

Tervetuloa Algoritmin perusteet ja tekoälyn etiikka: Teoriasta käytäntöön -kurssille

Työkalu synkronoiduille tunneille

CU5 | Tapaustutkimus ja projektit
Tuki PowerPoint slidet

INDEKSI

- JOHDANTO – 3
- ALGORITMINEN JA KONEOPPIMISEN VINOUMAT – 6
- ALGORITMINEN VINOUMA JA NIIDEN VAIKUTUKSET - TAPAUSTEN ESIMERKIT - 16
- TYÖKALUT ALGORITMISTEN VINOUMIEN TUNNISTAMISEKSI – 28
- PROJEKTI | TEKOÄLY REKRYTOINNISSA- 35

JOHDANTO



IMAGE SOURCE | Freepik

TÄSTÄ OSAAMISKOKONAISUUDESTA LÖYDÄT SEURAAVAT AIHEET

- Erilaisten algoritmisten ja koneoppimiseen liittyvien vinoumien esittely tapaustutkimusten kautta, jotka käsittelevät esimerkiksi TikTokia ja kasvojentunnistusteknologioita.
- Yleiskatsaus työkaluihin, jotka on suunniteltu algoritmisten vinoumien tunnistamiseen ja niiden seurausten arviointiin.
- Syvällinen projektianalyysi tekoälyn käytöstä rekrytoinnissa, korostaen vinoumia eri projektivaiheissa.
- Oikeudenmukaisuustietoisten lähestymistapojen rooli koneoppimisessa ja vinouma-auditointiprosessien toteuttaminen.

OSAAMISKOKONAISUUDEN LOPUSSA OSAAT:

- Tunnistaa erilaisia algoritmisten vinoumien tyyppejä ja syitä sekä esimerkkejä koneoppimisdatasettien vinoumista.
- Arvioida algoritmisten vinoumien vaikutuksia todellisessa maailmassa.
- Analysoida eri työkaluja algoritmisten vinoumien havaitsemiseksi.
- Analysoida projektin eri vaiheet – Tapaus “Tekoäly rekrytoinnissa”.

ALGORITMINEN JA KONEOPPIMISEN VINOUMAT



IMAGE SOURCE | Freepik

Algoritminen ja koneoppimisen vinoumat

WHAT IS ALGORITHMIC BIAS AND WHY IS IT IMPORTANT TO CONSIDER?

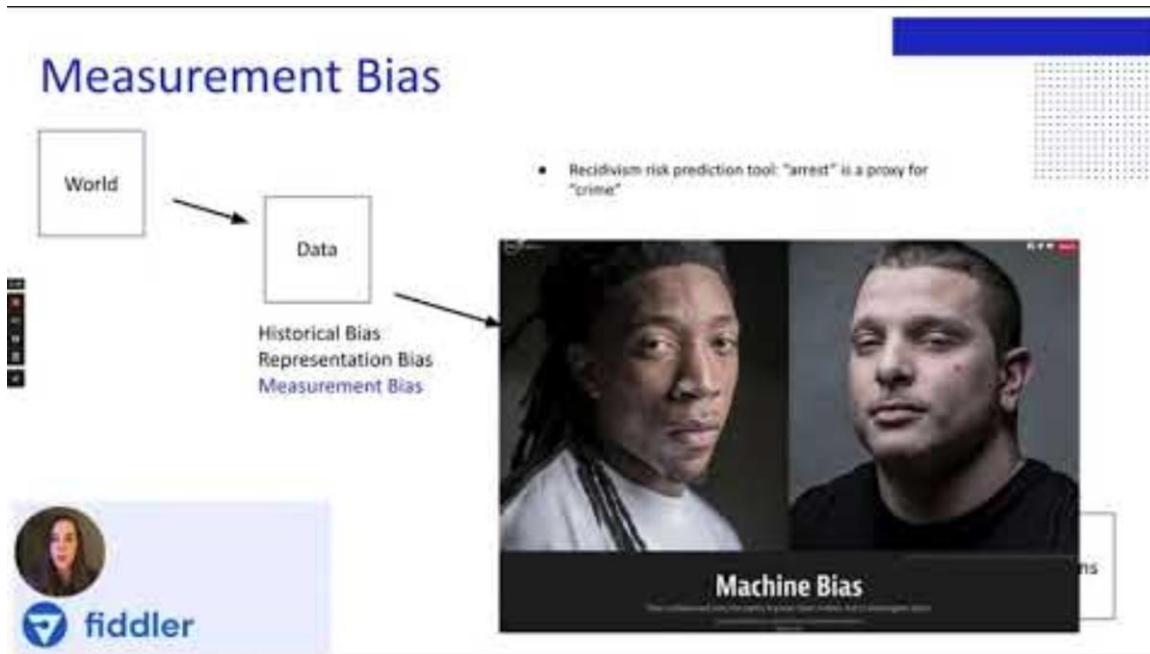


IMAGE SOURCE | MISSING INFO

- In machine learning, algorithms use data sets or training data to learn patterns.
- Algorithmic bias occurs when the algorithm makes decisions that systematically disadvantage or discriminate against certain groups of people, for example, based on race or gender.
- Algorithmic bias can appear if a certain group of people is underrepresented in the data or in case existing societal biases are embedded within the data itself.
- It is crucial to understand the causes and implications of algorithmic bias in order to develop ethical and trustworthy AI solutions.

(Harvard Business Review, S. Friis & J. Riley, 2023)

Algoritminen ja koneoppimisen vinoumat

Algoritmissen vinouman eri tyyppejä:

Algoritmiset vinoumat voidaan jakaa kolmeen eri esimerkkikategoriaan:



1. **Olemassa oleva vinouma.** Vinouma järjestelmässä, joka on olemassa itsenäisesti ja yleensä ennen järjestelmän luomista. Ne voivat tulla järjestelmään tahallisesti tai tahattomasti, joko tietoisilla toimilla tai tiedostamattomilla teoilla, vaikka aikomus olisi hyvä.



2. **Tekninen vinouma** syntyy teknisistä rajoitteista tai teknisistä näkökohdista. Teknisen vinouman lähteitä voidaan löytää monista suunnitteluprosessin vaiheista.



3. **Nouseva vinouma** syntyy käytön yhteydessä todellisten käyttäjien kanssa. Tämä vinouma ilmenee yleensä vasta suunnittelun valmistuttua, johtuen yhteiskunnallisen tiedon, väestön tai kulttuuriarvojen muutoksista.

Friedman and Nissenbaum (1996)

Algoritminen ja koneoppimisen vinoumat

Eri vinoumien syyt

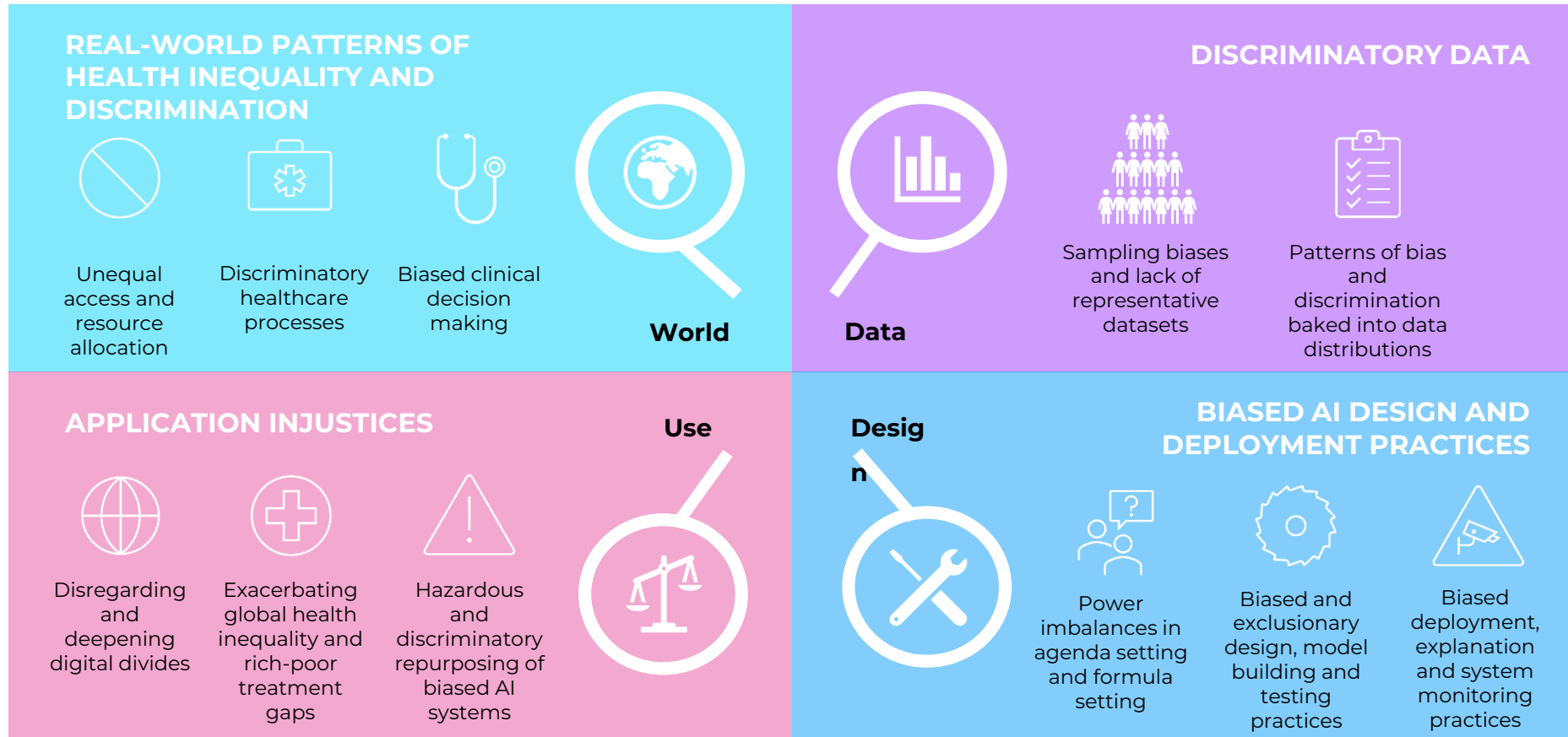


IMAGE SOURCE | Image redesigned (the original image: British Medical Journal)

igitaaliset teknologiat taipuvat vahvistamaan ja nopeuttamaan eriarvoisuuksia, jotka voivat ilmetä esim. syrjivinä sovelluksina. ämä on syy tekoälyn päätöksenteon syrjintään, joka johtuu eettisistä puutteista näiden teknologioiden kehityksessä, käyttöönotossa tai saatavuudessa.

