ja piirrevalintojen perusteluista on oleellista. Näiden dokumenttien tulisi olla sidosryhmien ja sääntelyelinten saatavilla tekoälyjärjestelmän oikeudenmukaisuuden arvioimiseksi.
- **Vinoumien tunnistaminen:** Vinoumien tunnistaminen tarkoittaa mahdollisten puolueellisuuksien havaitsemista datassa ennen sen käyttöä tekoälymallien koulutuksessa. Tavoitteena on korjata vinoumat ennakoivasti ennen kuin ne vaikuttavat mallin oikeudenmukaisuuteen. Tilastollisten analyysien ja visualisointityökalujen käyttö datan jakautumisen tarkastelemiseksi eri väestöryhmissä on keskeistä. Tämä voi tarkoittaa erikoistuneiden ohjelmistojen tai algoritmien käyttöä, jotka osoittavat, missä kohdin data ei ehkä ole edustavaa tai missä tulokset saattavat vaikuttaa suhteettomasti tiettyihin ryhmiin.

Tekoälyjärjestelmän oikeudenmukaisuus alkaa jo siitä, miten data kerätään, käsitellään ja valmistellaan ennen kuin se edes päätyy algoritmien koulutusvaiheeseen. Huolellisesti toteutetut käytännöt datan käsittelyssä ja esikäsittelyssä voivat merkittävästi vähentää vinoumien syntymisen riskiä. Nämä toimet eivät ainoastaan paranna tekoälysovellusten oikeudenmukaisuutta ja luotettavuutta, vaan myös lisäävät käyttäjien ja sidosryhmien luottamusta erityisesti monimuotoisissa ja säännellyissä ympäristöissä.

9. Nousevat suuntaukset tekoälyn oikeudenmukaisuudessa

Tekoälyteknologioiden kehittyessä ne kohtaavat yhä monimutkaisempia tietoa-aineistoja ja muuttuvia sääntely-ympäristöjä. Uudet teknologiat, kuten selitettävä tekoäly (XAI, *explainable AI*) ja hajautettu oppiminen (*federated learning*), tarjoavat mahdollisuuksia parantaa tekoälyn oikeudenmukaisuutta lisäämällä läpinäkyvyyttä päätöksentekoprosesseihin sekä mahdollistamalla hajautetun datankäsittelyn, joka tukee yksityisyydensuojaa. Tulevaisuuden haasteita ovat muun muassa:

- **Monimutkaiset datavuorovaikutukset:** Suurten tietomassojen aikakaudella suurien ja monimutkaisten aineistojen sisäisten vuorovaikutusten ja riippuvuuksien ymmärtäminen on entistä haastavampaa, mutta samalla keskeistä oikeudenmukaisuuden varmistamiseksi.
- **Mukautuvat sääntelyvaatimukset:** Yleisen tietoisuuden lisääntyessä tekoälyn liittyvistä vinoumista, sääntelyviranomaiset tulevat todennäköisesti asettamaan tiukempia oikeudenmukaisuusvaatimuksia. Tämä edellyttää, että tekoälyjärjestelmät ovat mukautuvia ja läpinäkyviä oikeudenmukaisuustoimissaan.
- **Selitettävyyys ja tulkittavuus:** Selitettävyyys ja tulkittavuus tarkoittavat sitä, että tekoälyjärjestelmän toiminta ja päätökset voidaan esittää ymmärrettävästi ihmisille. Tavoitteena on varmistaa, että sidosryhmät ymmärtävät, miten tekoäly tekee päätöksiä — mikä on keskeistä luottamuksen ja vastuun kannalta. Tämä kehityssuunta sisältää menetelmien ja työkalujen kehittämistä, joiden avulla tekoälyn päätöksentekoa voidaan havainnollistaa ja selittää. Esimerkiksi piirteiden tärkeys pisteytys, päätöspuukaaviot ja malliriippumattomat selitystekniikat auttavat avaamaan monimutkaisten mallien, kuten syvien neuroverkkojen, toimintaa.
- **Vastahyökkäykset ja puolustuskeinot:** Vastahyökkäykset tarkoittavat tekoälyjärjestelmän syötteiden manipulointia siten, että järjestelmä tekee virheellisiä päätöksiä. Näiden hyökkäysten torjumisen tavoitteena on tehdä

tekoälyjärjestelmistä entistä kestävämpiä ja turvallisempia, sekä suojata niitä haitallisilta syötteiltä, jotka voisivat hyödyntää mallin vinoumia tai heikkouksia. Puolustuskeinoin kuuluu muun muassa hyökkäyskoulutus, jossa malli altistetaan harjoitteluvaiheessa manipuloiduille esimerkeille oppiakseen vastustamaan niitä. Muita lähestymistapoja ovat säännölliset tietoturva-arvioinnit sekä mallien päivittäminen uusien hyökkäystapojen tunnistamiseksi ja torjumiseksi.

- **Moniulotteinen oikeudenmukaisuus:** Moniulotteinen oikeudenmukaisuus viittaa vinoumien käsittelymiseen, kun ne liittyvät samanaikaisesti useisiin väestöllisiin ja sosioekonomisiin ominaisuuksiin, kuten esimerkiksi etniseen taustaan, sukupuoleen, ikään ja vammaisuuteen. Tavoitteena on varmistaa kokonaisvaltainen oikeudenmukaisuus, joka huomioi ihmisten monikerroksisen identiteetin ja sen vaikutukset. Tämä edellyttää moniulotteisten oikeudenmukaisuusmittarien kehittämistä sekä moniulotteisia data-analyysimenetelmiä vinoumien arvioimiseksi ja vähentämiseksi useiden ominaisuuksien risteyskohdissa.
- **Oikeudenmukaisuus kehittyvissä teknologioissa:** Tekoälyn integroituminen kehittyviin teknologioihin, kuten autonomisiin ajoneuvoihin, älykkääseen terveydenhuoltoon ja muut tietotekniset verkkolaitteet (IoT), tekee oikeudenmukaisuuden varmistamisesta näissä sovelluksissa entistä tärkeämpää. Tavoitteena on sisällyttää oikeudenmukaisuus jo teknologian suunnittelun alkuvaiheisiin, jotta voidaan ehkäistä vinoumia, joilla voisi olla laajamittaisia yhteiskunnallisia vaikutuksia. Tämä sisältää vaikutustenarviointia, vinoumatarkastuksia sekä kunkin teknologian käyttötarkoitukseen sovitettuja oikeudenmukaisuusohjeita.
- **Eettiset näkökulmat ja hallinto:** Tekoälyn eettiset näkökohdat ja hallinto tarkoittavat puitteiden ja ohjeistojen luomista, joiden avulla varmistetaan, että tekoälyä kehitetään ja käytetään yhteiskunnallisten arvojen ja eettisten periaatteiden mukaisesti. Tavoitteena on tarjota sääntelyyn ja eettiseen valvontaan perustuva ohjausmekanismi, joka ohjaa tekoälyn kehitystä ja käyttöä. Tämä tarkoittaa tekoälyn eettisten ohjeiden laatimista, eettisten toimikuntien perustamista sekä yhteistyötä lainsäätäjien kanssa oikeudenmukaisuuden, läpinäkyvyyden ja vastuun varmistamiseksi.
- **Ihmiskeskeinen suunnittelu:** Ihmiskeskeinen suunnittelu pyrkii luomaan tekoälyjärjestelmiä, jotka asettavat ihmisen hyvinvoinnin, arvot ja eettiset näkökulmat etusijalle. Tavoitteena on varmistaa, että tekoäly tukee ihmisten toimintaa ja parantaa elämänlaatua ilman, että se syrjäyttää tai heikentää ihmisen roolia päätöksenteossa. Tämä edellyttää vuorovaikutusta erilaisten käyttäjäryhmien kanssa jo suunnitteluvaiheessa, palautesilmukoiden hyödyntämistä tekoälyn jatkuvassa kehittämisessä sekä sitä, että tekoäly toimii **ihmisen tukena, ei korvaajana**.
- **Algoritmissen vinouman lievittäminen:** Algoritmissen vinouman lievittäminen tarkoittaa tekoälyalgoritmien sisäisten vinoumien

tunnistamista ja korjaamista epäoikeudenmukaisten lopputulosten estämiseksi. Tavoitteena on jatkuvasti kehittää ja tarkentaa tekoälymalleja siten, että ne olisivat mahdollisimman objektiivisia ja puolueettomia. Tämä sisältää vinoumien tunnistus- ja korjaustekniikoiden hyödyntämisen mallin koulutusprosessin aikana, kuten koulutusdatan uudelleenpainottamisen, malliparametrien säätämisen tarkkuuden ja oikeudenmukaisuuden tasapainottamiseksi sekä ulkoisten vinoumien lievittämiseen tarkoitettujen algoritmien käyttöä.

Nämä kehityssuunnat ja haasteet korostavat tekoälyn jatkuvasti muuttuvaa luonnetta sekä tarvetta jatkuvalla innovoinnille ja valvonnalle oikeudenmukaisuuden varmistamiseksi sitä mukaa, kun tekoäly sulautuu yhä tiiviimmin osaksi ihmisten arkea.

10. Johtopäätökset

Tekoälyn oikeudenmukaisuuden ja vinoumien tarkastelu on ratkaisevan tärkeää, kun tekoälyä integroidaan yhä laajemmin osaksi jokapäiväistä elämää ja kriittistä infrastruktuuria. Eri vinoumien määritelmistä ja vaikutuksista tekoälyjärjestelmissä aina niiden tunnistamiseen, mittaamiseen ja lievittämiseen tähtääviin strategioihin käy selväksi, että oikeudenmukaisuus tekoälyssä on monitahoinen haaste, joka vaatii laaja-alaista ja jatkuvaa huomiota.

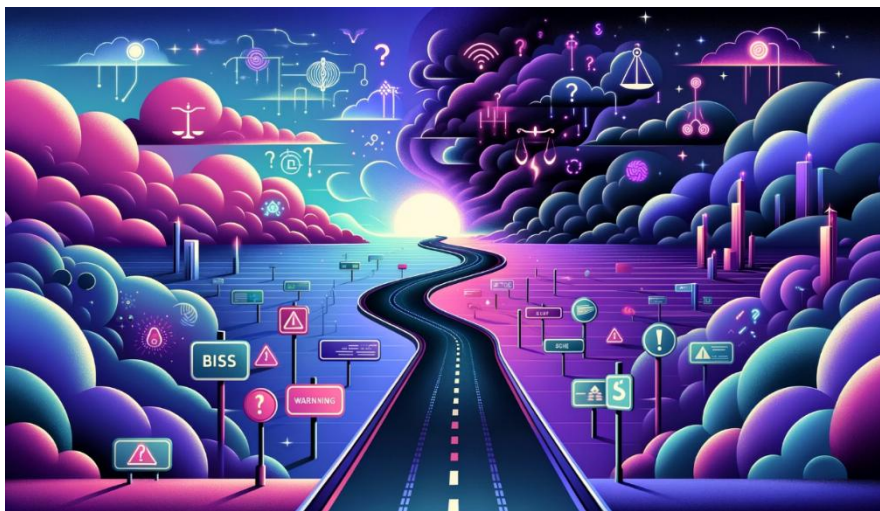


IMAGE SOURCE | Generated by DALL-E

Vinoumien tyyppien ja vaikutusten ymmärtäminen:

Ensimmäinen askel vinoumien lievittämiseksi on tunnistaa erilaiset vinoumatyypit, kuten otantavinouma, algoritmisen vinouma ja vahvistusvinouma, sekä ymmärtää niiden potentiaali ylläpitää tai pahentaa eriarvoisuutta. Reaaliaikaiset esimerkit, kuten kasvojentunnistuksen epätarkkuudet ja vinoutuneet rekrytointialgoritmit, ovat osoittaneet, kuinka haitallisia valvomattomat tekoälyvinoumat voivat olla.

Oikeudenmukaisuuden varmistamisen strategiat:

Tehokkaita keinoja oikeudenmukaisuuden edistämiseksi ovat esimerkiksi monimuotoisen datan kerääminen, osallistava otanta ja tekoälyjärjestelmien säännöllinen seuranta. Näillä toimilla pyritään kehittämään tekoälymalleja, jotka ovat edustavia ja oikeudenmukaisia sekä vähentävät systemaattisia virheitä, jotka voivat asettaa tietyt ryhmät epäedulliseen asemaan.

Nousevat suuntaukset ja haasteet:

Tekoälyn oikeudenmukaisuuden kenttä kehittyy jatkuvasti, ja uusia haasteita nousee esiin, kuten oikeudenmukaisuuden varmistaminen uusissa teknologioissa sekä puolustuskeinojen kehittäminen adversaarisia hyökkäyksiä vastaan. Samalla korostuvat myös selitettävyys ja ihmiskeskeinen suunnittelu, jotka ovat keskeisiä tekijöitä luotettavan tekoälyn rakentamisessa.

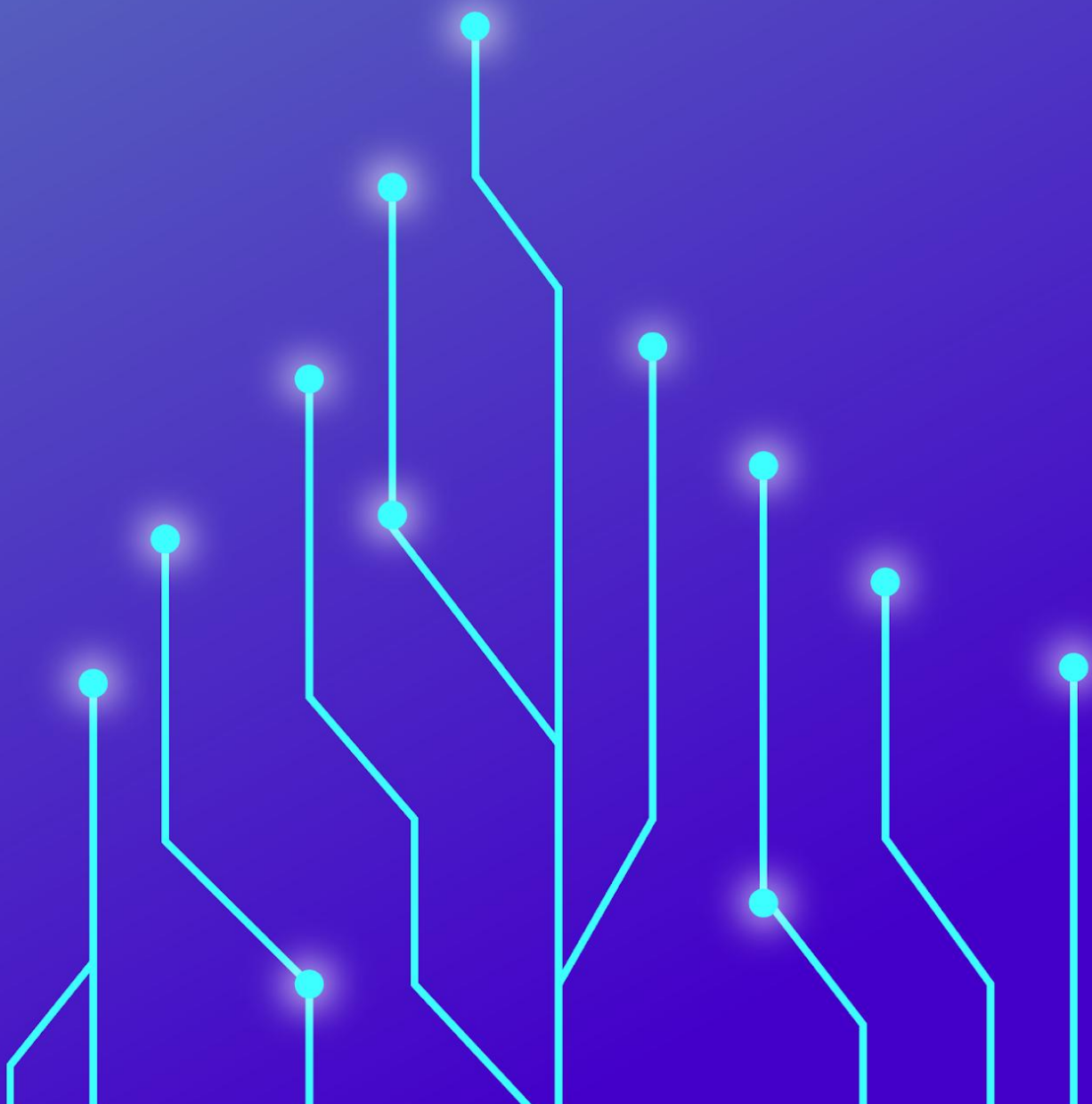
Eettiset näkökohdat ja hallinto:

Eettisen hallinnon merkitystä tekoälyn kehityksessä ei voi liioitella. Vahvojen eettisten ohjeistojen ja hallintorakenteiden luominen on keskeistä, jotta tekoälyteknologioiden käyttöönotto tapahtuisi vastuullisesti ja yhteiskunnallisia arvoja kunnioittaen.