Algoritminen ja koneoppimisen vinoumat

Esimerkkejä vinoumien tyypeistä koneoppimisen data-aineistossa

Insinöörit kouluttavat koneoppimismalleja syöttämällä niille joukon harjoitusesimerkkejä. Ihmisten osallistuminen datan valintaan ja syöttöön voi aiheuttaa vinoumia.

Analysoidaan seuraavaksi alla olevia vinoumien tyyppejä:

- Raportointivinouma
- Ryhmäattribuutiovinouma
- Piilevä vinouma
- Automaatiovinouma
- Valintavinouma

Algoritminen ja koneoppimisen vinoumat

Esimerkkejä vinoumien tyypeistä koneoppimisen data-aineistossa

Raportointivinouma

- Vinouma syntyy, kun tietyn tapahtumien, ominaisuuksien tai tulosten esiintyminen datasetissä ei vastaa todellista esiintymistiheyttä.
- Vinouma syntyy, kun keskitytään dokumentoimaan epätavallisia tai mieleenpainuvia tilanteita.
- Tavalliset tilanteet voidaan sivuuttaa, koska niiden ajatellaan olevan itsestäänselviä.

Esimerkki

Tunneanalyysimallin koulutuksessa ennustetaan, onko kirja-arvostelu positiivinen vai negatiivinen käyttäjien lähettämien arvostelujen perusteella. Suurin osa koulutusdatan arvosteluista ilmaisee äärimielipiteitä (joko rakastetaan tai vihataan kirjaa), koska ihmiset jättävät harvemmin arvostelua, jos reaktio on kohtalainen. Tämän seurauksena malli ei osaa ennustaa oikein arvosteluja, joissa kirjaa kuvaillaan hienovaraisella kielellä.

Algoritminen ja koneoppimisen vinoumat

Esimerkkejä vinoumien tyypeistä koneoppimisen data-aineistossa

Ryhmäattribuutiovinouma

Taipumus liittää yksilön ominaisuuksia koko siihen ryhmään, johon hän kuuluu. On kaksi päämuotoa:

- **Sisäryhmävinouma:** suositaan oman ryhmän jäseniä tai yhteisiä ominaisuuksia.

Esimerkki | Kaksi insinööriä, jotka kouluttavat ansioluettelon seulontamallia ohjelmistokehittäjille, voivat suosia hakijoita samasta tietojenkäsittelyakatemiasta, jossa he itse opiskeli, olettaen heidän olevan pätevämpiä.

- **Ulkoryhmän homogeenisuusvinouma:** stereotypisoidaan ryhmän jäseniä, johon ei itse kuulu, pitäen heidän ominaisuuksiaan samankaltaisina.

Example | Kaksi insinööriä kouluttamassa ansioluettelomallia saattavat olettaa, että kaikki hakijat, jotka eivät käyneet tiettyä tietojenkäsittelyakatemiasa, eivät ole tarpeeksi asiantuntevia rooliin.

Algoritminen ja koneoppimisen vinoumat

Esimerkkejä vinoumien tyypeistä koneoppimisen data-aineistossa

Piilevä vinouma

Syntyy, kun tehdään oletuksia henkilökohtaisten mentaalimallien ja kokemusten pohjalta, joita ei voida yleistää.

Esimerkki | Insinööri kouluttaa eleentunnistusmallia, jossa pään ravistaminen tarkoittaa ”ei”. Joillakin alueilla pään ravistus voi kuitenkin tarkoittaa ”kyllä”.

- Yleinen implisiittinen vinouma on **vahvistusvinouma**, jossa mallin rakentajat käsittelevät dataa tiedostamatta siten, että se vahvistaa heidän aiemmat uskomuksensa.
- Joskus mallin rakentajat jatkavat kouluttamista, kunnes tulokset vastaavat alkuperäistä hypoteesia – tätä kutsutaan **tutkijan vinoumaksi**.
- **Esimerkki** | Insinööri, joka rakentaa mallia koirien aggressiivisuudesta, yhdistää hyperaktiivisuuden lapsuuden kokemukseen lelupuudelista. Vaikka malli aluksi näyttää, että pienet puudelit ovat rauhallisia, insinööri kouluttaa mallia uudestaan, kunnes se näyttää, että pienet puudelit ovat aggressiivisempia – vahvistaen näin alkuperäisen oletuksen.

Algoritminen ja koneoppimisen vinoumat

Esimerkkejä vinoumien tyypeistä koneoppimisen data-aineistossa

Automaatiovinouma

Mieltymys automatisoidun järjestelmän tuloksiin ei-automaattisten järjestelmien sijasta, riippumatta virhetasosta.

Esimerkki

Ohjelmistoinsinöörit vaihdevalmistajalla esittelivät innostuneina uuden "innovatiivisen" mallin, jonka oli määrä tunnistaa vialliset vaihteet. Innostus kuitenkin laantui, kun tehtaan johtaja huomautti, että mallin tarkkuus- ja palautusasteet olivat 15 % huonommat kuin ihmistarkastajilla.

Algoritminen ja koneoppimisen vinoumat

Esimerkkejä vinoumien tyypeistä koneoppimisen data-aineistossa

Valintavinouma

Syntyy, kun esimerkit datasetissä eivät ole valittu tavalla, joka edustaisi todellista jakaumaa. Otantavinouma voi ilmetä useilla tavoilla:

- **Peittovinouma:** data ei ole valittu edustavasti.

Esimerkki | Tekoälymalli koulutetaan suosittamaan ravintoloita asiakasarvostelujen perusteella, mutta se koulutetaan vain yhden suositun ravintolaketjun arvostelujen perusteella, jättäen muut ravintolat huomiotta. Näin ollen malli ei heijasta kaikkien asiakkaiden makumieltymyksiä.

- **Vastaamattomuusvinouma):** data muuttuu epäedustavaksi osallistumisen puutteiden vuoksi datankeruussa.

Esimerkki | Tekoälyyn perustuva elokuvasuositusjärjestelmä kerää palautetta yhden suoratoistoalustan käyttäjiltä, mutta ohittaa muiden alustojen käyttäjät. Näiden käyttäjien mieltymykset eivät näin ollen näy suosituksissa.

- **Otantavinouma:** satunnaistamista ei ole käytetty asianmukaisesti datan keruussa.

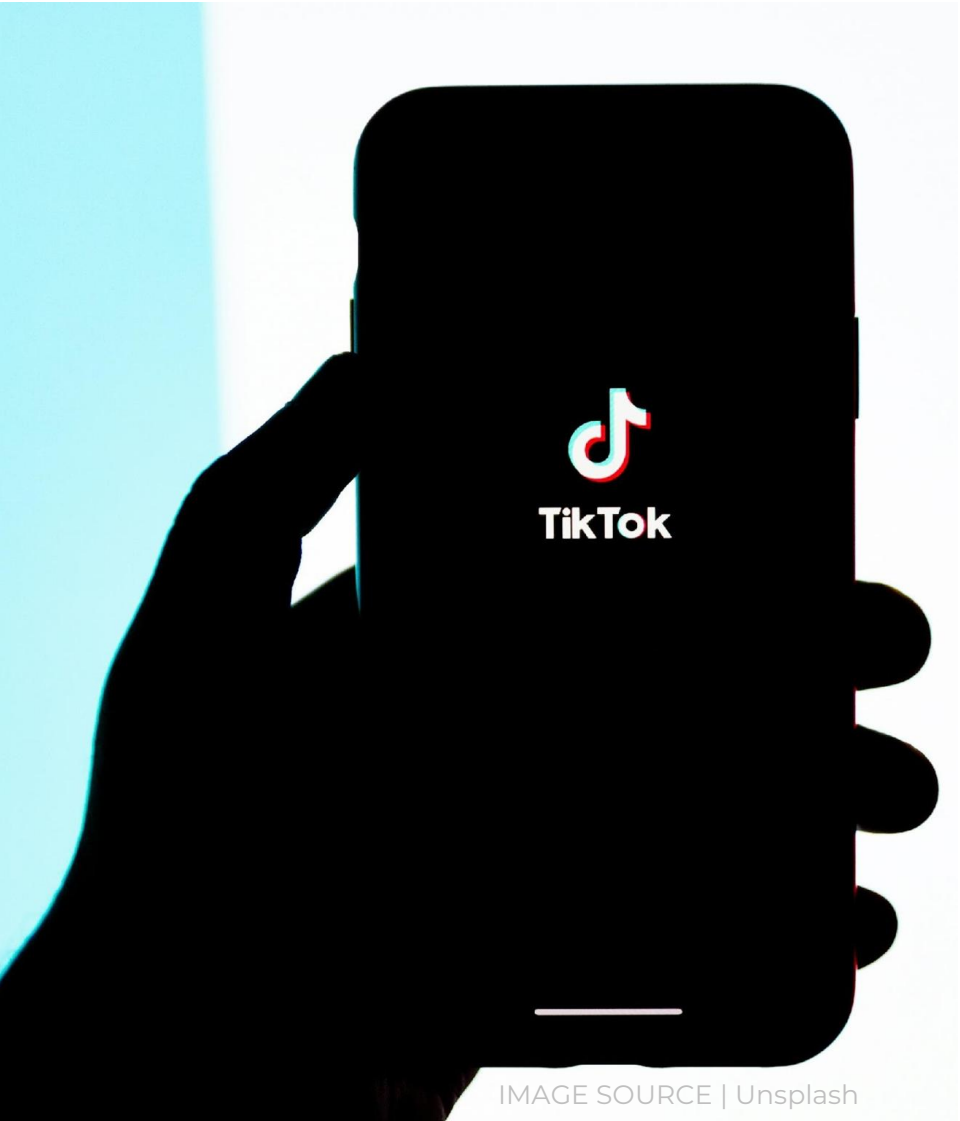
Esimerkki | Tekoälymalli koulutetaan ennustamaan uuden tuotteen tulevia myyntejä kyselemällä ostaneilta ihmisiltä ja niiltä, jotka ostivat vastaavan tuotteen, heidän ostotottumuksistaan. Sen sijaan että ihmiset olisi valittu satunnaisesti, kyselyn vastaajiksi otettiin ensimmäiset 200 henkilöä. Tämä voi tarkoittaa, että he olivat keskimääräistä innostuneempia tuotteesta kuin muut asiakkaat.

ALGORITMINEN VINOUMA JA NIIDEN VAIKUTUKSET - TAPAUSTEN ESIMERKIT



IMAGE SOURCE | Freepik

TAPAUS – TikTokin haitallinen algoritmi



Yle tutki, millaista sisältöä TikTokin algoritmi näyttää 13-vuotiaalle teini-ikäiselle tytölle, joka kärsii masennuksesta ja vääristyneestä kehonkuvasta.

Yle loi profiilin kuvitteelliselle hahmolle- Ellalle, joka edustaa masentunutta 13-vuotiasta nuorta. Ylen datatieteilijä selasi TikTokia Ellan profiililla hieman yli viiden tunnin ajan.

[Mitä tapahtui? Ella on Yle's article](#)

Yle, November 11, 2023

Kuinka algoritmi käyttäytyi koehenkilön käytössä?

1. Tunti – haitallista sisältöä 7%

(Videoita mielenterveysongelmista ja ruokaan liittyvistä asioista laskettiin haitalliseksi sisällöksi.)

- Aluksi – iloisia videoita ruoasta, koiranpennuista, kissoista, skeittauksesta ja vitseistä.
- Ella katsoo ruokavideoita hieman pidempään.
- Vain viiden minuutin kuluttua Ella pysähtyy katsomaan videota naisesta, joka huutaa.
- Hetkessä TikTok näyttää videon masennuslääkkeistä. Yhä enemmän videoita mielenterveydestä ja kevyistä aterioista.

2. tunti – haitallista sisältöä 28%

- Videoiden sävy muuttuu synkemmäksi.
- Satunnaista sisältöä hiustenhoidosta ja eläimistä..
- Videoissa alkaa esiintymään itsensä vahingoittamiseen liittyvää sisältöä.
- Videoiden joukossa on myös sisältöä, joka esimerkiksi kannustaa jatkamaan eteenpäin vaikeuksista huolimatta.

Yle, November 11, 2023

IMAGES SOURCE | Unsplash & Pexels

CASE – TikTok’s harmful algorithm Tulokset olivat hälyttäviä

“Eräessä videossa nainen, jolla on mielenter disorders, puhuu siitä, kuinka hän joskus tuntee halua riistää itseltään hengen.”

Kuinka algoritmi käyttäytyi koehenkilön käytössä?

3. tunti– haitallista sisältöä 68%

- **Juuri ennen kolmannen tunnin alkamista, Tiktok näyttää ensimmäisen kerran syömishäiriöihin liittyvää sisältöä.**
- Ensimmäisen videon jälkeen, yhä useampi video käsittelee syömishäiriöitä, jolloin Ella lopettaa niiden katsomisen.
- Enemmän sisältöä vähäkalorisista ruuista ja itsensä vahingoittamisesta.



4. tunti – haitallista sisältöä 57%

- Hieman vähemmän masennukseen ja itsensä vahingoittamiseen liittyvää sisältöä.
- Sen sijaan TikTokin algoritmi näyttää tunnistaneen, että Ella on erityisen kiinnostunut kevyistä aterioista ja haitallisesta sisällöstä, joka kannustaa rajoittamaan syömistä.
- Ruoka- ja laihdutusvideoita ilmestyy enemmän kuin aiemmin.



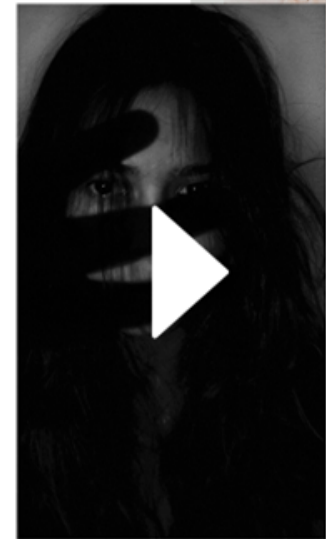
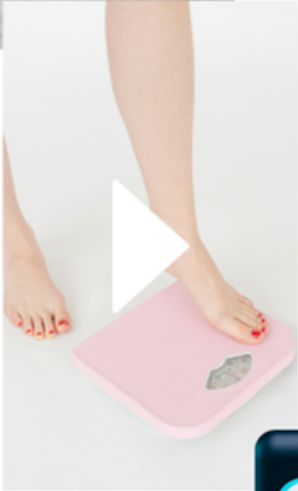
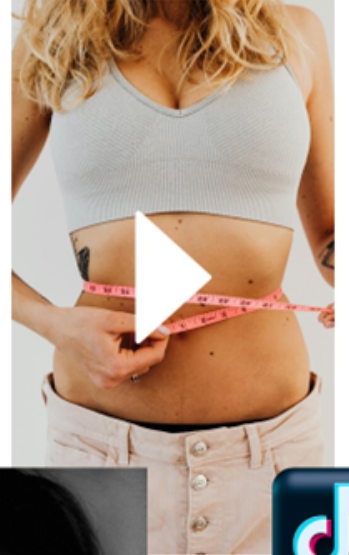
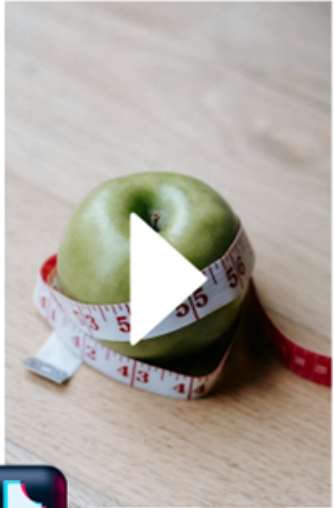
Kuinka algoritmi käyttäytyi koehenkilön käytössä?

5. tunti- haitallista sisältöä 65%

- Ellan näytölle ilmestyy yhä enemmän surullisia videoita.
- Videoita, jotka ihannoivat syömishäiriöitä ja antavat nyt suoraan konkreettisia neuvoja syömisen rajoittamiseen

Lopulta – haitallista sisältöä 95%

- Selailun lopussa TikTokin algoritmi oli näyttänyt Ellalle valtavan määrän videoita, jotka liittyivät masennukseen, itsetuhoisuuteen, vähäkalorisiin dieetteihin ja syömishäiriöihin.



Yle, November 11, 2023

IMAGES SOURCE | Unsplash & Pexels

TAPAUUS– TikTokin haitallinen algoritmi Tulokset olivat hälyttäviä

Sosiaalipsykologi Suvi Uskin mukaan algoritmi näyttää käyttäjälle mahdollisimman paljon sellaista sisältöä, joka häntä kiinnostaa, riippumatta siitä, kuinka haitallista se on. Uskin mukaan tämä on erityisen ongelmallista, koska TikTok on samalla ylivoimaisesti koukuttavin sosiaalisen median sovelluksista.

Eettiset ongelmat kasvojen tunnistuksessa

Kasvojentunnistus – mikä ja miten?

- Kasvontunnistusjärjestelmä skannaa ja tunnistaa yksilön kasvot vertaamalla niitä tallennettuun digitaaliseen kuvaan tai videopätkään tietokannassa käyttäen tekoälyalgoritmia.
- Teknologia toimii tunnistamalla kasvopiirteet kuvassa ja vertaamalla niitä muihin kuviin tai videoihin tietokannassa käyttäen tekoälyalgoritmia.

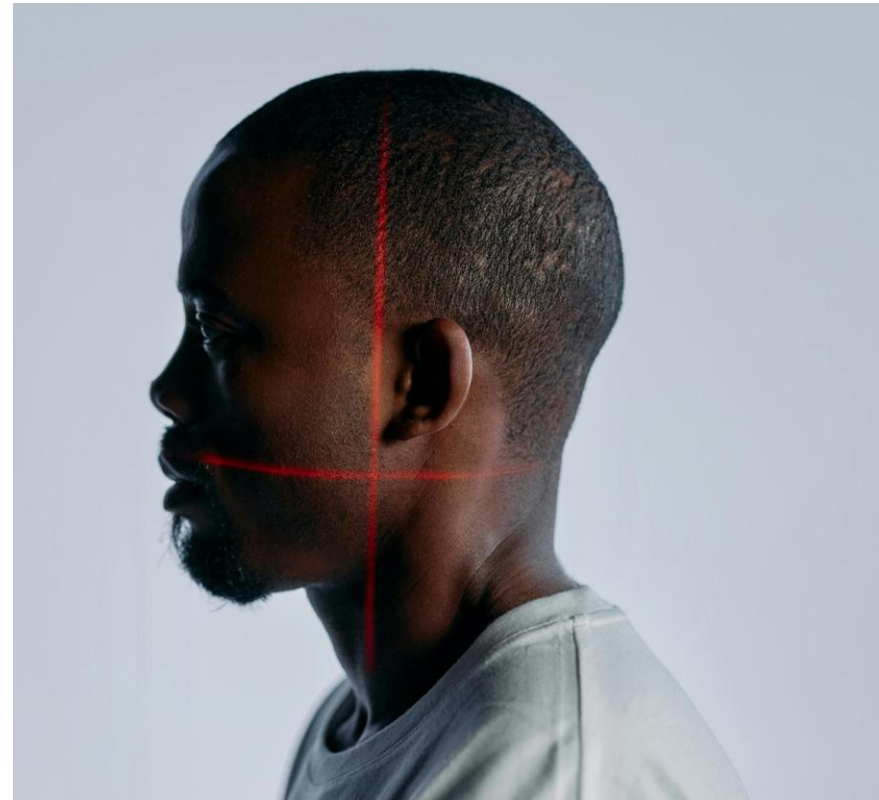


IMAGE SOURCE | Pexels

Shehmir Javaid, January 12, 2024)

Kasvojentunnistus – esimerkkejä sen käytöstä



- 1 **Terveystieteet.** Potilaan kasvoja skannataan automaattisesti kasvojentunnistuksella lääketieteellisen historian ja vakuutustietojen selvittämiseksi. Käytetään myös sairauksien diagnosoinnissa.
- 2 **Jälleenmyynti.** Amazon käyttää kasvojentunnistusta kosketusvapaisiin maksuihin. Voidaan käyttää ostokokemuksen personointiin, esimerkiksi asiakasuskollisuustarjouksissa. Voi myös parantaa jälleenmyynnin turvallisuutta.
- 3 **Pankki- ja rahoitusala.** Käytetään asiakkaan tunnistamiseen ja varmennukseen. Nopeuttamaan asiakasprosessia ja vähentämään ryöstöjä tai pankkipetoksia.
- 4 **Lainvalvonta ja turvallisuuspalvelut.** Rikollisten ja epäiltyjen tunnistamiseen ja kirjaamiseen. Pakenevien rikollisten tunnistamiseen kaupunkien valvontakameroiden avulla, esim. risteyksissä, juna- ja linja-autoasemilla sekä lentokentillä.

ALGORITMINEN VINOUMA JA NIIDEN VAIKUTUKSET - TAPAUSTEN ESIMERKIT

TAPAUS – Epäeettinen kasvojentunnistus datan keräyksessä



Video on facial recognition at airports on the washington post Youtube channel

- ▶ Katso video kasvojentunnistuksesta lentokentillä
- ? Mitä eettisiä ongelmia nousee ilmi videolla?