Jatkuva oppiminen ja sopeutuminen:

Koska sekä tekoälyteknologiat että yhteiskunnalliset normit muuttuvat, myös lähestymistapojemme oikeudenmukaisuuteen on kehityttävä. Kehittäjien, päättäjien ja muiden sidosryhmien on jatkuvasti koulutauduttava, oltava valppaina ja sopeuduttava muuttuviin olosuhteisiin varmistaakseen, että tekoälyjärjestelmät pysyvät oikeudenmukaisina ja hyödyttävät kaikkia.

Kurssi 5 | Tapaustutkimukset ja projektit



Sisältö

1. Johdanto	115
2. Erityyppiset vinoumat	116
3. Työkalut tekoälyratkaisun eettisyyden arviointiin	118
4. Projekt: Tekoälyratkaisu rekrytinnissa	119
5. Mahdollisten vinoumien tunnistaminen	123
6. Ehdokkaiden seulonta ja esivalinta: aineiston määrittely	134
7. Algoritmissen mallin kehittäminen	136
8. Kuinka mitata oikeudenmukaisuutta koneoppimiseratkaisuissa	140
9. Ratkaisun käyttöönotto	145
10. Toiminta ja seuranta	146
11. Johtopäätös	147
12. Lähteet	148

1. Johdanto

Tässä kurssiosiossa keskitytään vinoumien tarkasteluun käytännön kontekstissa. Projektina toimii hypoteettinen tilanne, jossa kansainvälinen yritys kehittää tekoälyratkaisua rekrytointiin. Tätä projektia käsitellään viimeisessä interaktiivisessa oppimistilaisuudessa yhdessä opiskelijoiden kanssa. Tarkoituksena on herättää keskustelua opiskelijoiden oppimisen tukemiseksi ja rohkaista heitä ilmaisemaan sekä soveltamaan aiemmilla kerroilla oppimaansa.

Algoritminen vinouma tarkoittaa algoritmisten järjestelmien tuottamia epäoikeudenmukaisia tai syrjiviä lopputuloksia. Tämä tapahtuu silloin, kun algoritmi suosii tai syrjii systemaattisesti tiettyjä yksilöitä tai ryhmiä tietyillä ominaisuuksilla – kuten etnisen taustan, sukupuolen tai sosioekonomisen aseman – perusteella. Koneoppimisalgoritmit käyttävät monenlaista dataa, kuten koulutusdataa, joiden pohjalta ne oppivat malleja, joita sovelletaan muihin yksilöihin tai ryhmiin ennusteiden tekemiseksi. Ongelmana on kuitenkin se, että algoritmit saattavat kohdella samankaltaisia yksilöitä tai tapauksia eri tavoin, mikä voi johtaa inhimillisten vinoumien toistamiseen tai jopa voimistamiseen – erityisesti suojattujen ryhmien kohdalla (Friedman & Nissenbaum, 1996; Lange & Duarte, 2018; Lee et al., 2019). Algoritmisten vinoumien syiden ja vaikutusten ymmärtäminen on keskeistä eettisten ja luotettavien tekoälyratkaisujen kehittämisessä.

Algoritmin vinoumien vaikutukset

Algoritminen vinouma voi ilmetä monin eri tavoin ja vaikuttaa eri ryhmiin eri tavoin. Lee, Resnick ja Barton (2019) ovat koonneet esimerkkejä, jotka osoittavat, kuinka vinouma voi olla peräisin useista lähteistä ja vaikuttaa ryhmiin joko tahattomasti tai tarkoituksellisesti.

- Sanojen assosiaatiot voivat heijastaa vinoumaa. Princetonin yliopiston tutkijat havaitsivat, että eurooppalaiset nimet koettiin myönteisemmin kuin afroamerikkalaiset nimet, ja sanat kuten "nainen" ja "tyttö" yhdistettiin useammin taiteisiin kuin tieteeseen ja matematiikkaan (Lee ym., 2019). Kuten kurssin alussa esitellyssä esimerkissä osoitettiin, assosiaatiot kuten 'sairaanhoitaja' naisille ja 'insinööri' miehille ovat usein seurausta vinoutuneesta datasta, joka on sisällytetty hakualgoritmeihin ja heijastaa yhteiskunnallisia stereotyyppioita. Tämä ylläpitää vinoumaa tekoälyssä ja koneoppimisessa, sillä ne perustuvat olemassa olevaan verkkosisältöön. Tämä muodostaa merkittävän ja jatkuvan haasteen tekoälyn ja koneoppimisen alalla.
- Verkkopohjaiset rekrytointityökalut. Amazon lopetti rekrytointialgoritmin käytön sukupuolivinouman vuoksi. Tekoälyohjelmisto rankaisi ansioluetteloita, joissa esiintyi sana "naisten", ja alensi niiden pisteytystä, jotka olivat peräisin naisten korkeakouluista.

- Myös verkkomainokset osoittavat vinoumaa. Harvardin tutkija Latanya Sweeney havaitsi, että afroamerikkalaisiin nimiin liittyvät haut toivat todennäköisemmin esiin mainoksia rikosrekisteripalveluista verrattuna valkoisiin nimiin.
- Kasvojentunnistusteknologia voi myös olla vinoutunutta. MIT:n tutkija Joy Buolamwini havaitsi, että kaupalliset järjestelmät epäonnistuvat usein tunnistamaan tummempia ihonsävyjä, koska niitä on koulutettu pääasiassa vaaleaihoisten ja miespuolisten kasvojen perusteella.
- Jopa rikosoikeuden algoritmit voivat olla vinoutuneita. Tuomareiden käyttämän COMPAS-algoritmin, jonka avulla ennustetaan takuusijoituksia, havaittiin olevan vinoutunut afroamerikkalaisia kohtaan, kuten ProPublica raportoi.

2. Erityyppiset vinoumat

Friedmanin ja Nissenbaumin (1996) mukaan algoritmiset vinoumat voidaan luokitella kolmeen päätyyppiin.

- Ensinnäkin, ennalta olemassa olevat vinoumat (**pre-existing bias**) viittaa vinoumiin, jotka ovat järjestelmässä jo ennen sen luomista, joko tarkoituksellisesti tai tahattomasti. Nämä vinoumat voivat joutua järjestelmään harkittujen toimien tai tiedostamattomien tekojen kautta, toisinaan jopa hyvistä aikomuksista huolimatta.
- Toiseksi, tekninen vinouma (**technical bias**) syntyy suunnitteluprosessin aikana esiintyvistä teknisistä rajoitteista tai ratkaisuksista. Useat suunnittelun osa-alueet voivat vaikuttaa teknisen vinouman syntyyn.
- Kolmanneksi, nouseva vinouma (**emergent bias**) ilmenee järjestelmän todellisessa käytössä oikeiden käyttäjien kanssa. Tämän tyyppinen vinouma ilmestyy yleensä vasta suunnitteluprosessin jälkeen, usein yhteiskunnallisen tiedon, väestön tai kulttuuristen arvojen muutosten seurauksena. (Friedman & Nissenbaum, 1996)

Esimerkkejä vinoumista koneoppimisen aineistoissa

Koneoppimismallit perustuvat vahvasti koulutusdataan, mutta ihmisen rooli datan valinnassa ja käsittelyssä voi aiheuttaa monenlaisia vinoumia. Yksi yleinen esimerkki on **raportointivinouma**, jossa aineistossa esiintyvien tapahtumien yleisyys ei vastaa todellista maailmaa. Tämä voi johtaa siihen, että mallit eivät suoriudu tavallisista tilanteista hyvin, koska ne painottuvat liikaa epätavallisten tilanteiden kuvaamiseen. (Google for Developers, 2022)

Toinen vinouman muoto on ryhmän **attribuutiovinouma**, jossa suositetaan omaa ryhmää tai stereotypisoidaan muita ryhmiä niiden ominaisuuksien perusteella. **Implisiittinen vinouma**, joka juontaa juurensa yksilöllisistä kokemuksista ja ajatusmalleista, voi edelleen vaikeuttaa tilannetta – se voi esimerkiksi johtaa **vahvistusvinoumaan** tai **tutkijavinoumaan** mallin koulutuksen aikana. (Google for Developers, 2022)

Automaatiovinouma tarkoittaa sitä, että automaattisia järjestelmiä suositetaan ihmisten tekemien arvioiden sijaan, riippumatta niiden tarkkuudesta.

Valintavinouma taas syntyy, kun data kerätään ei-edustavalla tavalla. Molemmat ilmiöt muodostavat merkittäviä haasteita pyrittäessä vähentämään vinoumia koneoppimismalleissa. (Google for Developers, 2022)

Todellisia esimerkkejä

Kun algoritmisten vinoumien teoreettinen viitekehys ja sen vaikutukset on ymmärretty, on aiheellista tarkastella todellisia esimerkkejä, jotka havainnollistavat vinoumien seurauksia. Ensimmäinen tapaus koskee TikTokin haitallista algoritmia, ja toinen käsittelee kasvojentunnistusteknologian eettisiä ongelmia.

TikTokin haitallinen algoritmi

Suomen Yleisradion (Yle) tekemä paljastava tutkimus toi esiin TikTokin algoritmin haitalliset vaikutukset. Yle toteutti kokeen, jossa seurattiin sisältöä, jota tarjottiin fiktiiviselle 13-vuotiaalle tytölle nimeltä Ella, jolla oli masennusta ja kehonkuvaongelmia. (YLE, 2023)

Tutkimuksessa Ylen datatieteilijät käyttivät TikTokia Ellan profiililla viiden tunnin ajan. Aluksi Ella näki iloisia videoita ruoasta, lemmikeistä ja vitseistä. Jo muutamassa minuutissa algoritmi alkoi näyttää mielenterveyteen liittyvää sisältöä, kuten masennuslääkkeitä. Ajan myötä sisällön sävy synkkeni huomattavasti: TikTok alkoi näyttää videoita itsemurhasta, itsensä vahingoittamisesta ja syömishäiriöistä. Tutkimuksen lopussa peräti 95 % Ellalle näytetystä sisällöstä arvioitiin haitalliseksi – mukaan lukien syömishäiriöiden ihannoitiin ja ohjeet ruoan rajoittamiseen. (YLE, 2023)

Sosiaalipsykologi Suvi Uski nosti esiin algoritmin huolestuttavan toimintamallin: se asettaa käyttäjien sitoutumisen etusijalle, riippumatta sisällön mahdollisista haittavaikutuksista. Tämä paljastus korostaa TikTokin algoritmien merkittäviä eettisiä ongelmia. (YLE, 2023)

Viattomalta vaikuttanut selailu johti nopeasti vaaralliseen kierteseen, joka heijasti Ellan haavoittuvaa tilaa. Erityisen huolestuttavaa oli algoritmin kyky tunnistaa nopeasti Ellan kiinnostus masennukseen ja epäterveisiin

käyttäytymismalleihin – ja sen taipumus vahvistaa näitä taipumuksia sen sijaan, että se olisi ohjannut häntä kohti myönteisiä ja tukevia resursseja. Tapaus nostaa esiin tärkeitä kysymyksiä algoritmien suunnittelijoiden ja alustojen ylläpitäjien eettisestä vastuusta käyttäjien hyvinvoinnin turvaamisessa. Ellan tapaus korostaa kiireellistä tarvetta lisätä algoritmien järjestelmien läpinäkyvyyttä, vastuullisuutta ja eettistä harkintaa – erityisesti kun kyseiset järjestelmät vaikuttavat merkittävästi käyttäjien käyttäytymiseen ja mielenterveyteen.

Epäeettinen kasvojentunnistusdatan keruu

Kasvojentunnistusteknologia on yleistynyt nopeasti eri aloilla, terveydenhuollosta lainvalvontaan (Javaid, 2024). Sen laaja käyttöönotto on kuitenkin herättänyt merkittäviä eettisiä huolia, erityisesti liittyen datan keruuseen sekä yksityisyyden ja kansalaisoikeuksien loukkaamiseen. Yksi merkittävä esimerkki epäeettisestä kasvojentunnistusdatan keruusta tuli ilmi *The Washington Postin* raportissa 7. heinäkuuta 2019. Raportin mukaan liittovaltion viranomaiset, kuten FBI ja ICE, olivat käyttäneet osavaltioiden ajokorttitietokantoja kasvojentunnistuksen tarkoituksiin, selaamalla miljoonien amerikkalaisten kuvia ilman heidän suostumustaan tai tietoaan (Harwell, 2019a).

Tämä huolestuttava paljastus tuo esiin kasvojentunnistusteknologian laajemmat riskit vinoumalle ja syrjinnälle. *The Washington Post* raportoi 19. joulukuuta julkaistusta liittovaltion tutkimuksesta, joka vahvisti kasvojentunnistusjärjestelmien rodullisen vinouman esiintyvyyden. Tutkimuksessa havaittiin, että aasialaiset ja afroamerikkalaiset henkilöt joutuivat jopa 100 kertaa todennäköisemmin väärin tunnistetuiksi kuin valkoiset miehet – riippuen käytetystä algoritmista ja hakutypistä (Harwell, 2019b).

3. Työkalut tekoälyn eettisyyden arviointiin

Tekoälyjärjestelmien kehityksessä – erityisesti aloilla, joilla vinoumat voivat olla herkkiä, kuten rekrytoinnissa – on tärkeää hyödyntää työkaluja, jotka on suunniteltu tunnistamaan ja arvioimaan mahdollisia vinoumia. Nämä arviointityökalut ovat keskeisiä sen varmistamisessa, että tekoälysovellukset toimivat oikeudenmukaisesti ja läpinäkyvästi. Näiden työkalujen avulla kehittäjät voivat paremmin ymmärtää, miten vinoumat voivat vaikuttaa järjestelmän lopputuloksiin, ja ryhtyä ennaltaehkäiseviin toimiin niiden vaikutusten vähentämiseksi. Tämä lähestymistapa parantaa tekoälyn luotettavuutta ja oikeudenmukaisuutta sekä lisää käyttäjien luottamusta osoittamalla sitoutumista eettisiin käytäntöihin.

Suunnittelutyökalut ovat erityisen tärkeitä eettisessä tekoälykehityksessä, koska ne mahdollistavat kokonaisvaltaisen ymmärryksen eri sidosryhmien näkökulmista. Ne auttavat kehittäjiä huomioimaan empaattisesti, miten

tekoälyteknologiat vaikuttavat eri käyttäjäryhmiin. Näiden ryhmien vuorovaikutusten, käyttäytymisen ja tarpeiden kartoittaminen auttaa tunnistamaan ja lieventämään tekoälyjärjestelmiin sisältyviä mahdollisia vinoumia. Laaja sidosryhmien huomioiminen edistää inklusiivisuutta ja varmistaa, että tuotteiden kehitys linjautuu eettisten periaatteiden kanssa ennaltaehkäisemällä epäoikeudenmukaisia tai syrjiviä seurauksia. Tässä projektitapauksessa erilaisia työkaluja esitellään suunnitteluprosessin eri vaiheissa.

Vaikka arviointityökaluja on olemassa runsaasti, tämä kurssi keskittyy esittelemään vain kolme niistä. Projektissa käytetyt suunnittelutyökalut on kuvattu tarkemmin projektiselosteessa, käsikirjan luvussa E: *Potential Biases*.