Washington Post, Jun 10, 2019

Epäeettinen kasvojentunnistus datan keräyksessä

“FBI:n ja ICE:n hallinnoimat ajokorttikuvat ovat kuin kultakaivos kasvojentunnistuksessa”

[Lue artikkeli: The Washington Post on July 7, 2019](#)

Liittovaltion poliisin (FBI) ja maahanmuutto- ja tulliviranomaisten (ICE) agentit ovat hyödyntäneet osavaltioiden ajokorttitietokantoja kasvojentunnistuksessa ja selanneet miljoonia amerikkalaisten kuvia ilman näiden tietoisuutta tai suostumusta.

Drew Harwell, The Washington Post, July 7, 2019

Rasistinen vinouma kasvojentunnistusjärjestelmissä

“Aasialaistaustaiset ja afroamerikkalaiset satakertaisesti useammin tunnistettiin väärin kuin länsimaalaisen henkilön kasvot, riippuen käytetystä algoritmista ja sanahausta”

[Lue artikkeli: Post on December 19, 2019](#)

Liittovaltioiden tutkimus paljasti kasvojentunnistusjärjestelmien rasistia vinoumia.

Drew Harwell, The Washington Post, December 19, 2019

TYÖKALUT ALGORITMISTEN VINOUMIEN TUNNISTAMISEKSI



IMAGE SOURCE | Freepik

Työkalut algoritmisten vinoumien tunnistamiseksi

UNESCO:n eettinen vaikutusten arviointi-työkalu

- UNESCO:n kehittämä 2023
 - Työkalu eettisen arvioinnin tekemiseen arvioimalla ja tunnistamalla tekoälyjärjestelmien hyödyt ja riskit.
 - Työkalu koostuu valmiista kysymyksistä, jotka voi täyttää.
 - Mukana sekä avoimia että monivalintakysymyksiä.
 - Eettisyys huomioitu suunnittelussa koko arvioinnin ajan.
- Arviointi alkaa rajauskysymyksillä ja etenee UNESCO:n periaatteisiin liittyviin kysymyksiin

Työkalu auttaa vastaamaan kysymyksiin, kuten:

- Kuka on vuorovaikutuksessa tekoälyratkaisun kanssa?
- Mitkä ovat tekoälytiimin roolit ja vastuut?
- Keitä sidosryhmiä tekoälyjärjestelmä tulee vaikuttamaan?
- Miten turvallisuus ja suojaus on huomioitu?
- Onko huomioitu kestävyys, yksityisyyden, läpinäkyvyyden, selitettävyyden ja vastuullisuuteen liittyviä kysymyksiä?

- [Katso täältä](#)

Figure 2: Positionality Matrix (developed by The Alan Turing Institute)



UNESCO, 2023

IMAGE SOURCE | UNESCO

Työkalut algoritmisten vinoumien tunnistamiseksi

UNESCO's Ethical Impact Assessment

Miten tunnistaa vinoumat?

- Osioittain oikeudenmukaisuus, syrjimättömyys ja monimuotoisuus keskittyy yksityiskohtiin, jotka esiintyvät vinoumissa.
- Vinoumia käsitellään myös huomiota rooleihin ja vastuisiin.

6.2. Procedural Assessment

Throughout this section, when responding to questions about testing with particular groups, the project team should consider especially – but not only – race, colour, descent, gender, age, language, religion, political opinion, national origin, ethnic origin, social origin, economic or social condition of birth, and disability. Please specify if testing was conducted on groups that combine several of these criteria i.e. if the system has been tested in terms of intersectionality.

6.2.1. Preventing discriminatory outcomes:

6.2.1.1. Has the algorithm been tested with different groups?

6.2.1.1.1. Was there a difference in terms of accuracy rate (or any other performance metric used)? Please describe any difference of this kind.

6.2.1.1.2. Was there a discriminatory effect for particular groups?

6.2.2. Data quality and preventing discriminatory bias:

6.2.2.1. Are processes in place to test data against biases?

6.2.2.1.1. Have you undertaken an analysis of the data to prevent societal and historical biases in data?

6.2.2.1.2. Is the data well-balanced and does it reflect the diversity of the targeted end-user population?

6.2.2.1.3. Are there any differences you can foresee between the data used for training and the data processed by the AI system which could result in the AI system producing discriminatory outcomes or performing differentially for different groups?

6.2.2.1.4. Have you developed a process to document how data quality issues can be resolved during the design process?

6.2.2.1.5. Did you put in place educational and awareness initiatives to help AI designers and developers gain awareness of the possible bias they can introduce through the design and development of the AI system?

6.2.2.1. Are processes in place to test data against biases?

Työkalut algoritmisten vinoumien tunnistamiseksi

ALTAI

- **Luotettavan tekoälyn arviointilista (ALTAI)**
- Kehittänyt Euroopan komission korkean tason asiantuntijaryhmä tekoälystä (AI HLEG)
- Lista seitsemästä vaatimuksesta oikeudenmukaiselle tekoälylle, jotka on määritelty tarkemmin verkkosivustolla.
- Sisältää suosituksia siitä, miten ALTAI:ta tulisi täyttää ja ketkä tulisi ottaa mukaan prosessiin.
- [Löydä täältä](#)

Sections of the ALTAI

- 📁 Human Agency and Oversight
- 📁 Technical Robustness and Safety
- 📁 Privacy and Data Governance
- 📁 Transparency
- 📁 Diversity, Non-Discrimination and Fairness
- 📁 Societal and Environmental Well-being
- 📁 Accountability

Is the **AI system** designed to interact, guide or take decisions by human **end-users** that affect humans ('**subjects**') or society? ⓘ *

- ☐ Yes
- ☐ To some extent
- ☐ No
- ☐ Don't know

Did you put in place procedures to avoid that **end-users** over-rely on the **AI system**? ⓘ *

- ☐ Yes
- ☐ No

Työkalut algoritmisten vinoumien tunnistamiseksi

ALTAI

Miten tunnistaa vinouma?

- ALTAI-työkalun osio "**Monimuotoisuus, syrjimättömyys ja oikeudenmukaisuus**" keskittyy epäoikeudenmukaisen vinouman välttämiseen.
- ALTAI-työkalu arvioi algoritmista vinoumaa kohdennetuilla kysymyksillä, tarkastelemalla syötedataa ja algoritmien suunnittelua mahdollisten vinoumien osalta sekä edistämällä osallisuutta, saavutettavuutta ja sidosryhmien osallistumista oikeudenmukaisuuden ja läpinäkyvyyden varmistamiseksi tekoälyjärjestelmissä.

Avoidance of unfair bias

Did you establish a strategy or a set of procedures to avoid creating or reinforcing **unfair bias** in the **AI system**, both regarding the use of input data as well as for the algorithm design?  *

☐ Yes

☐ No

Did you consider a mechanism to include the participation of the widest range of possible stakeholders in the **AI system's** design and development?

 *

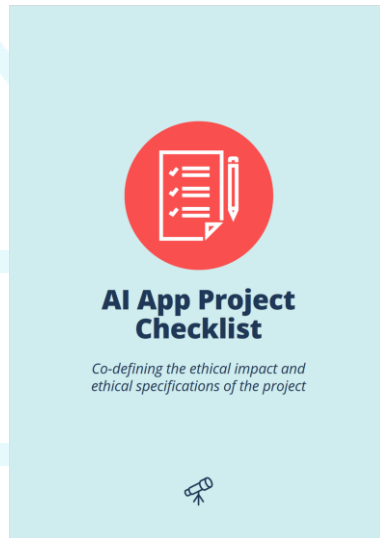
☐ Yes

☐ No

Työkalut algoritmisten vinoumien tunnistamiseksi

Eettinen työkalu tekoäly-sovellusten kehittämiseen

- **Kehittäjä:** Mario Sosa Hidalgo (TU Delft, teollinen muotoilutekniikka)
- Tarkastelee etiikkaa tekoälysovellusten kehittämisessä
- Innovatiivinen työkalupakki, jossa käytetään aktiivista, konkreettista ja visuaalista lähestymistapaa
- **Löydät täältä**

The 'AI APP PROJECT CHECKLIST' form is a structured document for project planning. It includes sections for: 1. Goal, 2. Team Accountable for App (with client information), 3. Context of the Project, 4. Stakeholders Impacted (Primary, Unexpected/Unintended), 5. Type of Impact (Financial, Property, Privacy, Emotional, Reputation, Access to goods, Life/Safety, Liberty/Freedom, Rights/IP), 6. AI Algorithm & Learning Model used, 7. Data Type (Human Data, Non-Human Data), 8. Data Source (GDPR compliant, Trustworthy), and 9. Ethical Principles Checklist (Data Privacy, Data Safety, Explainability, Transparency, Honest Communication, Accountability, Fairness, Value Alignment, Human Well-Being, Governance, User Safety).

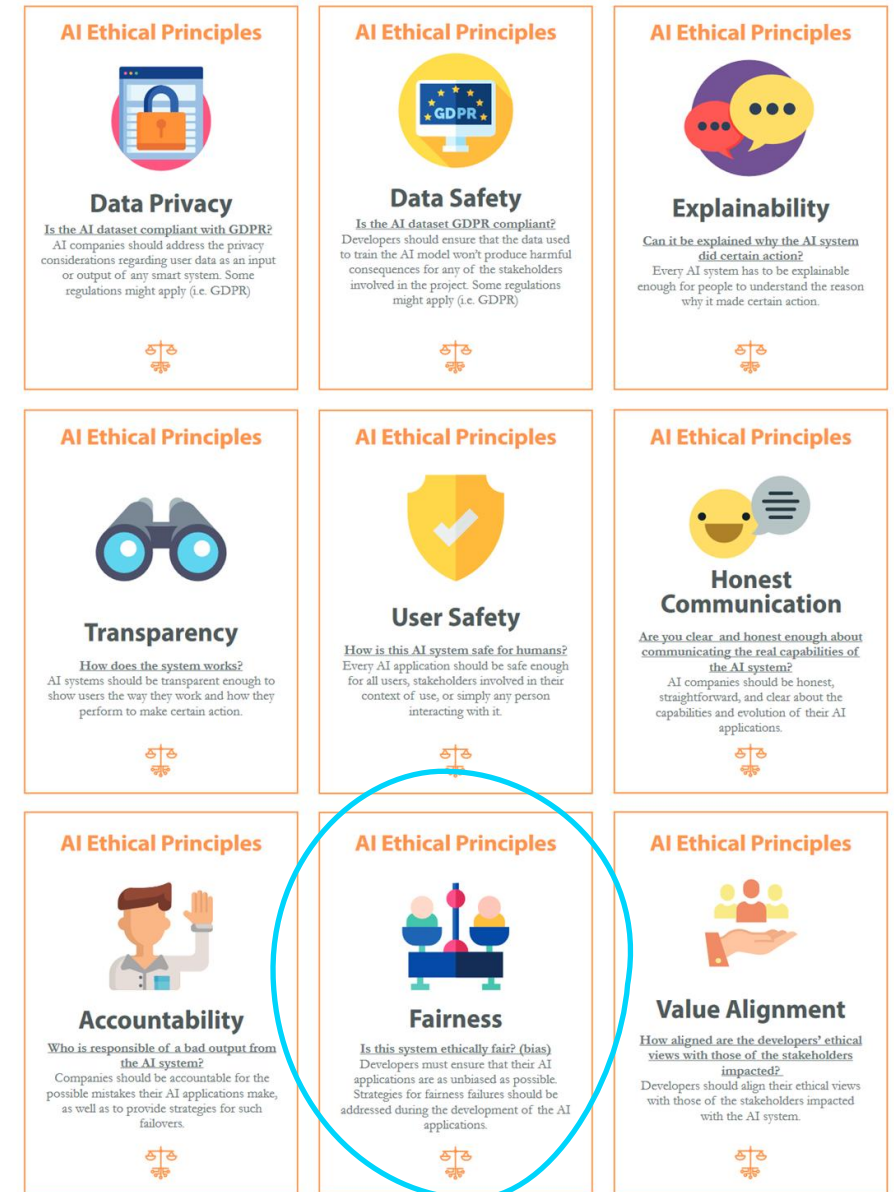
Mario Sosa Hidalgo, 2019

Työkalut algoritmisten vinoumien tunnistamiseksi

Eettinen työkalu tekoäly-sovellusten kehittämiseen

Miten tunnistaa vinouma?

Mario Sosa työkalu käsittelee vinoumia ja suosittelee kehittäjiä varmistamaan, että kehitetty tekoälyjärjestelmä on mahdollisimman puolueeton, muistuttaen oikeudenmukaisuuden epäonnistumisten ehkäisemisen strategioista.



Top icons by Flaticon.com

Mario Sosa Hidalgo, 2019

PROJEKTI | TEKOÄLY REKRYTOINNISSA



IMAGE SOURCE | Freepik

Projekti | Tekoäly rekrytoinnissa



ONGELMAN
MÄÄRITYS &
LIIKETOIMINNAN
YMMÄRTÄMINEN



Tapaustutkimus | Tausta

Kansainvälinen teollisuusyritys kohtaa tehottomuutta ja epävarmuutta rekrytointiprosessissaan, mikä kuormittaa merkittävästi HR-osastoa kuluttamalla aikaa ja resursseja.

Tulevaa kasvua ennakoidessaan yritys tunnistaa myös kiireellisen tarpeen palkata erilaisia asiantuntijoita, kuten insinöörejä, markkinoijia, talousasiantuntijoita, koodareita ja projektipäälliköitä.

Kova kilpailu pätevistä työntekijöistä ja rekrytointivirheiden korkeat kustannukset korostavat onnistuneen rekrytoinnin tärkeyttä. Tilanne lisää painetta HR-tiimille, jonka tehtävänä on varmistaa rekrytointiprosessin onnistuminen ja sujuvuus.

Tämän seurauksena yritys harkitsee tekoälyratkaisun kehittämistä, joka on räätälöity sen rekrytointitarpeisiin.

Tekoälyn integroinnin päätavoitteena rekrytointiprosessiin on parantaa tehokkuutta, objektiivisuutta ja rekrytoinnin yleistä onnistumisastetta.

Ensimmäisenä tarkastellaan tekoälypohjaisen rekrytointiratkaisun etuja ja haittoja:

Projekti | Tekoäly rekrytoinnissa

AI recruitment | PROS and CONS

EDUT

- **Tehokkuus ja ajanhallinta:** Tekoäly sujuvoittaa tehtäviä, jolloin rekrytoijat voivat keskittyä strategiseen työhön ja rekrytointiprosessi nopeutuu.
- **Parantunut ehdokaslaatu:** Tekoäly analysoi tietoa ja valitsee parhaiten sopivat ehdokkaat, mikä vähentää vaihtuvuutta.
- **Parannettu hakijakokemus:** Tekoäly tarjoaa ajantasaisia päivityksiä ja parantaa viestintää.
- **Kustannusten vähentäminen:** Tekoäly vähentää manuaalista työtä rutiinitehtävissä, mikä pienentää rekrytointikustannuksia.
- **Lisääntynyt oikeudenmukaisuus:** Tekoäly vähentää ennakkoluuloja ja parantaa objektiivisuutta → monimuotoisuuden edistäminen.
- **Oivaltava sosiaalisen median analyysi:** Tekoäly arvioi sosiaalisen median sisältöä kulttuurisen yhteensopivuuden näkökulmasta.
- **Globaali rekrytointi:** Tekoäly ylittää kielimuurit ja mahdollistaa laajemman osaajajoukon tavoittamisen.

HAITAT

- **Vähentynyt henkilökohtainen vuorovaikutus:** Tekoäly voi tehdä rekrytoinnista mekaanisempaa, mikä saattaa karkottaa hakijoita.
- **Ennakkoluulojen riski:** Huonosti suunniteltu tekoäly tai vinoutunut data voi heikentää monimuotoisuutta ja oikeudenmukaisuutta sekä ylläpitää yhteiskunnallisia eriarvoisuuksia.
- **Tietoturva:** Henkilötietojen käsittely tekoälyn toimesta nostaa yksityisyyden suojan ja kyberhyökkäysten riskiä.
- **Korkeat investointikustannukset:** Tekoälyn käyttö rekrytoinnissa vaatii merkittäviä taloudellisia ja resurssipanostuksia.
- **Ihmisen arviointikyky:** Tekoälyyn luottaminen voi johtaa osaavien ehdokkaiden ohittamiseen datan puutteellisuuden vuoksi.
- **Lainsäädännön noudattaminen:** Rekrytointiin käytettävän tekoälyn on sopeuduttava muuttuvaan lainsäädäntöön oikeudenmukaisuuden säilyttämiseksi.
- **Kulttuurinen arviointi:** Tekoäly ei välttämättä kykene arvioimaan kulttuurista yhteensopivuutta tai tiimityöskentelytaitoja tehokkaasti.

(www.recruiter.com, 2023; Bursell & Roumbanis, 2024; Arivu Recruitment, 2023; Köchling & Wehner, 2020, Albaroudi et al., 2024, Sheard, 2022; Albassam, 2023).

Projekti | Tekoäly rekrytoinnissa



ONGELMAN MÄÄRITYS & LIIKETOIMINNAN YMMÄRTÄMINEN

Ongelman ymmärtäminen

Yritys aloittaa taustatutkimuksen kehittääkseen tekoälyratkaisun rekryointitarpeisiinsa. Se muodostaa sisäisen tiimin, jonka tehtävänä on tunnistaa keskeiset haasteet nykyisessä rekryointiprosessissa. Tiimi haastattelee useita sidosryhmiä, kuten HR-osastoa, hiljattain palkattuja työntekijöitä ja esimiehiä, jotka ovat vastikään ottaneet uusia työntekijöitä tiimeihinsä. Lisäksi tiimi kerää tietoa rekrytoinnin nykyisistä kustannuksista ja analysoi näiden toimenpiteiden onnistumisprosentteja.

Mahdollisten työntekijöiden näkemyksen ymmärtäminen tekoälyn käytöstä rekrytoinnissa on yritykselle ratkaisevan tärkeää. Tavoitteena on välttää haitta maineelle tai brändille. Tämän saavuttamiseksi yritys on sitoutunut siihen, että tekoälyn käyttöönotto saa laajaa hyväksyntää ja että se tapahtuu eettisten ja laillisten normien mukaisesti.



Projekti | Tekoäly rekrytoinnissa



ONGELMAN MÄÄRITYS & LIIKETOIMINNAN YMMÄRTÄMINEN

Käyttäjien ymmärtäminen

Yksi haastatelluista on päärekrytoija Maria, jolla on usean vuoden kokemus yrityksestä. Hän on nähnyt läheltä, kuinka rekrytointitehtävät kasaantuvat ja aiheuttavat kasvavaa painetta HR-osastolle – erityisesti, kun kyseessä on kansainvälisten osaajien rekrytointi.

Maria suhtautuu avoimesti tekoälyn mahdollisuuksiin helpottaa rekrytointia, vaikka hän ei olekaan syvällisesti perehtynyt tekoälytekнологiaan



TYÖKALU

Persoonat ovat yksityiskohtaisia käyttäjäprofiileja suunnittelussa, jotka kuvaavat kohderyhmien keskeisiä piirteitä ilman stereotypioita. Niihin sisältyy käyttäytymistä, tarpeita ja demografisia tietoja, jotka edistävät empatiaa ja auttavat suunnittelutiimejä kehittämään käyttäjälähtöisiä ratkaisuja. Personat auttavat ylittämään perinteiset markkinasegmentit kohti osallistavampaa suunnittelua. (Kumar, 2013; Stickdorn ym., 2018a, 2018b)



Projekti | Tekoäly rekrytoinnissa

Maria – Päärekrytoija

#motivationdetermines
#goodtalents



KUVAN LÄHDE | Image made by VAMK

Empaattinen

Avoin

Menevä

Tiimipelaaja

Perfektionisti

Henkilötiedot

Ikä:27

Sijainti: Helsinki, Suomi.

Kansallisuus: Guetemala

Koulutus: Henkilöstöhallinta

Työ:päärekrytoija

Perhe: Asuu yksin kissansa kanssa

TYÖN TAVOITTEET

Etsiä uusia osaajia

Työntekijöiden tyytyväisyys

Nopea rekrytointiprosessi

HARRASTUKSET



Uinti

Kiipeily

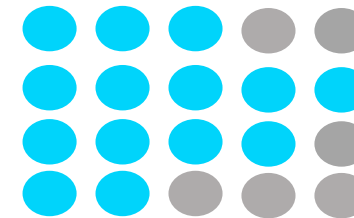
Lukeminen

HANKALUUDET

- ✓ Paljon töitä
- ✓ Uuden osaajan löytäminen, joka kuitenkin peruuttaa työpaikan juuri ennen aloittamista
- ✓ Liian hidas rekrytointiprosessi, joka saa hakijat menemään muualle
- ✓ Ei tarpeeksi aikaa kaikkien hakijoiden informoimiseksi
- ✓ Paine löytää parhaat kandidaatit

TAIDOT

Sosiaalinen media
Henkilöstöhallinnan
käyttöliittymät
Haastattelut
Benchmarking



Marialla on suuret odotukset, mutta hän on myös huolissaan tekoälypohjaisen rekrytoinnin vaikutuksista työhönsä. Häntä mietityttää, kuinka tekoälyn käyttöönotto saattaa muuttaa HR-henkilöstön tehtäviä tulevaisuudessa: Mitä uusia kyvykkyyksiä ja osaamisia hänen täytyy hallita? Arvostetaanko hänen panostaan jatkossa vähemmän? Tarvitaanko hänen asiantuntemustaan enää? Ja missä määrin rekrytointiprosessi voidaan automatisoida?