UNESCO:n Ethical Impact Assessment -työkalu, kehitetty vuonna 2023, tarjoaa jäsenmäärältään lähestymistavan tekoälyjärjestelmien eettisten vaikutusten arviointiin. Työkalu sisältää monipuolisia kysymyksiä, sekä avoimia että monivalintakysymyksiä, jotka ohjaavat käyttäjää kattavan arviointiprosessin läpi. Työkalu korostaa eettisiä näkökulmia jo prosessin alkuvaiheessa alkaen rajauskysymyksistä ja edeten UNESCO:n periaatteisiin. Se käsittelee keskeisiä teemoja, kuten sidosryhmien vaikutuksia, tiimin vastuuta, turvallisuutta, yksityisyyttä, läpinäkyvyyttä, selitettävyyttä, vastuullisuutta ja kestävyyttä. (UNESCO, 2023)

UNESCO:n työkalun osio *Fairness, Non-Discrimination, and Diversity* keskittyy erityisesti algoritmisen vinouman tunnistamiseen – erityisesti tekoälyjärjestelmissä käytetyn datan osalta. Työkalu pureutuu vastuisiin ja rooleihin varmistaa, että kaikki järjestelmään sisältyvät vinoumat tunnistetaan ja käsitellään asianmukaisesti. (UNESCO, 2023)

Assessment List for Trustworthy AI (ALTAI) on Euroopan komission asettaman tekoälyn korkean tason asiantuntijaryhmän (AI HLEG) kehittämä työkalu. Se tarjoaa jäsenmäärältään viitekehyksen tekoälyn luotettavuuden arviointiin. Työkalu määrittelee seitsemän vaatimusta luotettavalle tekoälylle ja tarjoaa ohjeet arvioinnin suorittamiseen sekä tunnistaa keskeiset sidosryhmät, jotka tulisi ottaa mukaan prosessiin. (ALTAI, 2020)

ALTAn osio *Diversity, Non-discrimination and Fairness* keskittyy epäoikeudenmukaisen vinouman välttämiseen. Se arvioi algoritmista vinoumaa kohdennettujen kysymysten kautta tarkastelemalla syötedataa ja algoritmin suunnittelua. Samalla se edistää osallistavuutta, saavutettavuutta ja sidosryhmien osallistamista oikeudenmukaisuuden ja läpinäkyvyyden varmistamiseksi tekoälyjärjestelmissä. (ALTAI, 2020)

Mario Sosa Hidalgon kehittämä *Design of an Ethical Toolkit for the Development of AI Applications* tarjoaa innovatiivisen, konkreettisen ja visuaalisen lähestymistavan eettisyyden integroimiseksi tekoälysovellusten kehitysprosessiin. Työkalu keskittyy vinoumien tunnistamiseen ja suosittelee

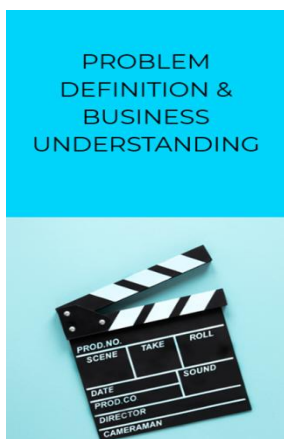
kehittäjille strategioita epäoikeudenmukaisuuden minimoimiseksi, jotta tekoälyjärjestelmistä tulisi mahdollisimman puolueettomia. (Hidalgo, 2019)

4. Projekti: Tekoälyratkaisu rekrytinnissa

Tässä osiossa keskitytään algoritmisen vinouman käsittelyyn tekoälyjärjestelmissä. Se esittelee projektin, joka perustuu hypoteettiseen tilanteeseen, jossa kansainvälinen yritys kehittää tekoälyratkaisua rekrytointiin.

Projektin kulkua ohjaa kertomus, joka esitetään sinisellä rajatuissa tekstilaatikoissa. Nämä laatikot toimivat keskeisinä lähtötietoina projektille. Lisäksi osio tarjoaa kattavaa taustatietoa, joka auttaa ymmärtämään kontekstia ja tekoälysovelluksen kehittämistä rekrytointitarkoituksiin.

Tavoitteena on aktivoida opiskelijoita tunnistamaan ja lieventämään vinoumia, joita voi syntyä tekoälyteknologioiden suunnittelun ja käyttöönoton aikana.



Tilanteen taustoitus

Kansainvälinen teollisuusyritys kohtaa tehottomuutta ja epävarmuutta rekrytointiprosessissaan, mikä kuormittaa henkilöstöhallintoa kuluttamalla runsaasti aikaa ja resursseja.

Yritys ennakoi tulevaa kasvua ja tiedostaa samalla kiireellisen tarpeen palkata erilaisia asiantuntijoita, kuten insinöörejä, markkinoinnin ammattilaisia, talousasiantuntijoita, koodareita ja projektipäälliköitä.

Kova kilpailu osaavista työntekijöistä ja rekrytointivirheisiin liittyvät korkeat kustannukset korostavat onnistuneen rekrytinnin merkitystä. Tilanne lisää painetta HR-tiimille, jonka tehtävänä on varmistaa rekrytointiprosessin sujuvuus ja onnistuminen.

Tämän seurauksena yritys selvittää mahdollisuutta kehittää tekoälyratkaisu, joka on räätälöity sen rekrytointitarpeisiin. Tekoälyn integroinnin ensisijaisena

Tekoälyä (AI) integroidaan yhä enenevässä määrin rekrytointiprosesseihin, ja se tarjoaa monia mahdollisuuksia rekrytointikäytäntöjen parantamiseen. Vaikka tekoälyn käyttö rekrytinnissa tuo mukanaan merkittäviä etuja, siihen liittyy myös tiettyjä haasteita. On tärkeää, että organisaatiot tunnistavat ja harkitsevat huolellisesti molempia puolia. Ymmärtämällä sekä mahdolliset hyödyt että haitat yritykset voivat tehdä perusteltuja päätöksiä siitä, miten tekoälyteknologioita kannattaa ottaa käyttöön. Tällainen tasapainoinen lähestymistapa varmistaa, että

samalla kun tekoälyn avulla tehostetaan ja kehitetään rekrytointia, myös siihen liittyvät seuraukset ja riskit otetaan asianmukaisesti huomioon.

PROBLEM
DEFINITION &
BUSINESS
UNDERSTANDING



Ongelman ymmärtäminen

Yritys aloittaa taustatutkimuksen tekoälyratkaisun kehittämiseksi ja muodostaa sisäisen tiimin, jonka tehtävänä on tunnistaa keskeiset haasteet nykyisessä rekrytointiprosessissa. Tiimi haastattelee useita sidosryhmiä, mukaan lukien HR-osaston henkilöstöä, äskettäin palkattuja työntekijöitä sekä esihenkilöitä, jotka ovat hiljattain perehdyttäneet uusia tiimiläisiä. Lisäksi tiimi kerää tietoa nykyisistä rekrytointikustannuksista ja analysoi rekrytointien onnistumisastetta.

Mahdollisten työntekijöiden käsitysten ymmärtäminen tekoälyn käytöstä rekrytoinnissa on yritykselle ratkaisevan tärkeää. Tavoitteena on välttää maineen tai brändin vahingoittuminen. Tämän saavuttamiseksi yritys on sitoutunut varmistamaan, että tekoälyn käyttöönotto saa laajaa hyväksyntää ja noudattaa sekä eettisiä että oikeudellisia periaatteita.

Keskustelujen ja selvitysten jälkeen tekoälypohjaisen rekrytointiratkaisun hyödyt ja haitat on listattu seuraavasti:

Tekoälyn hyödyt rekrytoinnissa

Tekoäly mullistaa rekrytointiprosesseja automatisoimalla rutiinitehtäviä, mikä vapauttaa rekrytoijien aikaa strategisempiin tehtäviin. Tämä automaatio parantaa paitsi hakemusten käsittelyn tehokkuutta myös lyhentää merkittävästi rekrytointiprosessiin kuluvaa aikaa. (Albaroudi ym., 2024; www.recruiter.com, 2023)

Lisäksi tekoälyn kyky analysoida laajoja tietoaaineistoja mahdollistaa parhaiten sopivien ehdokkaiden tunnistamisen, mikä parantaa rekrytointien laatua. Tämä voi vähentää vaihtuvuutta varmistamalla paremman yhteensopivuuden työn ja työntekijän välillä. Tekoälyn tarjoama nopea palaute ja ajantasaiset päivitykset tehostavat viestintää ja pitävät hakijat ajan tasalla valintaprosessin aikana. Tämä antaa rekrytoijille enemmän aikaa keskittyä valittuihin ehdokkaisiin, mikä parantaa hakijakokemusta. (Arivu Recruitment and Consulting, 2023; Sheard, 2022; www.recruiter.com, 2023)

Tekoälyn käyttö rekrytoinnissa voi myös vähentää kustannuksia vähentämällä manuaalisen työn tarvetta. (www.recruiter.com, 2023) Lisäksi tekoäly pyrkii vähentämään inhimillisiä vinoumia, mikä lisää objektiivisuutta,

johdonmukaisuutta ja oikeudenmukaisuutta rekrytoinnissa. Tämä on erityisen hyödyllistä ryhmille, joilla on haasteita työmarkkinoille pääsyssä, koska tekoäly voi parantaa heidän mahdollisuuksiaan työllistyä. (Bursell & Roumbanis, 2024; Köchling & Wehner, 2020; Sheard, 2022)

Tekoälyn analyysikyky ulottuu myös sosiaalisen median seulontaan, jossa se arvioi ehdokkaiden profiileja kulttuurisen sopivuuden ja työn vaatimusten näkökulmasta. Tekoäly voi myös havaita epäasiallista sisältöä, varmistaen, että rekrytointi on yrityksen arvojen mukaista. Lisäksi tekoälytyökalut auttavat ylittämään kielimuurit, mikä mahdollistaa laajemman ja monimuotoisemman kansainvälisen osaajajoukon tavoittamisen. (Albaroudi ym., 2024)

Lopuksi, tekoälyn odotetaan edistävän työpaikkojen monimuotoisuutta poistamalla vinoumia rekrytointiprosessista. (Sheard, 2022) Kaikki nämä edut korostavat tekoälyn mullistavaa vaikutusta rekrytointiin ja lupaavat tehokkaampaa, oikeudenmukaisempaa ja osallistavampaa työnhakua.

Tekoälyn huolet rekrytoinnissa

Vaikka tekoäly tuo lukuisia tehokkuusetuja rekrytointiin, sen käyttöönottoon liittyy myös huolia ihmiskontaktin vähentymisestä ja vuorovaikutuksen persoonattomuudesta. Tämä voi tehdä rekrytointikokemuksesta mekaanisen tuntuisen, mikä voi vieraannuttaa hakijoita ja heikentää kokemuksen laatua. (www.recruiter.com, 2023)

Lisäksi vinoumien ja syrjinnän riski on merkittävä huolenaihe. Ellei tekoälyä suunnitella ja kouluteta huolellisesti tai jos se perustuu vinoutuneeseen opetusaineistoon, se voi tahtomattaan tuoda rekrytointiprosessiin uusia vinoumia. Tämä voi heikentää monimuotoisuutta, osallisuutta ja oikeudenmukaisuutta rekrytoinnissa. (Albassam, 2023; www.recruiter.com, 2023)

Tietosuojan ja tietoturvaan liittyvät huolenaiheet ovat myös keskeisiä, sillä tekoälyjärjestelmät käsittelevät valtavia määriä henkilötietoja. Tämä herättää kysymyksiä yksityisyydensuojasta, tietoturvasta ja kyberhyökkäysten riskistä, mikä korostaa vahvojen suojausmenetelmien ja läpinäkyvän tietojen käytön tarvetta. (Albassam, 2023; www.recruiter.com, 2023)

Tekoälyn käyttöönotto rekrytoinnissa vaatii myös resursseja ja investointeja. Erityisesti pienille yrityksille nämä kustannukset voivat muodostaa esteen tekoälyn hyödyntämiselle rekrytointiprosesseissa. (www.recruiter.com, 2023)

Huolta herättää myös inhimillisen harkinnan puute ja mahdolliset tarkkuusongelmat. Jos rekrytoinnissa luotetaan yksinomaan tekoälyyn ilman ihmisen valvontaa, voidaan jättää huomaamatta päteviä ehdokkaita esimerkiksi puutteellisen datan tai siirrettävien taitojen tunnistamattomuuden vuoksi. Tekoälyn tarkkuutta voivat myös heikentää hakijat, jotka muokkaavat profiiliaan

näyttääkseen pätevämmiltä. (Albassam, 2023; Arivu Recruitment and Consulting, 2023)

Tekoälyn käyttöä rekrytoinnissa säätelevä oikeudellinen kehys on jatkuvasti kehittyvä, ja yritysten on mukautettava toimintaansa oikeudenmukaisuuden varmistamiseksi. Tämä korostaa jatkuvaa valppautta ja mukautumiskykyä tekoälytyökalujen käytössä. (Fernandes, 2021; www.recruiter.com, 2023)

Tekoälyn rajoitukset kulttuurisen sopivuuden tai tiimityötaitojen arvioinnissa voivat johtaa siihen, että potentiaalisesti erinomaiset ehdokkaat jäävät palkkaamatta. Albassam (2023) korostaa näiden usein työn onnistumisen kannalta keskeisten, mutta tekoälyn vaikeasti mitattavien tekijöiden huomioimista.

Markkinajohtajuus ja tekoälytyökalujen läpinäkyvyyttä – erityisesti Yhdysvaltalaisissa yksityisyhtiöissä kehitettyjen järjestelmien osalta – vaikeuttavat vinoumien tunnistamista ja ulkoista arviointia. Näiden järjestelmien omistusoikeuteen perustuva luonne estää niiden oikeudenmukaisuuden ja tehokkuuden arviointia ja parantamista. (Sheard, 2022)

Lopuksi huolta herättää automaattisten rekrytointijärjestelmien mahdollinen vaikutus yhteiskunnallisen eriarvoisuuden lisääntymiseen. Nämä järjestelmät voivat toistaa olemassa olevia vinoumia algoritmien sekä verkostojen ja epävirallisten rakenteiden merkityksen kautta työnhaussa. Tämä osoittaa, kuinka tärkeää on tarkasti poistaa tekoälypohjaisista järjestelmistä kaikki vinoumat. (Harvard John A. Paulson School of Engineering and Applied Sciences, 2023)

Kaikki nämä näkökohdat korostavat tekoälyn käytön moniulotteisuutta ja haastavuutta rekrytointiprosessissa, ja painottavat tasapainoisen lähestymistavan tarvetta, jossa huomioidaan tehokkuus, oikeudenmukaisuus ja osallisuus.

5. Mahdollisten vinoumien tunnistaminen

PROBLEM DEFINITION & BUSINESS UNDERSTANDING



Järjestelmän käyttäjien ymmärtäminen

Yksi haastatelluista henkilöistä on rekrytointipäällikkö Maria, joka on työskennellyt yrityksessä useita vuosia. Hän on nähnyt itse rekrytointitehtävien kertymisen ja sen aiheuttaman kasvavan paineen HR-osastolle, joka vain kasvaa, kun rekrytoidaan globaaleja osajia.

Maria on avoinmielinen ja kiinnostunut tekoälyn mahdollisuuksista auttaa rekrytointitehtävissä, vaikka hän ei ole syvällisesti perehtynyt tekoälytekologiaan.

Työkalu: Persoonia

Persoonia käytetään suunnittelussa luomaan yksityiskohtaisia profiileja, jotka edustavat tietyn tyyppisiä käyttäjiä tai käyttäjäryhmiä. Nämä profiilit sisältävät olennaisia piirteitä ilman että turvaudutaan stereotypioihin. Sisällyttämällä realistisia yksityiskohtia, kuten käyttäytymistä, tarpeita ja demografista tietoa, persoonista tulee samaistuttavia ja hyödyllisiä erilaisten käyttäjäryhmien ymmärtämisessä. Ne auttavat suunnittelu- ja kehitystiimejä rakentamaan empatiaa, yhtenäistämään toimintaansa ja kehittämään ratkaisuja, jotka vastaavat tarkasti käyttäjien tarpeita. Persoonat ovat erityisen arvokkaita, koska ne auttavat purkamaan perinteisiä markkinointisegmenttejä, mikä johtaa osallistavampaan ja tehokkaampaan suunnitteluun. (Kumar, 2013; Stickdorn ym., 2018a, 2018b)

HR-ammattilaisen Marian näkökulman vinoumat

Persoonaa arvioitaessa on tärkeää huomioida, että Marian henkilökohtaiset näkemykset voivat vaikuttaa työkalun käyttöönottoa koskeviin suosituksiin ja mielipiteisiin. Tällä hetkellä on kiistelyä, onko tekoäly uhka vai mahdollisuus perinteisille rekrytointirooleille. HR-ammattilaiset saattavat nähdä tekoälypohjaiset rekrytointijärjestelmät uhkana omalle työlleen, mikä voi johtaa vastahakoisuuteen ottaa tekoälytyökaluja käyttöön (Chen, 2023). Marian persoonanperusteella mahdollisia yksilöllisiä vinoumia voivat olla esimerkiksi:

- Ikään liittyvä avoimuus tai epäluulo uutta teknologiaa kohtaan ja sen vaikutus perinteisiin rooleihin.
- Teknologisesti edistyksellisessä kaupungissa, kuten Helsingissä, asuminen voi vaikuttaa hänen näkemyksiinsä ja kallistaa niitä innovatiivisten ratkaisujen suuntaan.
- Hänen guatemalalainen taustansa saattaa vaikuttaa siihen, miten hän kokee tekoälyn vaikutukset monimuotoisuuteen ja osallisuuteen rekrytoinnissa.
- HR-johtamisen koulutus voi vahvistaa uskoa inhimillisen arvion merkitykseen, mikä saattaa aiheuttaa vastahakoisuutta tukeutua tekoälyyn kriittisissä rekrytointitehtävissä.
- Yksin asuminen ja siihen liittyvät henkilökohtaiset kokemukset voivat saada hänet arvostamaan ihmisten välistä vuorovaikutusta, mikä voi lisätä epäluuloa tekoälyn epäpersoonallista luonnetta kohtaan.
- Hänen työssään korostuvat tehokkuus ja henkilökohtainen vuorovaikutus hakijoiden kanssa, mikä voi olla ristiriidassa hänen näkemyksensä kanssa tekoälypohjaisesta rekrytoinnista.
- Turhautuminen suurista työkuormista ja hitaista prosesseista saattaa lisätä hänen avoimuuttaan tekoälyn tarjoamalle tehokkuudelle, mutta pelko vuorovaikutuksen katoamisesta ja etääntymisestä hakijoista voi aiheuttaa epäröintiä.
- Harrastukset kuten uinti, kiipeily ja lukeminen viittaavat tasapainoiseen ja monipuoliseen elämäntapaan, mikä voi näkyä toiveena vivahteikkaammasta lähestymistavasta rekrytointiin.
- Huoli siitä, että tekoäly muuttaa HR:n tehtäviä, voi saada hänet pohtimaan, miten se sopii hänen osaamiseensa ja vaikuttaako se hänen roolinsa tulevaisuuteen.

PROBLEM DEFINITION & BUSINESS UNDERSTANDING



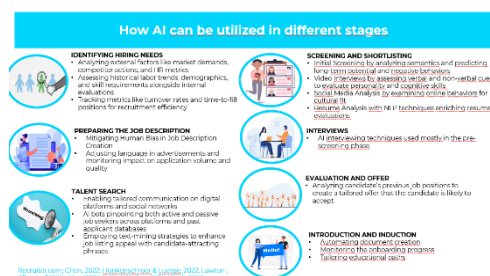
Ymmärrys rekrytointiprosessista

Saadakseen kattavan kuvan rekrytointiprosessista yritys laatii rekrytointimatkan, johon osallistuvat Maria ja rekrytointitiimi. Yhteistyössä linjaesimiesten kanssa he käyvät läpi jokaisen vaiheen tunnistaakseen merkittävät turhautumisen lähteet ja aikaa vievät vaiheet. Lisäksi he tarkastelevat tekoälyn mahdollista hyödyntämistä kussakin vaiheessa rekrytointiprosessin tehokkuuden ja vaikuttavuuden parantamiseksi.

Työkalu: Asiakaspolkukartta

Asiakaspolkukartta on visuaalinen työkalu, jota käytetään muotoilussa kuvaamaan käyttäjän vaiheittaista kokemusta palvelun kanssa. Se esittää käyttäjän näkökulman, yksilöiden mitä tapahtuu kussakin vuorovaikutuksen vaiheessa, mitä kosketuspisteitä on mukana ja millaisia haasteita tai esteitä käyttäjä voi kohdata. Asiakaspolkukartoissa on usein mukana myös käyttäjän tunnevasteita kuvaavia kerroksia, jotka korostavat positiivisia tai negatiivisia kokemuksia matkan eri vaiheissa. Karttoja voidaan käyttää sekä nykyisten kokemusten (nykytilan polkukartat) että toivottujen tulevaisuuden vuorovaikutusten (tulevaisuustilan polkukartat) havainnollistamiseen. Ne voivat vaihdella laajuudeltaan koko kokemuksen yleiskatsauksista tarkasti rajattuihin yksittäisiin hetkiin keskittyviin karttoihin. (Kumar, 2013; Stickdorn et al., 2018a)

Tässä projektissa asiakaspolkukartta esitetään aluksi nykytilan kuvauksena, jota seuraa tulevaisuustilan ennuste tekoälytyökalujen mahdollistamista rekrytointikäytännöistä. Tekoälyn käyttöä rekrytointiprosessin eri vaiheissa käsitellään tarkemmin luvussa "Tekoälyratkaisujen hyödyntäminen: rekrytointiprosessin vaiheiden ja niihin liittyvien vinoumien tarkastelu"





DESIGN PHASE



Henkilökohtaisten ennakkoluulojen tunnistaminen

Selkeän ongelmanmäärittelyn jälkeen yritys on päättänyt edetä tekoälypohjaisen ratkaisun toteuttamisessa. Sinut on nimitetty vastuulliseksi tekoälykehittäjäksi valitun teknologian käyttöönoton johtamiseen.

Yritys tunnistaa, että projektin onnistumisen kannalta on tärkeää koota monipuolinen ja eri sidosryhmiä edustava tiimi, jotta saadaan mukaan erilaisia näkökulmia.

Mahdollisten tiedostamattomien ennakkoluulojen käsittelemiseksi yritys on määrännyt, että kaikkien tiimin jäsenten on esitettävä harjoitukseen nimeltä

Sidosryhmät ja mahdolliset ennakkoluulot

Työkalu: Alan Turingin asemamatriisi

"Aseman" reflektointi— ymmärrys omasta sosiaalisesta ja kulttuurisesta taustasta — on keskeistä tekoälykehityksessä eettisten päätöksentekojen kannalta. Alan Turingin instituutti (2022) korostaa "asemamatriisin" tärkeyttä työkaluna arvioida, kuinka yksilölliset piirteet, kuten ikä, rotu ja koulutus, vaikuttavat tutkimukseen ja innovaatioon. Tämä itseymmärrys on avainasemassa estettäessä ennakkoluuloja ja valtasuhteiden epätasapainoa tekoälyjärjestelmissä. Olemalla tietoisia näistä henkilökohtaisista tekijöistä kehittäjät voivat varmistaa, että tekoälyratkaisut ovat oikeudenmukaisia ja kunnioittavat niitä moninaisia yhteisöjä, joita ne palvelevat, edistäen positiivista muutosta ilman epäoikeudenmukaisuuden jatkumista.

Alan Turinginasemamatriisi on pohdiskeleva kehys, joka paljastaa yksilön sosioekonomisten taustojen, kulttuuristen kontekstien ja elämäkokemusten vaikutuksen päätöksentekoon ja arvosteluun tiimissä. Se toimii työkaluna tunnistaa ja käsitellä erilaisten henkilökohtaisten piirteiden, kuten iän ja etnisyyden, vaikutuksia tutkimus- ja kehitysnäkökulmiin. Tämä matriisi on keskeinen tiedostamattomien ennakkoluulojen paljastamisessa ja valtasuhteiden hallinnassa, edistäen sitoutumista monimuotoisuuteen ja tasa-arvoon tekoälykehityksessä. (Alan Turing Institute, 2022)

Sidosryhmien huomioon ottaminen



Sisäisen sidosryhmän lisäksi, joka osallistuu projektiin, yrityksen täytyy ottaa huomioon myös muut siihen liittyvät sidosryhmät.

Tämän vuoksi yritys loi sidosryhmäkartan seuraten G.J. Millerin määritelmää tekoälyn kehitysprosessin sidosryhmien kartoittamisesta, jossa määritellään kehittämisen ja käytön sidosryhmät sekä ulkoiset sidosryhmät.

Karttaan perustuen projektipäällikön täytyy järjestää sidosryhmät ryhmiin tietääkseen, kuinka pitää sidosryhmät informoituina ja

Työkalu: Sidosryhmäkartta

Sidosryhmäkartta on visuaalinen työkalu, jota käytetään projektin kaikkien sidosryhmien roolien ja suhteiden tunnistamiseen ja esittämiseen. Se voi vaihdella yksinkertaisesta nelikentästä, joka kuvaa vaikutusvaltaa ja osallistumistasoa, aina yksityiskohtaiseen matriisiin, joka esittää sidosryhmien välisiä vuorovaikutussuhteita. Sidosryhmäkartta auttaa selkeyttämään sidosryhmädynamiikkaa ja on keskeinen verkostosuhteiden ymmärtämisessä ja hallinnassa. (Giordano ym., 2018; SDT, n.d)

Työkalu: Sidosryhmäanalyysimatriisi

Sidosryhmäanalyysimatriisi on visuaalinen työkalu, jota käytetään sidosryhmien luokitteluun ja hallintaan heidän vaikutusvaltansa ja kiinnostuksensa perusteella. Asettamalla sidosryhmät nelikenttään tämä matriisi auttaa priorisoimaan vuorovaikutusstrategiat. Yläneljässä vasemmalla ovat sidosryhmät, joilla on paljon vaikutusvaltaa mutta vähän kiinnostusta; heidän kohdallaan pyritään pitämään heidät tyytyväisinä kuormittamatta liikaa. Yläneljä oikealla sisältää sidosryhmät, joilla on sekä suuri vaikutusvalta että korkea kiinnostus, ja jotka vaativat tiivistä hallintaa. Alaneljän oikealla puolella ovat sidosryhmät, joilla on suuri kiinnostus mutta vähäinen vaikutusvalta; heitä tulisi pitää ajan tasalla säännöllisillä päivityksillä. Alaneljän vasemmalla ovat ne, joiden kiinnostus ja

vaikutusvalta ovat vähäisiä; heidän kohdallaan riittää vähäinen vuorovaikutus ja keskittyminen seurantaan. (Hoory & Botorff, 2022; Wallbridge, 2023)



Vinoumat rekryointivaiheissa; Työpaja

Kun sidosryhmät ovat selvillä, yritys päättää järjestää eettisen työpajan, jossa käydään läpi kaikki mahdolliset vinoumat eri rekryointivaiheissa.

Kenttään "Työskennellään yhdessä" (edellinen dia) kuuluvat sidosryhmät kutsutaan osallistumaan.

Millaisia post-it-lappuja sinä olisit täyttänyt?

Työkalu: Työpaja "workshop"

Työpaja on yhteisöllinen suunnittelutyökalu, jossa osallistujaryhmä osallistuu vuorovaikutteisiin toimintoihin ratkaistakseen tietyn ongelman tai työstääkseen projektia. Työpajaa johtaa yleensä fasilitaattori, ja sen kesto voi vaihdella muutamasta tunnista useisiin päiviin, riippuen tavoitteista ja tehtävien monimutkaisuudesta. Tämä menetelmä on tehokas ideoiden tuottamisessa, ongelmanratkaisussa sekä tiimityön ja innovoinnin edistämisessä ryhmän sisällä. (Wirtz, 2022)

Tekoälyratkaisujen paljastaminen: Rekryointiprosessin vaiheet ja niihin liittyvät vinoumat

Algoritmeja ja koneoppimisteknologioita käytetään yhä enemmän jo rekryointiprosessin alkuvaiheissa. Nämä ennakoivat lähestymistavat auttavat ennustamaan liiketoiminnan kehitystä analysoimalla ulkoisia tekijöitä, kuten markkinakysyntää, kilpailijoiden toimenpiteitä ja muutoksia kuluttajatarpeissa. Henkilöstöhallinnon mittareilla (HR-metriikoilla) on keskeinen rooli näiden ennusteiden laatimisessa. Niihin sisältyy historiallisten työvoimatrendien, väestörakenteen muutosten ja muuttuvien taitovaatimusten analyysi. Lisäksi sisäiset arviot, kuten työkuorman arvioinnit, esihenkilöhavainnot, henkilöstöpalautteet ja työvoiman demografiat, ovat olennaisia nykytilan ja

henkilöstön kehityksen ymmärtämisessä. Mittarit, kuten vaihtuvuusaste, tehtävien täyttöaika ja rekrytointikustannukset, ovat ratkaisevia organisaation rekrytointitehokkuuden arvioinnissa ja kehityskohteiden tunnistamisessa. (www.recruiter.com, 2023)

Ennakoivaan rekrytointiin liittyvät vinoumat voivat johtua useista lähteistä, kuten nykyisten työntekijöiden suoriutumistasosta tietyissä rooleissa, mikä voi johtaa ankkuroitumisvinoumaan. Esimerkiksi, jos erittäin hyvin suoriutuva työntekijä täytyy korvata, analytiikka voi ehdottaa yhden työntekijän palkkaamista. Kuitenkin lähtijän korkea suorituskky saattaa todellisuudessa edellyttää kahden työntekijän palkkaamista. Vastaavasti, heikosti suoriutuvan työntekijän perusteella voidaan virheellisesti päätellä, että tarvitaan lisää työvoimaa paikkaamaan tuottavuusvajetta. (Chen, 2023, s. 145) Lisäksi muutosvastarinta työntekijöiden keskuudessa voi johtaa status quo -vinoumaan, jossa organisaatorakenteita ei haluta mukauttaa, vaikka liiketoiminnalliset tarpeet sitä selvästi edellyttäisivät. (Sukernek, 2024)

Ennakoiva analyysi voi myös jättää huomiotta kriittisiä tekijöitä liian kapean tarkastelun vuoksi. Esimerkiksi toimialan laajempien trendien tai väestörakenteen muutosten, kuten väestön ikääntymisen, huomiotta jättäminen voi johtaa siihen, että yritys ei ole valmistautunut tuleviin työvoimatarpeisiin.

Kuten muutkin rekrytointiprosessin vaiheet, myös tämä vaihe on altis palautesilmukkavinoumalle. Tämä vinoumatyyppi voi syntyä, kun rekrytoinnissa käytettävät tekoälyjärjestelmät koulutetaan historiallisella datalla ja ne vahvistavat tahattomasti omia ennakkoluuloisia päätöksiään. Tämä tapahtuu dynaamisissa ympäristöissä, joissa tekoälyn tekemät päätökset vaikuttavat siihen dataan, josta se tulevaisuudessa oppii. Jos tekoäly, joka on jo ennakkoluuloinen aiemman datan vuoksi, jatkaa vinoutuneiden rekrytointipäätösten tekemistä, nämä päätökset palautuvat sen opetusaineistoon, mikä ylläpitää ja jopa voimistaa alkuperäisiä vinoumia. (Vivek, 2023, s. 0107)

Vinoumat työpaikkailmoituksissa

Tekoäly voi helpottaa tarkkojen työpaikkailmoitusten laatimista. Muokkaamalla ilmoitusten kieltä ja seuraamalla näiden muutosten vaikutusta hakemusten määrään ja laatuun organisaatiot voivat tehostaa ilmoittelunsa vaikuttavuutta. Lisäksi tekoäly voi tunnistaa, mitä osa-alueita yrityksestä – mukaan lukien sen kulttuuri ja saavutukset – tulisi korostaa potentiaalisille hakijoille parhaan mahdollisen vastauksen saamiseksi. (Chen, 2023, s. 140)

Työpaikkailmoituksilla on keskeinen rooli yrityksen tarpeiden viestimisessä tuleville työntekijöille. Ilmoituksissa esiintyvät vinoumat voivat kuitenkin estää pätevien hakijoiden hakeutumisen tehtävään. Tällaiset vinoumat liittyvät usein

ikään, sukupuoleen tai koulutustaustaan. (Ongig, 2024)

Mielenkiintoista on, että artikkeleissa ehdotetaan tekoälyn voivan vähentää joitakin näistä ongelmista tarkastelemalla työpaikkailmoituksia ja tunnistamalla kieltä, joka saattaa aiheuttaa vinoumaa. Vaikka tekoälyn tuottamat työpaikkailmoitukset voivat sisältää vähemmän vinoumia kuin ihmisten kirjoittamat, on olemassa riski, että tekoäly ylläpitää olemassa olevia ennakkoluuloja, mikäli työkaluja ei suunnitella huolellisesti. (CareerExperts, 2023) Esimerkiksi ankkuroitumisvinoumaa voi esiintyä esimiehen aiempien rekrytointipäätösten perusteella, jolloin nykyisten huipputyöntekijöiden ominaisuudet ja taidot vaikuttavat uusien työkuvausten laatimiseen. (Chen, 2023, s. 145)

Lisäksi tekoälyn ehdotukset tiettyjen termien, kuten 'kunnianhimoinen' tai 'itsevarma johtaja', käytöstä työpaikkailmoituksissa voivat tahattomasti suosia tai karkottaa tiettyjä väestöryhmiä, ja siten aiheuttaa vinoumaa (Lawton, 2022). Jos tekoälyjärjestelmät koulutetaan aiemmilla työpaikkailmoituksilla, ne saattavat toistaa sellaista kieltä ja vaatimuksia, jotka ovat historiallisesti houkutelleet vähemmän monimuotoista hakijajoukkoa – ja näin ylläpitää tätä kehitystä.

Vinoumat osaajien etsinnässä

Erilaisia tekoälymenetelmiä hyödynnetään jo varhaisessa vaiheessa potentiaalisten hakijoiden kanssa tapahtuvassa vuorovaikutuksessa. Algoritmit mahdollistavat kohdennetun viestinnän digitaalisilla alustoilla ja sosiaalisessa mediassa, ja tekoälybotteja käytetään sekä aktiivisten että passiivisten työnhakijoiden tunnistamiseen eri alustoilla tai yrityksen aiempien hakijoiden tietokannoissa. Lisäksi tekoäly voi hyödyntää tekstinlouhintaa työpaikkailmoitusten houkuttelevuuden lisäämiseksi käyttämällä ilmaisuja, jotka houkuttelevat eniten hakijoita. Tekoäly voi myös käyttää sävymittaria (tone metric) antaakseen suosituksia ilmoitusten muokkaamiseksi osallistavammiksi. (Hunkenschroer & Luetge, 2022, s. 992)

Tekoälypohjaiset työpaikkailmoitukset sisältävät luontaisesti syrjinnän riskin. Työnantajat voivat kohdentaa ilmoituksia tiettyjen kriteerien, kuten kielitaidon, koulutustaustan, työkokemuksen tai jopa iän ja sukupuolen mukaan, mikä voi estää tiettyjä ryhmiä näkemästä ilmoituksia ja siten edistää työpaikkojen yksimuotoisuutta (Chen, 2023, s. 144; Sheard, 2022, s. 624). Esimerkiksi Verizonin Facebook-ilmoitus, joka kohdistettiin 25–36-vuotiaille henkilöille, jotka asuvat tai ovat vierailleet Yhdysvaltain pääkaupungissa ja ovat kiinnostuneita rahoituksesta, saattoi tahattomasti sulkea pois muita potentiaalisia työnhakijoita. (Sheard, 2022, s. 624)

Algoritmit, joiden tavoitteena on tehostaa ilmoittelua, suosivat usein väestöryhmiä, jotka ovat aiemmin osoittaneet suurinta vuorovaikutusta ilmoitusten kanssa, mikä kaventaa hakijakunnan monimuotoisuutta ajan myötä. (Köchling & Wehner, 2020). Rekrytointivinjaumia vahvistavat entisestään rekrytoijien mieltymykset tiettyihin ilmoituslustoisiin sekä taipumus vuorovaikuttaa eniten näiden alustojen käyttäjien kanssa. Tämä voi johtaa monimuotoisuuden puutteeseen, koska eri alustat houkuttelevat erilaisia käyttäjiä. (Lawton, 2022).