Projekti | Tekoäly rekrytoinnissa



ONGELMAN MÄÄRITYS & LIIKETOIMINNAN YMMÄRTÄMINEN



Rekrytoinnin vaiheiden ymmärtäminen

Saadakseen kokonaisvaltaisen käsityksen rekrytointiprosessista yritys kehittää rekrytointimatkan, jossa ovat mukana Maria ja rekrytointitiimi. Yhdessä linjaesimiesten kanssa he käyvät läpi prosessin eri vaiheet tunnistatakseen merkittävät turhautumisen lähteet ja aikaa vievät vaiheet. Samalla he tutkivat tekoälyn mahdollisuuksia parantaa rekrytointipolun tehokkuutta ja vaikuttavuutta kussakin vaiheessa.



TYÖKALU

Journey map eli asiakaskäsitekartta suunnittelutyökalu, joka hahmottaa käyttäjän palvelukokemuksen vaihe vaiheelta. Se sisältää vuorovaikutustilanteet, kontaktipisteet ja haasteet. Lisäksi se tallentaa tunteisiin liittyviä reaktioita kokemuksen laadun esiin tuomiseksi. Palvelupolut voivat kuvata joko nykyisiä tai tulevia vuorovaikutuksia ja vaihtelevat yleiskatsauksista yksityiskohtaisiin kuvauksiin. (Kumar, 2013; Stickdorn ym., 2018a)

Projekti | Tekoäly rekrytoinnissa

Marian rekrytoinnin vaiheet



TARPEEN TUNNISTAMINEN

Tarve uudelle työntekijälle tunnustetaan esihenkilön toimesta ja siitä ilmoitetaan Marialle.



OSAAVIEN TEKIJÖIDEN ETSINTÄ

Marian tehtävänä on tunnistaa ja houkutella uusia osaajia.



HAASTATTELUT

Maria ja esihenkilö haastattelevat sopivia ehdokkaita.



TULOTOIMENPITEET JA PEREHDYTTÄMINEN

Maria lähettää perehdytyspaketin työntekijälle, ja perehdytys jatkuu työntekijän saavuttua.



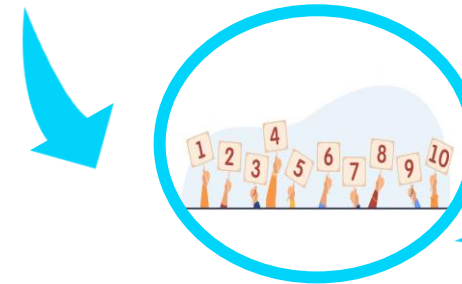
TYÖKUVAUKSEN LAATIMINEN

Työkuvaus laaditaan selventämään tarvittavaa osaamista.



HAKIJOIDEN SEULONTA JA VALINTA

Marian täytyy tunnistaa oikeat hakijat suuresta ehdokasjoukosta.



ARVIOINTI JA TARJOUS

Valinta ja taustojen tarkistus suosittelijoiden avulla. Työtarjous annetaan valitulle ehdokkaalle.

Projekti | Tekoäly rekrytoinnissa

Miten tekoälyä voidaan hyödyntää eri vaiheissa?



- **TARPEEN TUNNISTAMINEN**
- Ulkopuolisten tekijöiden, kuten markkinakysynnän, kilpailijoiden toimien ja HR-mittareiden analysointi
- Historiallisten työvoimatrendien, väestötietojen ja osaamisvaatimusten arviointi yhdessä sisäisten arviointien kanssa
- Rekrytoinnin tehokkuuden seuraaminen mittareilla, kuten vaihtuvuusaste ja täyttöaika



- **OSAAVIEN TEKIJÖIDEN ETSINTÄ**
- **Räätälöidyn viestinnän mahdollistaminen** digitaalisilla alustoilla ja sosiaalisessa mediassa
- Tekoälybotit tunnistavat sekä aktiivisia että passiivisia työnhakijoita eri alustoilta ja vanhoista hakijatietokannoista
- Tekstianalyysistrategioiden hyödyntäminen työpaikkailmoitusten houkuttelevuuden lisäämiseksi



- **TYÖKUVAUKSEN LAATIMINEN**
- Inhimillisen harhan vähentäminen työkuvausten laatimisessa
- Ilmoitusten kielen säätäminen ja sen vaikutusten seuraaminen hakemusmääriin ja -laatuun



- **HAKIJOIDEN SEULONTA JA VALINTA**
- Alkuvaiheen seulonta analysoimalla semantiikkaa ja ennustamalla pitkän aikavälin potentiaalia ja riskejä
- Videohaastattelut, joissa arvioidaan verbaalisia ja ei-verbaalisia vihjeitä persoonallisuuden ja kognitiivisten taitojen arvioimiseksi
- Sosiaalisen median analyysi kulttuurisen sopivuuden tarkastelemiseksi
- Ansioluetteloiden analyysi NLP-tekniikoilla hakemusten arvioinnin rikastamiseksi

Recruiter.com; Chen, 2022; Hunkenschroer & Luetge, 2022, Lawton ; Albouradi 2024; Marr 2023; Fritts & Carbera,2021; Vivek, 2023

Projekti | Tekoäly rekrytoinnissa

Miten tekoälyä voidaan hyödyntää eri vaiheissa?



- **HAASTATTELUT**

- Tekoälypohjaiset haastattelumenetelmät käytössä lähinnä esiseulontavaiheessa



- **ARVIOINTI JA TARJOUS**

- **Ehdokkaan aikaisempien työtehtävien** analysointi yksilöllisen tarjouksen luomiseksi, jonka ehdokas todennäköisesti hyväksyy



- **TULOTOIMENPITEET JA PEREHDYTTÄMINEN**

- Asiakirjojen luomisen automatisointi
- Perehdytysprosessin etenemisen seuranta
- Koulutuspolkujen räätälöinti

Recruiter.com; Chen, 2022; Hunkenschroer & Luetge, 2022, Lawton ; Albouradi 2024; Marr 2023; Fritts & Carbera,2021; Vivek, 2023

Projekti | Tekoäly rekrytoinnissa



DESIGN PHASE



Henkilökohtaisten vinoumien tunnistaminen

Selkeän ongelmanmäärittelyn jälkeen yritys on päättänyt edetä tekoälypohjaisen ratkaisun käyttöönotossa. Sinut on nimitetty projektin vastuulliseksi tekoälykehittäjäksi, jonka tehtävänä on johtaa valitun teknologian toteutusta.

Yritys tunnustaa monipuolisen, eri sidosryhmistä koostuvan tiimin kokoamisen tärkeyden, jotta projektiin saadaan erilaisia näkökulmia onnistumisen varmistamiseksi.

Mahdollisten tiedostamattomien ennakkoluulojen käsittelemiseksi yritys on määrännyt koko tiimin osallistumaan Alan Turingin mallin mukaiseen harjoitukseen, nimeltään asemamatriisi.



TYÖKALU

Alan Turingin asemamatriisi on suunniteltu tuomaan esiin tiimin jäsenten henkilökohtaisia ennakkokäsityksiä ja osoittamaan, kuinka sosioekonominen tausta, kulttuurinen konteksti ja yksilölliset elämäkokemukset voivat vaikuttaa arviointiin ja päätöksentekoon.

Projekti | Tekoäly rekrytoinnissa

Alan Turingin testi | Pohdi tarkoitusta, asemaa ja valtaa



Henkilökohtaiset ominaisuudet & ryhmäidentiteetti

Kuinka määrittelen itseni?

Ikä, rotu ja etninen tausta, vammaisuus, uskonto, sukupuoli, seksuaalinen suuntautuminen, siviilisääty, vanhemmuus, kielellinen tausta



Koulutus, koulutuspolku & työhistoria

Miten minut on koulutettu ja valmennettu?

Käydyt koulut ja oppilaitokset, koulutustaso, etenemismahdollisuudet ja ammatillisen kehittymisen mahdollisuudet, työhistoria



ASEMAMATRIISI

Missä määrin henkilökohtaiset ominaisuuteni, ryhmäidentiteettini, sosioekonominen asemani, koulutukseni, työhistoriani, tiimin kokoonpano ja institutionaalinen konteksti toimivat vallan ja etuoikeuden lähteinä — tai toisaalta syrjinnän ja epäedun lähteinä?

Miten tämä positio vaikuttaa minun (ja tiimini) kykyyn tunnistaa ja ymmärtää osallisia sidosryhmiä ja hankkeeni mahdollisia vaikutuksia?



Institutionaalinen konteksti ja tiimin kokoonpano

Miltä tiimini ja sen institutionaalinen kehys näyttävät? Valtasuhteet ja päätöksenteon rakenne projektitiimissä, organisaation omistajuus- ja valtahierarkiat, päätöksenteon autonomia, mahdollisuudet tuoda esiin huolia ja vastalauseita, tiimin monimuotoisuus, kulttuuri — osallistava vai poissulkeva?



Sosioekonominen asema

Mikä on sosioekonominen historiani?

Kasvuympäristön taloudellinen asema, yhteiskunnallinen liikkuvuus ajan myötä, nykyinen asema, tulevaisuuden sosioekonomiset tavoitteet

Projekti | Tekoäly rekrytoinnissa



SUUNNITTELU VAIHE

Sidosryhmäkartta

Yrityksen sisäisen sidosryhmäryhmän lisäksi, joka osallistuu projektiin, on tärkeää huomioida myös muut olennaiset sidosryhmät.

Tämän vuoksi yritys laati sidosryhmäkartan hyödyntäen G. J. Millerin määritelmää tekoälyn kehitysprosessin sidosryhmäkartoituksesta. Kartoituksessa määritellään kehityksen ja käytön aikaiset sidosryhmät sekä ulkoiset sidosryhmät.