Vinoumat eivät synny pelkästään algoritmeista tai työnantajien käytännöistä – myös hakijoiden käyttäytyminen vaikuttaa. Potentiaalisten ehdokkaiden tunnistaminen sosiaalisen median tai muun verkkotoiminnan perusteella saattaa suosia niitä, jotka ovat aktiivisia verkossa tai osaavat rakentaa houkuttelevan digitaalisen läsnäolon. Tämä näkyy erityisesti kilpailuilla aloilla, kuten IT:ssä, missä kiireiset ammattilaiset eivät välttämättä päivitä LinkedIn-profiiliaan säännöllisesti. Toisaalta jotkut voivat tietoisesti rakentaa profiilejaan näyttääkseen houkuttelevammilta, mikä herättää kysymyksiä tällaisten digitaalisten persoonien aitoudesta. (Naakka, 2018, s. 46–47)

Tekoälypohjaisessa rekrytoinnissa läpinäkyvyys on merkittävä eettinen haaste. Algoritmien monimutkaisuus voi johtaa ns. ”musta laatikko” -ilmiöön, jossa päätösten perustelut jäävät hakijoille epäselviksi. (DigitalOcean, n.d.) Lisäksi herää eettisiä kysymyksiä hakijoiden sosiaalisen median profiilien ja muun verkkotoiminnan käytöstä rekrytoinnissa. Eri maissa sovellettavat tietosuojastandardit vaihtelevat, ja Euroopan yleinen tietosuoja-asetus (GDPR) on yksi tiukimmista. (Hunkenschroer & Luetge, 2022, s. 995)

Vinoumat hakijoiden seulonnessa ja valinnassa

Tekoälyteknologiat näyttelevät keskeistä roolia työnhakijoiden alkuvaiheen seulonnessa, jossa etusijalle asetetaan ne hakijat, joiden katsotaan olevan sopivimpia tehtävään. Aiemmin yritykset skannasivat ansioluetteloita tiettyjen avainsanojen perusteella. Nykyiset menetelmät ovat kuitenkin kehittyneempiä ja sisältävät muun muassa chatbotteja ja ansioluetteloiden jäsentämistyökaluja, jotka etsivät semanttisia yhteyksiä ja arvioivat muita pätevyyyksiä. Jotkin työkalut ennustavat myös hakijan tulevaa suoriutumista etsimällä viitteitä pitkäaikaisesta sitoutumisesta tai tuottavuudesta sekä negatiivisia merkkejä, kuten toistuvia myöhästymisiä tai kurinpidollisia ongelmia. (Hunkenschroer & Luetge, 2022, s. 992)

Hakijoiden valikointiprosessissa tekoälyä käytetään myös videohaastatteluiden analysointiin. Tekoälyavustajat esittävät etukäteen määritellyn joukon kysymyksiä ja arvioivat paitsi vastaukset myös hakijan äänensävyä, mikroilmeitä ja emotionaalisia reaktioita persoonallisuuspiirteiden selvittämiseksi. Lisäksi tekoälypohjaisia videopelejä käytetään arvioimaan ominaisuuksia kuten

riskinottohalukkuutta, suunnittelutaitoja, sinnikkyyttä ja motivaatiota. (Hunkenschroer & Luetge, 2022, s. 992)

Tekoälyratkaisut analysoivat myös hakijoiden digitaalista jalanjälkeä, kuten sosiaalisen median toimintaa ja muuta verkkokäyttäytymistä, käyttämällä kielellistä analyysia arvioidakseen henkilön sopivuutta yrityskulttuuriin. (Hunkenschroer & Luetge, 2022, s. 992) Sosiaalisen median lisäksi neuro-lingvististä ohjelmointia (NLP) hyödynnetään rekrytoinnin eri vaiheissa, mukaan lukien ansioluetteloiden analysointi, tarjoten syvällisempää tietoa hakijoista. (Albaroudi et al., 2024, s. 391)

Tekoälyllä on potentiaalia oikeudenmukaisemman rekrytointikäytännön edistämisessä vähentämällä tiedostamattomia ennakkoarvioita. Kun arvioinnissa keskitytään ennalta määriteltyihin kykyihin ja pätevyyskriteereihin, voidaan sivuuttaa demografiset tekijät kuten ikä, sukupuoli ja etnisyys, jotka saattaisivat muuten vaikuttaa rekrytoijan arvioon. Tämä voi lisätä työpaikkojen monimuotoisuutta, kun hakijoita arvioidaan heidän osaamisensa ja saavutustensa perusteella. Lisäksi tietyt tekoälysovellukset on suunniteltu erityisesti tukemaan yrityksiä monimuotoisuus- ja inklusiivisuustavoitteiden saavuttamisessa tutkimalla rekrytointitrendejä ja ehdottamalla muutoksia mahdollisten vinoumien korjaamiseksi. (DigitalOcean, n.d.)

Vaikka seulontaprosessit voivat vähentää joitain vinoumia, ne voivat myös tahattomasti luoda uusia, mikäli niitä ei toteuteta huolellisesti. Vinoumat voivat syntyä mallin suunnitteluperusteista, valituista muuttujista ja opetusaineistosta. (Hunkenschroer & Luetge, 2022, s. 994)

Kun koulutusaineisto pohjautuu olemassa olevien työntekijöiden tai suppean väestöryhmän tietoihin, jotka eivät edusta monimuotoista väestöä, se voi johtaa tahattomaan syrjintään aliedustettuja ryhmiä kohtaan. (Hunkenschroer & Luetge, 2022, s. 994)

Tekoälyjärjestelmät, jotka oppivat historiasta peräisin olevista vinoutuneista aineistoista, voivat vahvistaa näitä vinoumia edelleen. Esimerkiksi jos järjestelmä koulutetaan pääasiassa tietyn demografisen ryhmän ansioluetteloilla, se voi suosia hakijoita kyseisestä ryhmästä epäoikeudenmukaisesti. Tunnettu tapaus tapahtui vuonna 2018, kun Amazonin tekoälypohjainen rekrytointijärjestelmä suosi miespuolisia hakijoita. Tämä johtui siitä, että koulutusaineisto sisälsi pääosin miesten ansioluetteloita, ja algoritmi rankaisi piirteitä, joita yleisesti yhdistetään naisiin. (Albaroudi et al., 2024, s. 386; Hunkenschroer & Luetge, 2022, s. 994)

Lisäksi koulutusvinoumaa voi syntyä, jos algoritmi ei huomioi kaikkia työn kannalta olennaisia taitoja ja ominaisuuksia. Jos työkalu painottaa liikaa teknisiä taitoja eikä ota huomioon vuorovaikutuskykyä, voidaan ohittaa tärkeitä pätevyyskriteerejä kuten viestintätaidot. Yleistysvinouma (aggregation bias) voi myös ilmetä, kun tehdään virheellisiä yleistyksiä koko väestöstä ja näin suljetaan yksilöitä pois ryhmään kohdistuvien oletusten perusteella. Esimerkiksi algoritmi

voi olettaa, että nuoremmat henkilöt hallitsevat teknologian paremmin, ja näin sivuuttaa iäkkäämpiä, yhtä päteviä hakijoita. (Albaroudi et al., 2024, s. 386)

Hakijoiden digitaalisen jalanjäljen arviointi voi myös ylläpitää vinoumia, sillä ihmisten näkemykset sosiaalisen median käytöstä ja mahdollisuudet profiilien päivittämiseen vaihtelevat. Pohjois-Carolinan osavaltionyliopiston tutkimuksessa haastateltiin 61 HR-ammattilaista, ja tulokset osoittivat, että henkilökohtaisia kiinnostuksen kohteita, kuten vaeltamista tai joulun viettoa, pidettiin ammatillisia ominaisuuksia tärkeämpinä. Tällainen lähestymistapa suosii tiettyjä väestöryhmiä ja olettaa, että ominaisuudet kuten "energinen" tai "aktiivinen" ovat automaattisesti parempia, mikä voi syrjiä iäkkäitä tai vammaisia henkilöitä. Tekoälyn tuominen tähän prosessiin voi lukita nämä ennakkoluulot algoritmien sisälle. (Shipman, 2021)

Tunneanalytiikkaa hyödyntävien ohjelmistojen käyttö videohaastatteluissa edellyttää varovaista suhtautumista eri kielten intonaatioihin sekä etnisten ryhmien mahdollisiin järjestelmällisiin epäedullisiin asemointeihin. Lisäksi osa hakijoista voi kokea vaikeaksi käyttäytyä luonnollisesti kameran edessä, mikä voi johtaa työkalun tekemiin virheellisiin arviointeihin. (Hunkenschroer & Luetge, 2022, s. 995)

Hakijoiden tietojen käyttö rekrytoinnissa liittyy myös olennaisesti yksityisyydensuojaan ja suostumukseen, jotka ovat keskeisiä eettisiä kysymyksiä. Eri maiden sääntelyt poikkeavat toisistaan, ja Euroopan yleinen tietosuoja-asetus (GDPR) on yksi tiukimmista. Asetuksen tavoitteena on suojella EU-kansalaisten oikeuksia säätelämällä henkilötietojen keruuta, säilyttämistä ja käsittelyä, ja edellyttämällä tietoista suostumusta tietojen käyttöön. Haasteeksi kuitenkin jää, uskaltavatko hakijat kieltäytyä suostumuksesta, jos he pelkäävät sen vaikuttavan heidän mahdollisuuksiinsa saada työpaikka. (Hunkenschroer & Luetge, 2022, s. 995)

Vinoumat hakijoiden haastattelussa

Yksi syy tekoälypohjaisten haastattelumenetelmien käyttöön voi olla kielimuurit monikansallisessa rekrytoinnissa, mikä vaatii paikallista osaajahankintaa ja monikielisiä HR-ammattilaisia. Tekoäly vähentää tätä tarvetta mahdollistamalla erikielisten hakijoiden haastattelun ilman, että HR:n tarvitsee hallita jokaista paikallista murretta, kiitos luonnollisen kielen käsittelyn (Natural Language Processing, NLP) kehittymisen. Tämä teknologia antaa yrityksille mahdollisuuden kuroa tehokkaasti umpeen kielikuilua globaalissa rekrytoinnissa ilman kääntäjiä. (Albaroudi ym., 2024, s. 392)

Tekoälypohjaisia haastattelutyökaluja käytetään kuitenkin useammin ensimmäisen haastattelukierroksen automatisointiin, jolloin ne osallistuvat ehdokkaiden arviointiin. (Fritts & Cabrera, 2021, s. 792) Tästä syystä tekoälypohjaiset haastatteluratkaisut ja niiden mahdollisesti synnyttämät vinoumat on käsitelty edellisessä luvussa "Ehdokkaiden seulonta ja esivalinta".

Vivek (2023) ehdottaa, että ihmisten hienovarainen ymmärrys ja tunneäly puoltavat sitä, että lopullinen päätösvalta rekrytoinneissa tulisi säilyttää ihmistyöntekijöillä. Fritts & Cabrera (2021) huomauttavat myös, että tekoälyn hyödyntäminen ehdokkaiden tunnistamisessa, seulonnassa, haastattelussa ja perehdytyksessä voi johtaa epäinhimillistymiseen, kun henkilökohtainen vuorovaikutus jää pois. Tämä voi olla eettisesti ongelmallista, sillä rekrytointialgoritmeihin upotetut keinotekoiset arvot voivat heikentää työntekijän ja työnantajan välistä suhdetta riisumalla siitä inhimillisen vuorovaikutuksen perinteiset piirteet.

Vinoumat arvioinnissa ja työtarjouksen laatimisessa

Kun ehdokas on valittu, yrityksen on tehtävä hänelle työtarjous. Tekoälyalgoritmit voivat arvioida ehdokkaan aiempia työsuhteita ja muodostaa tarjouksen, jonka ehdokas todennäköisesti hyväksyy. Nämä työkalut voivat kuitenkin pahentaa olemassa olevia epätasa-arvoja aloituspalkoissa ja myöhemmissä palkkakehityksissä eri sukupuolten, etnisten ryhmien ja muiden demografisten tekijöiden välillä. Tämä johtuu siitä, että työkalut perustuvat usein historiallisiin palkkatietoihin, jotka voivat sisältää sisäänrakennettuja vinoumia. (Lawton, 2022)

Vinoumat perehdytyksessä

Tekoälyn hyödyntäminen perehdytyksessä voi vähentää henkilöstöhallinnon rutiinitehtäviä automatisoimalla asiakirjahallintaa, viestintäaikataulutusta ja käyttämällä chatbotteja palautteen ja kysymysten käsittelyyn alkuvaiheessa. Tekoäly arvioi myös uusien työntekijöiden edistymistä ja varmistaa koulutusvaatimusten noudattamisen samalla, kun se tarjoaa analytiikkaa perehdytysprosessin kehityksestä (Marr, 2023).

Tekoälypohjaiset käyttöliittymät voivat räätälöidä koulutuspolkuja yksilöllisten taitojen ja oppimistyylien mukaan varmistaen, että uudet työntekijät saavat tarvitsemansa tuen onnistumiseen. HR-osastot voivat hyödyntää automatisoiduista vuorovaikutuksista ja palautteesta saatavaa tietoa ymmärtääkseen uusien työntekijöiden osaamista ja kehitystarpeita. Ennakoiva analytiikka voi tunnistaa uusien työntekijöiden tarpeita etukäteen ja tarjota oikea-aikaisia resursseja ja tukea (Marr, 2023).

Tekoälyä voidaan käyttää myös uuden työntekijän ja työmentorinsa yhdistämiseen heidän taitojensa, kiinnostuksen kohteidensa ja asiantuntemuksensa perusteella (Featured, 2023). Kuten aiemmissa vaiheissa, myös tekoälypohjaisessa perehdytyksessä voi esiintyä vinoumia, jos koulutusdata perustuu historiallisiin aineistoihin, jotka sisältävät aiempia ennakkoluuloja. Lisäksi algoritmeihin voi huomaamatta sisältyä kehittäjiensä tiedostamattomia vinoumia koulutuspolkuja suunniteltaessa.

6. Ehdokkaiden seulonta ja esivalinta: aineiston määrittely



Ehdokkaiden seulonta ja esivalinta; aineiston määrittely

Vaikka yritys tarkastelee tekoälyn mahdollisuuksia ja vaikutuksia kaikissa rekrytoinnin vaiheissa, se keskittyy tällä hetkellä erityisesti tekoälyn hyödyntämiseen seulonta- ja esivalintavaiheessa. Tämä sisältää ansioluetteloiden (CV:iden) tarkastelun perusteellisen analyysin sekä ehdokkaiden sovittamisen oikeisiin työtehtäviin.

Tekoälytyökalujen avulla pyritään automatisoimaan ansioluetteloiden seulontaprosessi siten, että työvaatimukset täyttävät ehdokkaat tunnistetaan nopeasti ja tarkasti. Tällaisen teknologian käyttöönoton odotetaan mahdollistavan henkilöstöresurssien strategisemmän hyödyntämisen, parantavan esivalittujen ehdokkaiden laatua sekä lisäävän koko rekrytointiprosessin tehokkuutta sovittamalla ehdokkaiden osaaminen ja taustat yrityksen tarpeisiin.

Ensimmäisenä vaiheena yrityksen tulee määritellä ne tietoaaineistot (datasetit), jotka muodostavat rekrytointijärjestelmän perustan ja joista algoritmit ammentavat tietonsa.

Tekoälyn kouluttaminen

Koulutusdatalla on keskeinen rooli algoritmien kehittämisessä. Nämä aineistot muodostavat algoritmien perustan. Tavoitteena on, että koulutusaineistot toimivat niin sanottuna "maastototuutena" eli faktapohjana, joka ohjaa algoritmia tulkitsemaan ja käsittelemään suuria tietomääriä. Ne opettavat algoritmia ymmärtämään syötteiden (järjestelmään syötettyjen tietojen) ja lopputulosten (algoritmin tekemät ennusteet tai päätökset) välisiä yhteyksiä.

Merkittävä haaste syntyy, jos koulutusdata sisältää ennakkoluuloja tai vinoumia. Nämä voivat heijastaa historiallisia epätasa-arvoja, ennakkokäsityksiä tai olla seurausta puutteellisesta tai harhaanjohtavasta tiedonkeruusta. Kun algoritmia koulutetaan vinoutuneella datalla, se omaksuu nämä vinoumat ja toistaa niitä tekemissään ennusteissa ja päätöksissä. Tätä ilmiötä kutsutaan usein nimellä "bias in, bias out" – eli jos syötedata (bias in) on vinoutunutta, myös mallin tuottamat lopputulokset (bias out) ovat vinoutuneita. Tämä voi johtaa epäoikeudenmukaisiin, syrjiviin tai virheellisiin lopputuloksiin, jolloin ehdokkaita valitaan tai hylätään vinoutuneiden kriteerien – ei heidän todellisten ansioidensa tai potentiaalinsa – perusteella. (Sheard, 2022)

Tekoälyn kouluttaminen yrityksen sisäisellä datalla

Tekoälysovelluksia ansioluetteloiden seulontaa varten kouluttaessaan yrityksillä on käytössään useita sisäisiä tietolähteitä, jotka voivat olla keskeisiä oppimisprosessin kannalta. (Fernandes, 2021; Ma, 2024; Sheard, 2022)

Yksi tällainen lähde on työntekijöiden suorituskyydata yhdessä työhistoriaan liittyvien tietojen kanssa. Tämä sisältää sisäisiä arviointitietoja ja aikaisempien rekrytointien tuloksia. Lisäksi yritykset voivat hyödyntää vuosien varrella kertyneitä ansioluetteloita, jotka sisältävät runsaasti hakemuksia suoraan työpaikkoihin.

Toinen arvokas tietolähde on niin sanotut "väärät negatiiviset tapaukset", eli tiedot ehdokkaista, jotka hylättiin virheellisesti aiemmissa rekrytoinneissa. Näiden tapausten analysointi voi antaa oivalluksia siitä, millaisia virheellisiä malleja tekoälyn tulisi välttää. Yritykset voivat myös poimia CV:istä usein esiintyviä avainsanoja ja ilmaisuja, joita voidaan käyttää tekoälyn kyvyn parantamiseen oikeanlaisen osaamisen tunnistamiseksi.

Lopuksi myös koulutustaustaan liittyvät tiedot – joita hakemuksissa lähes aina esiintyy – voidaan hyödyntää tekoälyn kyvykkyyden lisäämiseksi akateemisten pätevyyksien ja niiden merkityksen arvioinnissa eri työtehtäviin.

Tekoälyn kouluttaminen yrityksen ulkoisella datalla

Mahdollisia ulkoisia tietolähteitä tekoälyn kouluttamiseen on esitetty useissa tutkimuksissa, ja niihin kuuluu useita alustoja ja tiedonkeruumenetelmiä. (Banerjee, 2022) Sosiaalisen median alustat tarjoavat runsaasti tietoa, kun käyttäjät jakavat henkilökohtaisia kiinnostuksen kohteita, kokemuksia ja ammatillisia saavutuksiaan. Näistä voidaan saada viitteitä ehdokkaan persoonallisuudesta, taidoista ja ammatillisista verkostoista (Albaroudi et al., 2024; Albassam, 2023; Chaker, 2018; Filtered, n.d.; Heymans, 2022; Shipman, 2021).

Lisäksi ammatilliset verkkosivustot, kuten LinkedIn, tarjoavat yksityiskohtaisia profiileja, jotka sisältävät työhistorian, koulutuksen, taidot, suositukset ja saavutukset. Näiden profiilien kautta voidaan muodostaa kattava kokonaiskuva potentiaalisista hakijoista.

Lopuksi yritykset voivat tehdä internet-hakuja, joiden avulla löydetään julkisesti saatavilla olevia tietoja hakijoista esimerkiksi uutisartikkeleista, blogeista, julkaisuista tai muista verkkoläsnäolon muodoista. Tämä laaja tietokenttä tarjoaa arvokasta kontekstia henkilön pätevyyksistä, saavutuksista ja luonteesta. (Chaker, 2018).

7. Algoritmisen mallin kehittäminen



Työpaikkailmoitus; avainsanojen valinta algoritmia varten

Yritys on hiljattain julkaissut työpaikkailmoituksen ja saanut siihen satoja hakemuksia. Tekoälyn perustuva seulonta- ja esivalintaprototyyppi skannaa hakijoiden ansioluettelot etsien avainsanoja ja muuta tietoa, joiden uskotaan olevan yhteydessä onnistuneisiin rekrytointeihin. Näitä voivat olla esimerkiksi työkokemus, työnimikkeet, aiemmat työnantajat, yliopistot ja tutkinnot. Tämän pohjalta järjestelmä luo

Tekoäly tehostaa merkittävästi hakijoiden seulontaprosessia, joka on perinteisesti yksi rekrytoinnin aikaa vievimmistä vaiheista. Rekrytointiprosessin

"seulontavaiheessa" hakijoiden arviointi kokemuksen, taitojen ja muiden olennaisten ominaisuuksien perusteella on keskeistä haastattelulistan laatimisessa. Tekoälyä voidaan hyödyntää ansioluetteloiden skannaamiseen avainsanojen ja muiden onnistuneisiin rekrytointeihin liitettyjen tietojen, kuten työnimikkeiden, aiempien työnantajien, oppilaitosten ja tutkintojen, perusteella. Näin sopivien ehdokkaiden tunnistaminen nopeutuu. (Dennis, 2023; Fernandes, 2021; Sheard, 2022)

Tekoälytyökalujen avulla rekrytoijat voivat nopeasti tunnistaa ja priorisoida parhaat ehdokkaat, mikä tehostaa prosessia ja parantaa keskeisiä rekrytointimittareita, kuten rekrytointiaikaa. Tekoälyjärjestelmät karsivat automaattisesti pois hakijat, jotka eivät täytä ennalta määriteltyjä kriteerejä, erittelevät ansioluetteloista oleellisen kokemuksen ja pisteyttävät hakijat heidän soveltuvuutensa perusteella. Tämä automatisointi mahdollistaa sen, että rekrytoijat voivat keskittyä paremmin prosessin myöhempiin vaiheisiin. Lisäksi tekoäly voi jo alkuvaiheessa seulonnan yhteydessä skannata ansioluettelot ja tunnistaa ne hakijat, jotka täyttävät tietyt vaatimukset. Koneoppimisalgoritmit arvioivat näiden hakijoiden taidot ja pätevyyden suhteessa työtehtävän vaatimuksiin ja kokoavat listan sopivista ehdokkaista. (ResumeMent, 2023; Dennis, 2023)

Työpaikkailmoitus sisältää keskeiset tiedot tehtävästä, kuten vaaditut taidot, pätevyudet ja kokemuksen, joita tekoäly käyttää ansioluetteloiden seulontaan. Ilman yksityiskohtaista työpaikkailmoitusta tekoälyltä puuttuisivat tarvittavat kriteerit, joiden perusteella arvioida ansioluotteloita.

Tehtäväkuvaus: Viestintävastustaja

Etsimme viestintävastustajaa vastaamaan sisällön, kuten tiedotteiden, blogien ja sosiaalisen median julkaisujen tuottamisesta yrityksemme puolesta. Lisäksi tehtäviin kuuluu median ja kampanjoiden seuranta sekä osallistuminen sisäisiin ja ulkoisiin tapahtumiin.

Tehtävät:

- Kehittää ja hallinnoida monipuolista sisältöä (tiedotteet, blogit, sosiaalisen median julkaisut) yrityksen strategisten tavoitteiden mukaisesti; seurata medianäkyvyyttä ja kampanjoita.
- Tukea sisäisten ja ulkoisten viestintästrategioiden toteutusta sekä osallistua yrityskuvan hallintaan.
- Järjestää markkinointitapahtumia ja tarjota kattavaa hallinnollista tukea, kuten ylläpitää tapahtumakalentereita ja päivittää yhteystietolistoja.
- Koota ja valmistella esityksiä ja raportteja, seurata projektien etenemistä ja medianäkyvyyttä.

Vaatimukset:

- Suoritettu alempi korkeakoulututkinto viestinnän, markkinoinnin tai vastaavalta alalta; erinomainen suullinen ja kirjallinen viestintätaito.
- Vahva osaaminen sosiaalisen median strategioissa ja mediayhteistyössä sekä luova ja innovatiivinen työote.
- Hyvät organisointikyvyt ja tarkkuus yksityiskohdissa; kykenee hoitamaan useita tehtäviä samanaikaisesti ja ylläpitämään hyviä ihmissuhteita.
- Hallitsee toimisto-ohjelmistoja ja suunnitteluohjelmia, kuten Photoshop ja InDesign, sekä tuntee hyvin eri sosiaalisen median alustat.

Tekoälyn kehitys on edennyt yksinkertaisesta avainsanojen vastaavuudesta pidemmälle. Aluksi yritykset käyttivät algoritmeja ansioluetteloiden

skannaamiseen ennalta valittujen avainsanojen tai -ilmaisujen perusteella. Nykyiset tekoälyteknologiat, kuten chatbotit ja kehittyneet ansioluetteloiden jäsennystyökalut, arvioivat semanttista merkitystä ja siihen liittyviä termejä määrittääkseen tarkemmin hakijan pätevyyden. Lisäksi koneoppimisalgoritmeja käytetään ennustamaan tulevaa työsuoriutumista erilaisten indikaattorien, kuten työsuhteiden keston ja tuottavuuden, perusteella. Nämä työkalut voivat myös ehdottaa hakijoille sopivimpia avoimia työpaikkoja heidän profiiliensa perusteella (Hunkenschroer & Luetge, 2022).

Tällä kurssilla käytämme kuitenkin yksinkertaista avainsanavastaavuutta esimerkkinä havainnollistaaksemme, kuinka tekoälyyn perustuva ansioluettelo-seulontasovellus käsittelee hakemuksia.

Luettelo ehdokkaiden seulonnassa käytetyistä erityisistä avainsanoista

- "Alempi korkeakoulututkinto viestinnässä"
- "Alempi korkeakoulututkinto markkinoinnissa"
- "Sisällöntuotanto"
- "Blogit"
- "Sosiaalisen median julkaisut"
- "Median seuranta ja kampanjaseuranta"
- "Viestintästrategioiden toteutuksen tukeminen"
- "Yrityskuvan hallinta"
- "Markkinointitapahtumien järjestäminen"
- "Kattavan hallinnollisen tuen tarjoaminen"
- "Yhteystietolistojen päivittäminen"
- "Esitysten ja raporttien kokoaminen"
- "Projektien seuraamisen ja medianäkyvyyden seuranta"
- "Luova ja innovatiivinen työote"
- "Vahvat organisointitaidot ja tarkkuus yksityiskohdissa"
- "Kyvykäs hoitamaan useita tehtäviä samanaikaisesti"
- "Hyvät ihmissuhdetaidot"
- "Toimisto- ja suunnitteluohjelmistojen osaaminen"
- "Photoshop" and "InDesign"
- "Eri sosiaalisten medioiden tuntemus"

Sheardin (2022) mukaan ansioluettelot tarkastetaan yleensä ihmisten toimesta vain silloin, kun ne vastaavat työpaikkailmoituksen kriteerejä, kun taas muut saatetaan sivuuttaa kokonaan. Tämä herättää huolen siitä, että merkittävä osa ansioluetteloista voidaan hylätä ennen aikaisesti.

Seuraavaksi keskitymme kahden hakemuksen analysointiin ymmärtääksemme, miten tekoälysovelluksen prototyyppi arvioi hakijoita seulontavaiheessa. Tämä analyysi auttaa tunnistamaan, onko valintaprosessiin mahdollisesti vaikuttavia vinoumia.

MODEL DEVELOPMENT AND TRAINING



Kahden hakijan CV:n vertailu avainsanojen muotoilun näkökulmasta

Yritys järjestää työpajan tutkiakseen tekoälypohjaisen CV-seulonnan yksityiskohtia oman toimintansa tapausesimerkin avulla.

Työpajan aikana analysoidaan kahta ansioluetteloa rinnakkain, jotta voidaan havaita, miten sanavalintojen ja muotoilun erot voivat tahattomasti johtaa pätevien hakijoiden sivuuttamiseen.

JOHN DOE

Communications assistant

EXPERIENCE

Communications assistant
ABC Corporation
2019 - Present

- Developed and managed diverse content, including **media releases, blogs, and social media posts**.
- Supported the implementation of comprehensive communications strategies.
- Organized multiple **marketing events** and **maintained event calendars**.
- Compiled **presentations** and tracked **project progress**, ensuring alignment with strategic goals.

EDUCATION

University of communications
Bachelor's degree in Marketing
2014-2018

SKILLS SUMMARY

- Proficient in Photoshop and InDesign**
- Skilled in social media strategies and media relations**
- Excellent verbal and written communication abilities
- Strong organizational skills and attention to detail**

About Me

A dedicated communications professional with a **Bachelor's degree in Marketing**, seeking the role of Communications Assistant to leverage extensive experience in **content creation, social media management**, and event coordination to contribute to the company's strategic goals.

+123-456-7890
hello@doeland.com
123 Anywhere St., Any City

JANE DOE

Digital Content Creator

EDUCATION

CREATIVE UNIVERSITY
BSc in Communications Studies 2018

WORK EXPERIENCE

Digital Content Creator; Creative Media Co
2019 - Present

- Spearheaded **content** initiatives across digital channels, crafting engaging articles and dynamic social engagements.
- Drove strategy for public relations efforts, enhancing brand visibility.
- Led the logistics for promotional and networking gatherings, ensuring smooth execution.
- Synthesized data into compelling **reports** and visuals for team updates.

COMPETENCIES

- Advanced user of digital design tools and content management systems.
- Crafted engaging online dialogue and cultivated media partnerships.
- Articulate communicator, both in written formats and oral presentations.
- Excelling in project orchestration and meticulous in administrative tasks.

PROFILE

Passionate communicator with an academic background in Communications Studies and hands-on experience in crafting engaging narratives and managing digital platforms. Eager to bring my toolkit of creative dissemination and stakeholder engagement to the Communications Assistant position.

CONTACT ME

(123) 456-7890
Jane@doe.com
123 Anywhere St., Any City, State, Country 12345

Tämä esimerkki havainnollistaa, miten sanamuotojen ja ilmaisujen erot kahdessa ansioluettelossa voivat vaikuttaa tekoälyn suorittaman seulonnan lopputulokseen. Vaikka toisessa ansioluettelossa olisikin tarvittava pätevyys, se saattaa jäädä huomiotta, jos kuvaukset eivät vastaa riittävän tarkasti yrityksen seulontaa varten määrittämiä avainsanoja. Tämä korostaa sitä, kuinka tärkeää on sovittaa ansioluettelon kieli tekoälyjärjestelmän avainsanoihin.

Näkökulma	CV 1	CV 2
Työnimikkeet ja terminologia	Käyttää vakiintuneita termejä kuten "Viestintäavustaja"	Käyttää nimikkeitä kuten "Digitaalisen sisällön tuottaja"
Akateeminen tausta	Käyttää selkeästi "alempi korkeakoulututkinto markkinoinnista"	Ilmoittaa "Yhteisöviestinnän kandidaatti (BSc)"
Tehtävien ja taitojen kuvaus	Suora vastaavuus avainsanoihin kuten "Sisällöntuotanto", "Tiedotteet"	Käyttää vaihtelevaa kieltä, kuten "vetovoimaisten tarinoiden rakentaminen", "johti <u>sisältöhankkeita</u> "
Tekniset taidot ja ohjelmistot	Mainitsee avainsanat "Photoshop" and "InDesign"	Yleinen maininta "digitaalisista <u>suunnittelutyökaluista</u> ja sisällönhallintajärjestel mistä"
Suorat <u>avainsanavastineet</u>	Vastaa tarkasti määriteltyihin avainsanoihin	Käyttää erilaista sanamuotoa, joka ei välttämättä vastaa tarkasti seulonnassa käytettyihin avainsanoihin

8. Kuinka mitata oikeudenmukaisuutta koneoppimisratkaisuissa



Ratkaisun oikeudenmukaisuuden testaaminen

Algoritmi-prototyypin arviointi paljasti epäjohdonmukaisuuksia ansioluetteloiden analysoinnissa, mikä johti parannuksiin CV-attribuuttien kattavamman ymmärtämisen saavuttamiseksi. Malli on nyt valmis testausvaiheeseen ennen käyttöönottoa todellisessa ympäristössä.

Vaikka arvioitavia osa-alueita on useita, keskitytään tässä sovelluksen oikeudenmukaisuuden testaamiseen, jotta voidaan selvittää prosessin aikana syntyneiden mahdollisten ennakkoluulojen laajuus. Oikeudenmukaisuusmittareita käytetään mittaamaan, kuinka paljon sovelluksen päätöksissä esiintyy puolueellisuutta.