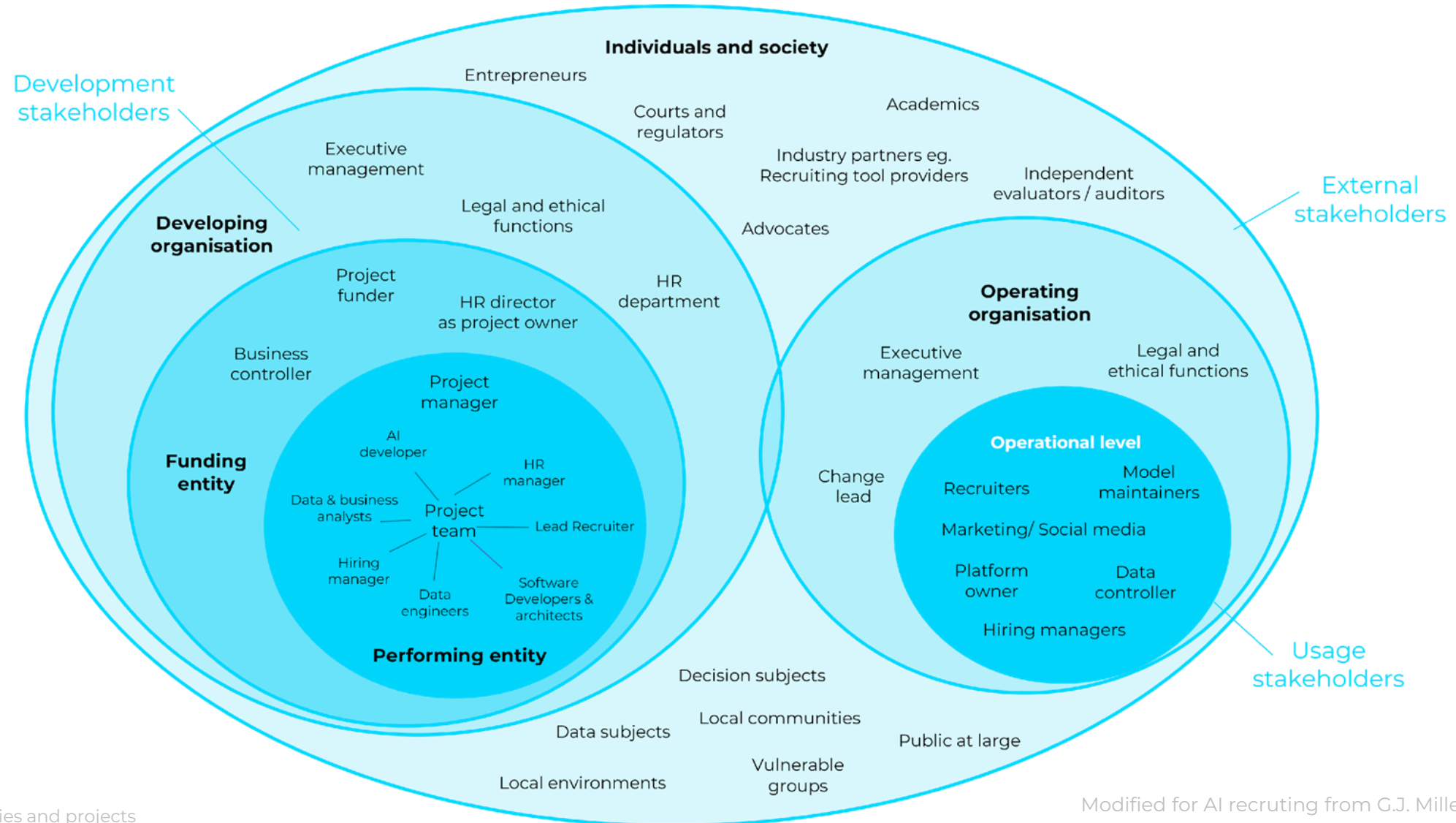
TYÖKALU

Sidosryhmäkartta on visuaalinen työkalu, jonka avulla tunnistetaan kaikki projektin osapuolet ja kuvataan heidän roolinsa ja suhteensa toisiinsa. Kartta voi olla yksinkertainen nelikenttä, jossa esitetään sidosryhmien vaikutusvalta ja osallistumisaste, tai se voi olla yksityiskohtaisempi matriisi, joka kuvaa sidosryhmien välisiä vuorovaikutuksia. Sidosryhmäkartta auttaa ymmärtämään sidosryhmädynamiikkaa ja on keskeinen verkostosuhteiden hallinnassa. (Giordano et al., 2018; SDT, n.d.-b)



Projekti | Tekoäly rekrytoinnissa

Sidosryhmäkartta: Tekoäly-avusteinen rekrytointiprojekti



Projekti | Tekoäly rekrytoinnissa



DESIGN PHASE

Sidosryhmän analyysimatriisi

Projektipäällikön tulee kartan pohjalta järjestellä sidosryhmät ryhmiin, jotta hän tietää, miten sidosryhmille tiedotetaan ja millainen vaikutus heillä on projektiin.



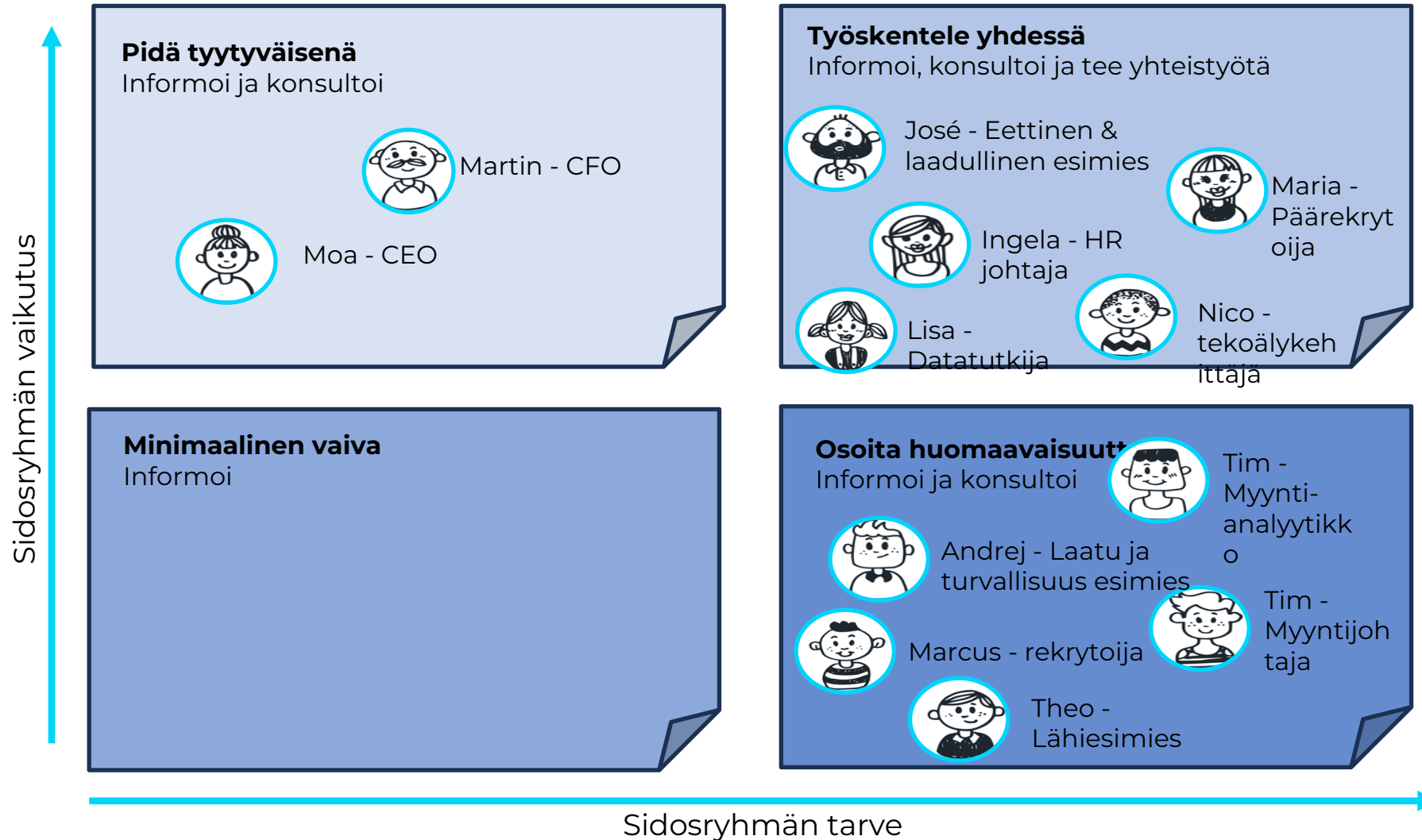
TYÖKALU

Työpaja on **yhteistoiminnallinen suunnittelutyökalu**, jossa osallistujat tekevät yhdessä vuorovaikutteisia tehtäviä ratkaistakseen tietyn ongelman tai edistääkseen projektia. Työpajoja ohjaa yleensä vetäjä, ja niiden kesto voi vaihdella muutamasta tunnista useisiin päiviin tavoitteista ja tehtävien monimutkaisuudesta riippuen. Työpaja on tehokas menetelmä ideoiden tuottamiseen, ongelmien ratkaisemiseen sekä tiimityön ja innovoinnin tukemiseen. (Wirtz, 2022)



Projekti | Tekoäly rekrytoinnissa

Sidosryhmän analyysimatriisi



Projekti | Tekoäly rekrytoinnissa



SUUNNITTELU VAIHE



Vinoumat rekrytoinnin eri vaiheissa - työpaja

Kun sidosryhmät ovat selvillä, yritys päättää järjestää eettisen työpajan, jossa käydään läpi mahdolliset ennakkoasenteet eri rekrytointivaiheissa.

Kentästä "Työskennellään yhdessä" (edellinen dia) kutsutaan sidosryhmiä osallistumaan työpajaan.

Millaisia post-it-lappuja olisit itse täyttänyt?

TYÖKALU

A **workshop** is a collaborative design tool where a group of participants engages in interactive activities to address a specific problem or work on a project. Typically facilitated by a leader, workshops can vary in duration from a few hours to several days, depending on the objectives and complexity of the tasks at hand. This method is effective for generating ideas, solving problems, and fostering teamwork and innovation within a group. (Wirtz, 2022)

Projekti | Tekoäly rekrytoinnissa



Vinoumat rekrytointitarpeen tunnistamisessa

ANKKUROINTIVINOUTUMA

Henkilö, jonka tilalle ollaan rekrytoimassa, vaikuttaa odotuksiin ja ennusteisiin rekrytointitarpeesta.

- HUONO suorittaja → Ennuste: tarvitaan lisää resursseja
- ERINOMAINEN suorittaja → Ennuste: tarvitaan vähemmän resursseja

PALAUTEVINOUTUMA

Tekoälyalgoritmi oppii omista aiemmista päätöksistään.

Jos rekrytointidatassa on vinoumaa, se voi ajan myötä voimistua tekoälyn alkuperäisten ennusteiden pohjalta.

STATUS QUO -VINOUTUMA

Organisaatiossa saattaa esiintyä vastustusta roolien muuttamiselle, vaikka liiketoiminnalliset tarpeet sitä edellyttäisivät.

DEMOGRAFINEN VINOUTUMA

Ennustavat mallit eivät välttämättä huomioi väestörakenteen muutoksia, kuten väestön ikääntymistä.

TOIMIALAVINOUTUMA

Liian kapea toimialanäkökulma voi estää laajempien tulevaisuuden tarpeiden tunnistamisen.

Projekti | Tekoäly rekrytoinnissa



Vinoumat työn kuvauksen laatimisessa

Tekoälyn luomat työpaikkailmoitukset tuotetaan usein vähentämään ihmisten **ennakkoluuloja rekrytoinnissa** → tekoäly voi tunnistaa kieltä, joka aiheuttaa vinoumia.