Jotta voidaan luoda järjestelmiä, jotka ovat oikeudenmukaisia kaikille ja vapaita sukupuoleen tai rotuun liittyvistä vinoumista, on tärkeää arvioida ratkaisun toimivuutta eri väestöryhmien keskuudessa. Vaikka ratkaisun oikeudenmukaisuutta voidaan arvioida useilla tavoilla, keskitytään tässä yhteydessä yksinomaan binäärisiin luokittelutesteihin. (Machine Learning and Fairness, 2021)

Oikeudenmukaisuus on noussut keskeiseksi aiheeksi koneoppimisessa sen soveltamisen yleistyessä eri aloilla. Aiheeseen on kohdistunut laajaa tutkimusta, koska koneoppimisjärjestelmät perustuvat voimakkaasti dataan, joka on ihmisten tuottamaa ja siten usein jo lähtökohtaisesti vinoutunutta. Oikeudenmukaisuuden määrittelyminen matemaattisesti on osoittautunut haastavaksi, eikä yhteisesti hyväksyttyjä määritelmiä ole syntynyt. Zhongin (2018) mukaan suurin osa oikeudenmukaisuuden määritelmistä kuitenkin keskittyy useisiin keskeisiin käsitteisiin:

- Tietämättömyys (Unawareness)
- Demografinen tasapaino (Demographic parity)
- Tasapuoliset todennäköisyydet (Equalized odds)
- Ennustettavuuden tasapaino (Predictive rate parity)
- Yksilön oikeudenmukaisuus (Individual fairness)
- Vastemuuttujaoikeudenmukaisuus (Counterfactual fairness)

Väestöllinen tasapaino, tasapuoliset todennäköisyydet ja ennustettavuuden suhteen pariteetti luokitellaan tyypillisesti "ryhmäoikeudenmukaisuuden" (group fairness) alaisuuteen. (Zhong, 2018) Tätä käsitettä tarkastellaan tarkemmin tässä osiossa erityisesti tekoälyn käytön näkökulmasta rekrytoinnissa.

Ryhmäoikeudenmukaisuuden lisäksi Castelnovo ym. (2022) jakavat oikeudenmukaisuuden luokat yksilölliseen oikeudenmukaisuuteen ja kausaalisuuteen perustuvaan oikeudenmukaisuuteen. Näihin kuuluvat myös Zhongin (2018) määritelmien mukaiset näkökulmat.

Sekaannusmatriisi

Yksi binäärisen luokittelun menetelmistä on sekaannusmatriisi (Machine Learning and Fairness, 2021). Oikeudenmukaisuuden mittareiden ymmärtämiseksi aloitetaan sekaannusmatriisin käsitteestä. Tämä työkalu tiivistää mallin tekemät ennusteet suhteessa todellisiin tuloksiin (ns. totuuteen), joihin

malli on opetettu. Sekamatriisi esittää oikeiden ja virheellisten ennusteiden määrät, tarjoten selkeän kuvan mallin toimivuudesta ja sen selitettävissä olevasta pätevyydestä (Saplicki, 2022).

MIEHET	Epäpätevä	Pätevä	NAISET	Epäpätevä	Pätevä
Hylkäys	15 (TN)	5 (FN)	Hylkäys	60 (TN)	20 (FN)
Palkkaus	20 (FP)	60 (TP)	Palkkaus	5 (FP)	15 (TP)

Kuvan lähde | Sekaannusmatriisi-esimerkki rekrytoinnissa
(Muokattu lähteestä Machine Learning and Fairness, 2021)

Sekaannusmatriisi auttaa meitä luokittelemaan rekrytointiprosessin tuloksia. Päteviksi arvioidut ja palkatut hakijat luokitellaan **oikeiksi positiivisiksi** (True Positives, TP), ja epäpäteviksi arvioidut ja hylätyt hakijat ovat **oikeita negatiivisia** (True Negatives, TN). Vastaavasti **väärät positiiviset** (False Positives, FP) ovat epäpäteviä hakijoita, jotka on virheellisesti palkattu, kun taas **väärät negatiiviset** (False Negatives, FN) ovat päteviä hakijoita, jotka on virheellisesti hylätty.

Sekaannusmatriisi luokittelee siis hakijat eri ryhmiin heidän luokittelutulostensa perusteella. Matriisin neljää arvoa hyödynnetään laskemaan koneoppimisen standardimittareita, kuten tarkkuus (accuracy) ja osumatarkkuus (precision), eri väestöryhmien välillä (Machine Learning and Fairness, 2021; Teodorescu, 2020).

Tarkasteltavassa rekrytointitilanteessa keskitytään erityisesti sukupuolijakaumiin. Vaikka myös muut väestölliset tekijät ovat tärkeitä, tämä esimerkki keskittyy yksinomaan sukupuoleen.

Oletetaan, että hakijoita on yhtä monta – 100 miestä ja 100 naista. Järjestelmä on tunnistanut tarkasti 75 henkilöä kummastakin ryhmästä joko päteviksi tai epäpäteviksi, mikä osoittaa, että tarkkuus on samanlainen molemmissa sukupuoliryhmissä (Machine Learning and Fairness, 2021).

MEN	Unqualified	Qualified	WOMEN	Unqualified	Qualified
Reject	15 (TN)	5 (FN)	Reject	60 (TN)	20 (FN)
Hire	20 (FP)	60 (TP)	Hire	5 (FP)	15 (TP)

KUVAN LÄHDE | Kuva 2 - Sekaannusmatriisin esimerkki rekrytoinnissa (Modified from Machine Learning and Fairness, 2021)

Väestöllinen tasapaino

Perustuen aiemmin esitettyyn dataan tarkastelemme, miten erilaiset oikeudenmukaisuuden mittarit soveltuvat tähän kontekstiin, alkaen väestöllisestä

tasapainosta. Tämä mittari tarkastelee vain luokittelijan antamaa tulosta, ottamatta huomioon todellisia luokkia. Väestöllinen tasapaino on yksi ehdotetuista mittareista automaattisen rekrytointijärjestelmän arvioinnissa (Machine Learning and Fairness, 2021).

Väestöllinen tasapaino, joka tunnetaan myös nimillä *independence* tai *statistical parity*, on oikeudenmukaisuuden kriteeri, jonka mukaan myönteisen lopputuloksen, kuten palkkauksen, todennäköisyys ei saa riippua suojellusta ominaisuudesta, kuten sukupuolesta (Zhong, 2018).

MIEHET	Epäpätevä	Pätevä	NAISET	Epäpätevä	Pätevä
Hylkäys	15 (TN)	5 (FN)	Hylkäys	60 (TN)	20 (FN)
Palkkaus	20 (FP)	60 (TP)	Palkkaus	5 (FP)	15 (TP)

Kuvan lähde | Väestöllinen tasapaino rekrytoinnissa
(Muokattu lähteestä Machine Learning and Fairness, 2021)

Rekrytointiesimerkissä (kuva 3), jos 80 sadasta mieshakijasta palkataan mutta vain 20 sadasta naishakijasta saa paikan, väestöllinen tasapaino ei täyty. Tämä saattaa kuitenkin olla hyväksyttävää, jos mieshakijat ovat todellisuudessa pätevämpiä kuin naishakijat, kuten tässä esimerkissä vaikuttaa olevan (Machine Learning and Fairness, 2021).

Demografisen tasapainon saavuttaminen on haastavaa, kun henkilöt kuuluvat useisiin suojeltuihin ryhmiin, koska tasaiset todennäköisyydet kaikissa ryhmissä eivät välttämättä ole mahdollisia. Tämä mittari voi myös tahtomattaan suosia vähemmän päteviä hakijoita, koska se tarkastelee vain ryhmätason oikeudenmukaisuutta, mikä voi johtaa yksilötason epäoikeudenmukaisuuteen. Esimerkiksi suurempi osuus vähemmän päteviä naishakijoita voi vähentää pätevien mieshakijoiden palkkausmahdollisuuksia (Teodorescu, 2020). Näin ollen tämä mittari tarkastelee ainoastaan lopputulosta, ei todellisia luokkia – eli se ei huomioi hakijoiden pätevyyttä. Mittarin mukaan palkattujen ja hakeneiden suhdeluvun tulisi olla samanlainen kaikissa ryhmissä (Machine Learning and Fairness, 2021).

Ennustettavuuden tasapaino (Predictive Parity)

Sufficiency-periaatteen mukaan henkilöiden, jotka saavat saman päätöksen mallilta, tulisi kohdata samanlaiset virheprosentit riippumatta heidän hermistä ominaisuuksistaan. Tämä käsite sisältyy ennustettavuuden pariteettiin (*predictive parity*), jota kutsutaan myös *outcome testiksi*. Se varmistaa, että virheprosentit ovat tasapainossa niiden henkilöiden välillä, jotka saavat saman päätöksen, ei niiden, joiden todelliset lopputulokset ovat samanlaisia.

Toisin kuin demografinen tasapaino, ennustettavuuden tasapaino huomioi sekä luokittelijan päätökset että todelliset lopputulokset. Se arvioi, onko tekoälyn palkkaamien hakijoiden pätevyyden todennäköisyys yhtenevä eri ryhmissä. Oletuksena on, että tämä todennäköisyys on lähes sama kaikissa ryhmissä (Castelnovo et al., 2022; Machine Learning and Fairness, 2021).

MEN	Unqualified	Qualified	WOMEN	Unqualified	Qualified
Reject	15	5	Reject	60	20
Hire	20	60	Hire	5	15

Kuvan lähde | Ennustettavuuden tasapaino rekrytinnissa
(Muokattu lähteestä Machine Learning and Fairness, 2021)

Tarkastellussa rekrytointiesimerkissä järjestelmä täyttää ennustettavuuden tasapainon kriteerit, koska kolme neljäsosaa kummankin ryhmän palkatuista hakijoista on todellisuudessa päteviä (Machine Learning and Fairness, 2021).

Väärin positiivisten ja negatiivisten tasapaino & Tasapuoliset todennäköisyydet (Equalized Odds)

Väärän positiivisen määrän tasapaino (*false positive rate balance*) tarkoittaa, että epäpäteväksi arvioidun hakijan virheellisen palkkaamisen todennäköisyyden tulisi olla sama eri väestöryhmissä. Yksinkertaisesti ilmaistuna järjestelmän tulisi tehdä yhtä paljon virheitä epäpäteviä hakijoita kohtaan riippumatta esimerkiksi sukupuolesta (Machine Learning and Fairness, 2021).

MIEHET	Epäpätevä	Pätevä	NAISET	Epäpätevä	Pätevä
Hylkäys	15	5	Hylkäys	60	20
Palkkaus	20	60	Palkkaus	5	15

Kuvan lähde | Väärin positiivisten suhde rekrytinnissa
(Muokattu lähteestä Machine Learning and Fairness, 2021)

Tässä esimerkissä (kuva 5) väärin positiivisten määrissä on eroja, sillä huomattavasti suurempi osuus epäpäteviä miehiä palkataan kuin naisia.

Väärän negatiivisen määrän tasapaino (*false negative rate balance*) on vastaava käsite väärille negatiivisille. Se tarkoittaa, että pätevien hakijoiden virheellisen hylkäämisen todennäköisyyden tulisi olla sama kaikissa ryhmissä.

MIEHET	Epäpätevä	Pätevä	NAISET	Epäpätevä	Pätevä
Hylkäys	15	5	Hylkäys	60	20
Palkkaus	20	60	Palkkaus	5	15

Kuvan lähde | Väärien negatiivisten suhde rekrytoinnissa
(Muokattu lähteestä Machine Learning and Fairness, 2021)

Tässä tapauksessa (kuva 6) väärien negatiivisten määrissä on selvä ero, sillä päteviä naishakijoita hylätään suuremmalla todennäköisyydellä kuin mieshakijoita.

Tasapuoliset todennäköisyydet (*equalized odds*) yhdistävät nämä kaksi ominaisuutta: järjestelmän tulee käsitellä sekä väärin hyväksytyjä epäpäteviä että väärin hylättyjä päteviä hakijoita tasapuolisesti eri ryhmissä. Tavoitteena on varmistaa, että samanlaiset hakijat saavat samanlaista kohtelua sukupuolesta tai muista tekijöistä riippumatta (Machine Learning and Fairness, 2021). Equalized odds on vaikeasti saavutettavissa, mutta sen toteutuessa se voidaan katsoa yhdeksi korkeimmista algoritmisen oikeudenmukaisuuden muodoista (Teodorescu, 2020).

Mittareiden väliset kompromissit (Trade-offs with metrics)

Järjestelmä ei voi samanaikaisesti täyttää ennustettavuuden pariteettia, väärien positiivisten määrän tasapainoa ja väärien negatiivisten määrän tasapainoa matemaattisten rajoitteiden vuoksi. Jos järjestelmä täyttää kaksi näistä kriteereistä, se väistämättä epäonnistuu kolmannessa (Machine Learning and Fairness, 2021).

Tämä teoreema korostaa eri oikeudenmukaisuustavoitteiden välisten kompromissien väistämättömyyttä. Kun pyritään oikeudenmukaisuuteen, törmätään usein ristiriitaisiin tavoitteisiin, eikä täydellinen tasapaino kaikkien mittareiden välillä ole realistinen.

Käytännössä päätöksentekijöiden on punnittava näitä kompromisseja tarkasti, otettava huomioon konteksti ja tehtävä perusteltuja valintoja sovelluksen erityistarpeiden mukaan. Vaikka täydellisyyttä ei voida saavuttaa, jatkuva pyrkimys oikeudenmukaisuuteen ja läpinäkyvyyteen on olennaista algoritmisten päätöksentekojärjestelmien kehittämisessä.

Lopuksi on tärkeää tunnustaa, että monet oikeudenmukaisuuden ulottuvuudet ovat niin sosio-tekniisesti monimutkaisia, ettei niitä voida mitata numeerisesti. Kaikkia oikeudenmukaisuuden osa-alueita ei siis voida tiivistää mittareiksi (Machine Learning and Fairness, 2021).

Yksilön oikeudenmukaisuus (Individual fairness)

Yksilötason oikeudenmukaisuus poikkeaa aiemmin käsitellyistä ryhmätason mittareista keskittymällä yksittäisten henkilöiden kohteluun. Tämän käsitteen mukaan samankaltaisia ominaisuuksia omaavien yksilöiden tulisi saada samankaltaiset lopputulokset.

Yksilöiden samankaltaisuuden mittaaminen on kuitenkin haastavaa. Kuvitellaan kolme työnhakijaa:

- A: kandidaatin tutkinto ja vuoden työkokemus alalta
- B: maisterin tutkinto ja vuoden työkokemus
- C: maisterin tutkinto mutta ei työkokemusta

On vaikea määrittää, onko A lähempänä B:tä vai C:tä – ja kuinka tätä samankaltaisuutta voitaisiin mitata – etenkin tilanteessa, jossa heidän työpanostaan ei voida vertailla, jos heitä ei palkata. (Zhong, 2018)

9. Ratkaisun käyttöönotto



Hakijoiden informoiminen tekoälyn käytöstä rekrytinnissa

Huolellisten sisäisten arviointien jälkeen yritys on valmis ottamaan seulontatyökalunsa käyttöön valikoidusti avoimissa tehtävissä kerätäkseen alkuvaiheen palautetta käyttäjiltä ja hakijoilta.

On erittäin tärkeää, että hakijat ovat jatkuvasti tietoisia siitä, että he ovat tekemisissä tekoälyohjatun rekrytointijärjestelmän kanssa – tämä edistää luottamuksen rakentumista. Tietoisuus tekoälytyökalujen hyödyistä tulisi luoda jo varhaisessa vaiheessa, mikä lisää hakijoiden halukkuutta käyttää vuorovaikutteista teknologiaa, kuten chatbotteja. Lisäksi tekoälyn perustuvien rekrytointiprosessien kehittäminen ja järjestelmän toiminnan selkeyttäminen ovat keskeisiä sen hyväksyttävyyden ja onnistumisen kannalta hakijoiden keskuudessa.

On myös varmistettava, että mukana on ihminen valvomassa prosessia ja puuttumassa tilanteisiin tarvittaessa. (Chen, 2023)

10. Toiminta ja seuranta



OPERATION AND MONITORING



Käyttöönoton jälkeen

Vaikka sisäisiä arviointeja oikeudenmukaisuudesta tehtiin ennen käyttöönottoa, yrityksen on oltava valmis myös riippumattomiin ulkoisiin auditointeihin.

On myös hyödyllistä hakea jatkuvasti palautetta käyttöönoton jälkeen ja luoda pysyviä vuorovaikutuskanavia sidosryhmien kanssa. Tämä tarkoittaa työntekijöiden pitämistä ajan tasalla ja heidän osallistamistaan kuulemis- ja osallistumisprosesseihin koko tekoälyjärjestelmän käyttöönoton ajan organisaatiossa. (HLEG Trustworthy AI)

Tekoälyn auditointi

Vaikka sisäiset oikeudenmukaisuuden mittaukset tehdään ennen käyttöönottoa, yrityksen on varauduttava myös ulkoiseen auditointiin.

Ahmedin (2020) mukaan tekoälysovelluksia arvioivien auditointien tulisi keskittyä kahteen pääkohtaan: tietosubjektien oikeuksien toteutumisen varmistamiseen sekä koneoppimiseen ja kyberturvallisuuteen liittyvien teknologisten riskien arviointiin. Auditointi alkaa sen laajuuden, tavoitteiden ja tekoälyn liittyvien organisaatoriskien määrittelystä, jotka dokumentoidaan yleensä riskien ja kontrollien matriisissa osana vakiintuneita viitekehyksiä.

Keskeisiä riskejä ovat muun muassa IT-strategioiden ja liiketoimintatavoitteiden välinen ristiriita, hallintorakenteiden tehottomuus sekä epäselvä vastuunjako. Sääntelyn näkökulmasta auditointien on ymmärrettävä tietosuojaperiaatteet ja niiden vaikutukset yksilöiden oikeuksiin esimerkiksi yleisen tietosuoja-asetuksen (GDPR) mukaisesti. (Ahmed, 2020)

Tekoälyn auditointiin on olemassa useita ohjeistuksia. Koska tekoäly muuttaa liiketoimintaa ja yhteiskuntaa, auditointien on varmistettava, että he ovat valmiita vastaamaan uusiin haasteisiin ja tarkistamaan, että tekoälyn hallinta- ja valvontajärjestelmät toimivat asianmukaisesti. (Ahmed, 2020)

11. Yhteenveto

Tämä kurssikokonaisuus keskittyi algoritmisten vinoumien tarkasteluun käytännön kontekstissa. Teemaa käsiteltiin hypoteettisen tilanteen kautta, jossa kansainvälinen yritys kehittää tekoälyratkaisua rekrytointiin.

Tekoälyteknologioiden käyttöönotto rekrytoinnissa tarjoaa merkittäviä etuja, kuten tehokkuuden parantamista ja mahdollisuuksia lisätä työpaikkojen monimuotoisuutta painottamalla pätevyyttä demografisten tekijöiden sijaan. Näiden teknologioiden käyttöönottamiseen liittyy kuitenkin myös merkittäviä eettisiä haasteita, erityisesti algoritmisten vinoumien tunnistamisessa ja vähentämisessä.

Tekoälyjärjestelmät voivat tahattomasti ylläpitää koulutusdatan sisältämiä olemassa olevia vinoumia. Nämä vinoumat voivat heijastaa historiallisia epätasa-arvoja tai yhteiskunnallisia eriarvoisuuksia, mikä voi johtaa syrjiviin käytäntöihin rekrytointiprosesseissa. Näiden vinoumien käsittely on paitsi eettinen velvollisuus, myös oikeudellinen välttämättömyys, sillä tekoälyn käyttöä rekrytoinnissa koskeva sääntely kehittyy jatkuvasti. Organisaatioiden on oltava valppaita ja ennakoivia näiden standardien noudattamisessa syrjinnän ehkäisemiseksi.

Näiden haasteiden tehokas torjuminen edellyttää kehittyneiden työkalujen käyttöä, jotka on suunniteltu erityisesti tekoälyjärjestelmien vinoumien tunnistamiseen ja korjaamiseen. Säännöllinen seuranta ja riippumattomat auditoinnit ovat olennaisia, sillä ne auttavat organisaatioita havaitsemaan uusia vinoumia ja säätämään järjestelmiään vastaavasti.

Lisäksi on tärkeää säilyttää tasapaino tekoälyn ohjaamien prosessien ja inhimillisen valvonnan välillä. Vaikka tekoäly voi merkittävästi tehostaa rekrytointia, ihmisen rooli on edelleen korvaamaton erityisesti tilanteiden tulkinnassa, joita tekoäly ei välttämättä ymmärrä oikein. Inhimillinen valvonta varmistaa, että rekrytointi pysyy sensitiivisenä ja osallistavana prosessina, joka huomioi inhimillisen kokemuksen ja arvojen hienovaraisuuden.

Yhteenvetona voidaan todeta, että vaikka tekoälyn käyttö rekrytoinnissa tarjoaa merkittäviä etuja, tämän teknologian vastuullinen hyödyntäminen on ratkaisevan tärkeää. Organisaatioiden tulisi jatkuvasti tunnistaa vinoumia ja käyttää kehittyneitä työkaluja niiden vähentämiseksi. Tällainen tasapainoinen lähestymistapa varmistaa, että rekrytointiprosessit noudattavat korkeimpia oikeudenmukaisuuden ja eettisen toiminnan standardeja, ja mahdollistavat samalla teknologisen kehityksen hyödyntämisen työelämän monimuotoisuuden ja tasa-arvon turvaamiseksi.

12. Lähteet

- Ahmed, H. S. A. (2020, December 21). *Auditing Guidelines for Artificial Intelligence*. ISACA. <https://www.isaca.org/resources/news-and-trends/newsletters/atisaca/2020/volume-26/auditing-guidelines-for-artificial-intelligence>
- Alan Turing Institute. (2022). *Reflect on Purpose, Positionality, and Power—Turing Commons*. <https://alan-turing-institute.github.io/turing-commons/skills-tracks/aeg/chapter5/reflect/>
- Albaroudi, E., Mansouri, T., & Alameer, A. (2024). A Comprehensive Review of AI Techniques for Addressing Algorithmic Bias in Job Hiring. *AI*, 5(1), 383–404. <https://doi.org/10.3390/ai5010019>
- Albassam, W. A. (2023). The Power of Artificial Intelligence in Recruitment: An Analytical Review of Current AI-Based Recruitment Strategies. *International Journal of Professional Business Review*, 8(6), e02089. <https://doi.org/10.26668/businessreview/2023.v8i6.2089>
- ALTAI. (2020). *The Assessment List for Trustworthy Artificial Intelligence*. <https://altai.insight-centre.org/>
- Arivu Recruitment and Consulting. (2023, July 13). *Pros and Cons of Using Artificial Intelligence (AI) in Recruiting* | LinkedIn. <https://www.linkedin.com/pulse/pros-cons-using-artificial-intelligence/>
- Banerjee, S. (2022, February 20). *Big data in Recruitment Sector* | LinkedIn. <https://www.linkedin.com/pulse/big-data-recruitment-sector-shantanu-banerjee-phd/>
- Bursell, M., & Roumbanis, L. (2024). After the algorithms: A study of meta-algorithmic judgments and diversity in the hiring process at a large multisite company. *Big Data & Society*, 11(1), 20539517231221758. <https://doi.org/10.1177/20539517231221758>
- CareerExperts. (2023, April 24). Reducing Bias in Hiring with AI-Powered Job Description Generation. *Career Experts*. <https://www.careerexperts.co.uk/management-leadership/ai-powered-job-description>
- Castelnovo, A., Crupi, R., Greco, G., Regoli, D., Penco, I. G., & Cosentini, A. C. (2022). A clarification of the nuances in the fairness metrics landscape. *Scientific Reports*, 12(1), 4209. <https://doi.org/10.1038/s41598-022-07939-1>
- Chaker, N. (2018, October 24). *Candidate Data Sources: The Recruiter's Checklist*. <https://beamery.com/resources/blogs/candidate-data-sources-the-recruiter-s-checklist>
- Chen, Z. (2023). Collaboration among recruiters and artificial intelligence: Removing human prejudices in employment. *Cognition, Technology & Work*, 25(1), 135–149. <https://doi.org/10.1007/s10111-022-00716-0>
- Dennis, J. (2023, May 22). *AI Recruiting: Uses, Advantages, & Disadvantages 2024*. TechnologyAdvice. <https://technologyadvice.com/blog/human-resources/ai-recruiting/>
- DigitalOcean. (n.d.). *How to Use AI in Hiring: Techniques and Tools*. Retrieved April 26, 2024, from <https://www.digitalocean.com/resources/article/ai-in-hiring>

- Featured. (2023, December 22). *How these 6 leaders are using AI-powered onboarding*. Fast Company. <https://www.fastcompany.com/90996367/how-leaders-using-ai-powered-onboarding>
- Fernandes, R. F. (2021). Big Data as a Tool to Enhance Recruitment Processes. *E-Journal of International and Comparative LABOUR STUDIES*, 10(03).
- Filtered. (n.d.). *What is AI Resume Screening and How Can it Benefit Your Enterprise?* / Filtered Blog. Retrieved March 14, 2024, from <https://www.filtered.ai/blog/ai-resume-screening-fd>
- Friedman, B., & Nissenbaum, H. (1996). Bias in Computer Systems. *ACM Transactions on Information Systems*, 14(3), 330–347.
- Fritts, M., & Cabrera, F. (2021). AI recruitment algorithms and the dehumanization problem. *Ethics and Information Technology*, 23(4), 791–801. <https://doi.org/10.1007/s10676-021-09615-w>
- Giordano, F., Morelli, N., Götzen, A. D., & Hunziker, J. (2018, June). *The stakeholder map: A conversation tool for designing people-led public services*. ServDes2018 - Service Design Proof of Concept, Politecnico di Milano.
- Google for Developers. (2022). *Fairness: Types of Bias / Machine Learning*. Google for Developers. <https://developers.google.com/machine-learning/crash-course/fairness/types-of-bias>
- Harvard John A. Paulson School of Engineering and Applied Sciences. (2023, December 6). *How Can Bias Be Removed from Artificial Intelligence-Powered Hiring Platforms?* <https://seas.harvard.edu/news/2023/06/how-can-bias-be-removed-artificial-intelligence-powered-hiring-platforms>
- Harwell, D. (2019a, July 7). FBI, ICE find state driver's license photos are a gold mine for facial-recognition searches. *Washington Post*. <https://www.washingtonpost.com/technology/2019/07/07/fbi-ice-find-state-drivers-license-photos-are-gold-mine-facial-recognition-searches/>
- Harwell, D. (2019b, December 19). Federal study confirms racial bias of many facial-recognition systems, casts doubt on their expanding use. *Washington Post*. <https://www.washingtonpost.com/technology/2019/12/19/federal-study-confirms-racial-bias-many-facial-recognition-systems-casts-doubt-their-expanding-use/>
- Heymans, Y. (2022, September 7). *Machine learning in recruitment: A deep dive*. <https://www.herohunt.ai/blog/machine-learning-in-recruitment-a-deep-dive>
- Hidalgo, M. A. S. (2019). *Design of an Ethical Toolkit for the Development of AI Applications* [Delft University of Technology]. <http://resolver.tudelft.nl/uuid:b5679758-343d-4437-b202-86b3c5cef6aa>
- Hoory, L., & Botorff, C. (2022, August 7). *What Is A Stakeholder Analysis? Everything You Need To Know – Forbes Advisor*. <https://www.forbes.com/advisor/business/what-is-stakeholder-analysis/>
- Hunkenschroer, A. L., & Luetge, C. (2022). Ethics of AI-Enabled Recruiting and Selection: A Review and Research Agenda. *Journal of Business Ethics*, 178(4), 977–1007. <https://doi.org/10.1007/s10551-022-05049-6>
- Javaid, S. (2024, January 12). *Facial Recognition: Best Practices & Use Cases in 2024*. AIMultiple: High Tech Use Cases & Tools to Grow Your Business. <https://research.aimultiple.com/facial-recognition/>

- Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13(3), 795–848. <https://doi.org/10.1007/s40685-020-00134-w>
- Kumar, V. (2013). *101 Design Methods—A Structured Approach for Driving Innovation in Your Organization*. John Wiley & Sons, Inc. <https://learning.oreilly.com/library/view/101-design-methods/9781118083468/cvi.xhtml>
- Lange, A. R., & Duarte, N. (2018, April 4). Understanding Bias in Algorithmic Design. *ASME ISHOW / IDEA LAB*. <https://medium.com/impact-engineered/understanding-bias-in-algorithmic-design-db9847103b6e>
- Lawton, G. (2022, August 29). *AI hiring bias: Everything you need to know* / *TechTarget*. HR Software. <https://www.techtarget.com/searchhrsoftware/tip/AI-hiring-bias-Everything-you-need-to-know>
- Lee, N. T., Resnick, P., & Barton, G. (2019, May 22). *Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms* / *Brookings*. <https://www.brookings.edu/articles/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/>
- Ma, Y. (2024). Intelligent Recommendation Method for Talent Resources Based on Big Data. *Intelligent Computing Technology and Automation*. <https://doi.org/10.3233/ATDE231176>
- Machine Learning and Fairness*. (2021, May 24). Microsoft Research webinars and tutorials. <https://youtu.be/ZtN6Qx4KddY?si=kWo9MxkNQ7ko1HUB>
- Marr, B. (2023, December 12). *AI-Enhanced Employee Onboarding: A New Era In HR Practices*. Forbes. <https://www.forbes.com/sites/bernardmarr/2023/12/12/ai-enhanced-employee-onboarding-a-new-era-in-hr-practices/>
- Naakka, S.-S. (2018). *BIG DATA KILPAILUETUNA SUORAHAKU- KONSULTOINNIN EHDOKASHAUISSA Enemmän, nopeammin ja parempia IT-osaajia big dataa?* [University of Turku]. <https://www.utupub.fi/bitstream/handle/10024/145309/Naakka%20Sara-Selia.pdf?sequence=1&isAllowed=y>
- Ongig. (2024). *Job Description Software* / *Ongig*. <https://www.ongig.com>
- ResumeMent. (2023, October 16). *From Resumes to Algorithms: The Role of AI in Screening Candidates* / *LinkedIn*. <https://www.linkedin.com/pulse/from-resumes-algorithms-role-ai-screening-candidates-resument/>
- Saplicki, C. (2022, November 11). Fairness Explained: Definitions and Metrics. *Medium*. <https://medium.com/ibm-data-ai/fairness-explained-definitions-and-metrics-9690f8e0a4ea>
- SDT. (n.d.). *Tools* / *Service Design Tools*. Retrieved April 18, 2024, from <https://servicedesigntools.org/tools.html>
- Sheard, N. (2022). Employment Discrimination by Algorithm: Can Anyone Be Held Accountable? *University of New South Wales Law Journal*, 45(2). <https://doi.org/10.53637/XTQY4027>

- Shipman, M. (2021, March 4). Social media checks can bring bias into hiring. *Futurity*. <https://www.futurity.org/cybertvetting-human-resources-bias-2527382-2/>
- Stickdorn, M., Hormess, M. E., Lawrence, A., & Schneider, J. (2018a). *This is Service Design Doing*. O'Reilly Media, Inc. <https://learning.oreilly.com/library/view/this-is-service/9781491927175/ch03.html>
- Stickdorn, M., Hormess, M., Lawrence, A., & Schneider, J. (2018b, July). *#TISDD Method Library*. This Is Service Design Doing. <https://www.thisisservicedesigndoing.com/methods>
- Sukernek, W. (2024). *How to Spot Bias in Hiring*. FloCareer. <https://blog.flocareer.com/how-to-spot-bias-in-hiring>
- Teodorescu, M. (2020). *Fairness Criteria | Exploring Fairness in Machine Learning for International Development | Supplemental Resources*. MIT OpenCourseWare. <https://ocw.mit.edu/courses/res-ec-001-exploring-fairness-in-machine-learning-for-international-development-spring-2020/pages/module-three-framework/fairness-criteria/>
- UNESCO. (2023). *Ethical impact assessment. A tool of the Recommendation on the Ethics of Artificial Intelligence*. UNESCO. <https://doi.org/10.54678/YTSA7796>
- Vivek, R. (2023). Enhancing diversity and reducing bias in recruitment through AI: A review of strategies and challenges. *Информатика. Экономика. Управление - Informatics. Economics. Management*, 2(4), 0101–0118. <https://doi.org/10.47813/2782-5280-2023-2-4-0101-0118>
- Wallbridge, A. (2023, April 13). *How To Conduct A Bullet-Proof Stakeholder Analysis Matrix In 4 Useful Steps*. TSW Training. <https://www.tsw.co.uk/blog/leadership-and-management/stakeholder-analysis-matrix/>
- Wirtz, D. (2022, August 3). *What Is a Workshop? (+2 Examples) | Facilitator School*. <https://www.facilitator.school/blog/what-is-a-workshop>
- www.recruiter.com. (2023, April 26). *Pros and Cons of Using AI in Recruiting*. Recruiter.Com. <https://www.recruiter.com/recruiting/pros-and-cons-of-using-ai-in-recruiting/>
- YLE. (2023, November 11). *Loimme 13-vuotiaan Ellan – Tiktok tarjosi hänelle sisältöä itsemurhasta ja kalorien laskemisesta*. Yle Uutiset. <https://yle.fi/a/74-20059318>
- Zhong, Z. (2018, October 22). *A Tutorial on Fairness in Machine Learning*. Medium. <https://towardsdatascience.com/a-tutorial-on-fairness-in-machine-learning-3ff8ba1040cb>



CharlTe



Co-funded by
the European Union



Universitat
de les Illes Balears



2022-1-ES01-KA220-HED-000085257