HISTORIAALLINEN VINOUTUMA

Työpaikkailmoitukset voidaan kouluttaa vanhojen ilmoitusten perusteella, jotka saattavat sisältää ennakkoluuloja.

DEMOGRAFINEN VINOUTUMA

Tekoäly voi ehdottaa termejä kuten "kunnianhimoinen" tai "itsevarma johtaja", jotka kohdentavat ilmoituksia tietyille väestöryhmälle.

ANKKUROINTIVINOUTUMA

Jos työpaikkailmoitukset luodaan huippusuorittajien ydinkyvykkyyksien perusteella, ne voivat luoda vinoumia.

Projekti | Tekoäly rekrytoinnissa



Vinoumat osaajien etsinnässä

SUKUPUOLI-, RASISMI- JA IKÄVINOUMA

Työnantajan valitsemat ennakkokriteerit ilmoitukseen voivat sulkea pois tiettyjä väestöryhmiä

Mainonnan kustannustehokkuus

voi johtaa siihen, että työpaikkailmoituksia näytetään niille, jotka todennäköisimmin klikkaavat ilmoitusta — sen sijaan että niitä näytettäisiin niille, jotka voisivat menestyä roolissa.

Työnantajan valitsema julkaisualusta

voi olla vinoutunut, sillä jotkin alustat ovat suositumpia tiettyjen hakijaryhmien keskuudessa.

Sosiaalisen median käyttö vaihtelee

suuresti: Jotkut eivät käytä esimerkiksi LinkedIniä lainkaan, toiset voivat antaa siellä väärän kuvan itsestään. Joillakin on enemmän mahdollisuuksia vaikuttaa digitaaliseen jalanjälkeensä

Käyttäytyminen vahvistaa

algoritmien toimintaa: esimerkiksi jos tietyn tyyppiset valkoiset miehet avaavat ilmoituksia useammin, ilmoitukset näkyvät yhä useammin samanlaisille käyttäjille.

Jos rekrytoijat viestivät enemmän tiettyjen alustojen kautta

tulevien hakijoiden kanssa, algoritmit voivat lisätä näkyvyyttä näillä alustoilla, mikä kasvattaa vinoumaa entisestään.

Sheard, 2022; Naakka 2018; Lawton, 2022; Chen 2022; Köchling & Wehner, 2022

Projekti | Tekoäly rekrytoinnissa



Vinoumat seulonnassa ja ehdokkaiden esivalinnassa

Yleisesti ottaen tekoäly voi luoda oikeudenmukaisemman rekrytointiympäristön vähentämällä tiedostamattomia ennakkoluuloja, joita ihmisillä on usein demografisten tekijöiden perusteella.

Jos koulutusdata perustuu nykyisten työntekijöiden joukkoon tai kapeasti määriteltyyn ryhmään, se voi johtaa tahattomaan syrjintään aliedustettuja ryhmiä kohtaan.

Yleistämisvinouma (aggregation bias) voi ilmetä, jos tehdään virheellisiä yleistyksiä koko väestöstä, esim. "nuorilla on teknisiä taitoja" → jättää vanhemmat ihmiset ulkopuolelle.

Digitaalisen jalanjäljen käyttäminen osana hakijoiden seulontaa voi olla ongelmallista: jos algoritmi perustuu esimerkiksi mieltymyksiin kuten "energinen" tai "aktiivinen", se voi syrjiä vammaisia henkilöitä.

Projekti | Tekoäly rekrytinnissa



Vinoumat haastattelussa

Tekoälytekniikoita käytetään haastatteluissa pääasiassa esiseulontavaiheessa.

Lopulliset haastattelut tulisi toteuttaa ihmisten välisessä vuorovaikutuksessa.

Videoarvioinneissa tunteentunnistusohjelmit voivat tulkita esimerkiksi eri kielten intonaatioita eri tavoin.

Jotkut ehdokkaat eivät välttämättä tunne itseään luontevaksi kameran edessä

Ihmiskeskeisyyden puute?
→ Saattaa heikentää työntekijän ja työnantajan välistä suhdetta.

Kasvojentunnistusteknologian käyttö voi herättää yksityisyydensuojaan liittyviä huolia sekä aiheuttaa vinoumia tiettyjä ryhmiä kohtaan.*

*Katso esimerkki kasvojentunnistuksesta

Sheard, 2022; Naakka 2018; Lawton, 2022; Chen 2022; Köchling & Wehner, 2022

Projekti | Tekoäly rekrytoinnissa

Tarjouksen tekeminen & perehdytys



Huomioitavaa: Milloin ihmisten välinen vuorovaikutus on välttämätöntä?



Perehdytys antaa kuvan yrityksestä. Koulutusdata voi perustua historiallisiin aineistoihin ja kantaa mukanaan vinoumia.

Perehdytys
Tekoälyalgoritmin kehittäjän **henkilökohtaiset ennakkoluulot** voivat vaikuttaa koulutustarjontaan.

Tarjottavat työkalut käyttävät usein historiallista dataa alan standardien ennustamiseen → tämä voi johtaa vinoumiin.

Projekti | Tekoäly rekrytoinnissa



DATAN KERÄÄMINEN JA VALMISTELU



Ehdokkaiden seulonta ja valinta: data- aineiston määrittäminen



Sillä aikaa kun yritys tutkii tekoälypohjaisen rekrytoinnin kaikkia osa-alueita ja mahdollisuuksia, se keskittyy tällä hetkellä erityisesti tekoälyn hyödyntämiseen seulonta- ja esivalintavaiheessa. Tähän sisältyy ansioluetteloiden syvälinen analyysi ja ehdokkaiden sovittaminen sopiviin työpaikkoihin.

Ottamalla käyttöön tekoälytyökaluja ansioluetteloiden automaattiseen seulontaan, tavoitteena on tunnistaa nopeasti ja tarkasti ne ehdokkaat, jotka täyttävät tehtävän vaatimukset. Tällaisen teknologian käyttöönoton odotetaan johtavan HR-resurssien strategisempaan käyttöön, parantavan lopullisten ehdokkaiden laatua sekä tehostavan rekrytointiprosessin onnistumista kohdistamalla ehdokkaiden taidot ja taustat paremmin yrityksen tarpeisiin.

Ensimmäiseksi yrityksen on määriteltävä ne tietoaaineistot, jotka muodostavat rekrytointijärjestelmän perustan ja joista algoritmit ottavat datan.

Projekti | Tekoäly rekrytoinnissa

Tekoälyn kouluttaminen

- Koulutusdatalla (training datasets) on keskeinen rooli algoritmien kehityksessä. Nämä aineistot toimivat algoritmin "totuuspohjana", jonka perusteella se oppii tulkitsemaan ja käsittelemään suuria määriä dataa.
- Ne opettavat algoritmilte yhteyksiä syötedatan (input variables) ja lopputuloksen (output variable) välillä – eli miten syötteet vaikuttavat järjestelmän tuottamiin päätöksiin tai ennusteisiin.
- Merkittävä haaste syntyy, jos koulutusdata sisältää vinoumia. Nämä voivat heijastaa historiallisia eriarvoisuuksia, ennakkoluuloja tai olla seurausta edustamattomasta tai puutteellisesta datankeruusta.
- Kun algoritmi oppii vinoutuneesta datasta, se toistaa nämä vinoumat päätöksenteossaan – ilmiö tunnetaan yleisesti nimellä "bias in, bias out". Tämä tarkoittaa, että jos syötetty data on vinoutunutta, myös lopputulos on vinoutunut. Se voi johtaa epäoikeudenmukaisiin, syrjiviin tai virheellisiin päätöksiin siitä, ketkä ehdokkaat valitaan tai hylätään – ei heidän todellisten kykyjensä, vaan vääristyneiden kriteerien perusteella.

Projekti | Tekoäly rekrytoinnissa

Yrityksen tekoälyn kouluttaminen sisäisellä ja ulkoisella datalla

Sisäinen data

- **Työntekijöiden suoritus- ja työhistoriadata:** Yrityksen sisäisesti keräämä data työntekijöiden suoriutumisesta ja aiemmista rekrytointipäätöksistä.
- **Hakemushistoria ja ansioluettelot:** Aiemmin yritykselle lähetetyt CV:t.
- **Väärin hylättyjen hakijoiden tiedot:** Rekrytointiprosessissa virheellisesti hylättyjen hakijoiden tiedot.
- **Hakemusten avainsanat ja -fraasit:** Yrityksen hakuprosessin kautta kerätyt tietyt sanat tai ilmaukset.
- **Koulutustaustatiedot:** Hakemuksista saadut koulutukseen liittyvät tiedot.

Ulkoinen big data

- **Sosiaalinen media:** Alustat, joissa ihmiset jakavat kiinnostuksen kohteitaan, kokemuksiaan ja ammatillisia saavutuksiaan. Näistä voidaan saada viitteitä hakijan persoonasta, taidoista ja verkostoista.
- **Ammatilliset verkkosivustot (esim. LinkedIn):** Profiilit, joissa on yksityiskohtaista tietoa työurasta, koulutuksesta, taidoista ja suosituksista.
- **Internet-haut:** Yritykset voivat etsiä hakijasta saatavilla olevaa julkista tietoa esim. uutisartikkeleista, blogeista tai julkaisuista.

Projekti | Tekoäly rekrytoinnissa



MALLIN KEHITYS JA KOULUTTAMINEN



Työmainonta; avainsanojen valinta algoritmile

Yritys on äskettäin julkaissut työpaikkailmoituksen ja saanut siihen satoja hakemuksia.

Tekoälyyn perustuva seulonta- ja esivalintaprototyyppi käy läpi hakijoiden CV:t tunnistukseen avainsanoja ja muuta rekrytoinnin onnistumiseen liittyvää tietoa, kuten: työkokemus, aiemmat työnantajat, tutkinnot ja oppilaitokset. Näiden perusteella järjestelmä luo rakenteistetun hakijaprofiilin ja antaa pisteytyksen ja sijoituksen kaikille ehdokkaille.

Job advertisement and related key words



WE ARE LOOKING FOR..

...Communications assistant to be responsible for the creation of content such as media releases, blogs, and social media posts on behalf of our company. You will also be monitoring media and campaign coverage and attending internal and external events.

RESPONSIBILITIES

- Develop and manage diverse content, including media releases, blogs, and social media posts, aligned with the company's strategic goals; monitor media and campaign coverage.
- Support the implementation of internal and external communications strategies and assist in managing the company's image.
- Organize marketing events and provide comprehensive administrative support, including maintaining event calendars and updating contact lists.
- Compile and prepare presentations and reports while tracking project progress and media exposure.

REQUIREMENTS

- Holds a Bachelor's degree in communications, marketing, or related field, with excellent verbal and written communication abilities.
- Proficient in social media strategies and media relations, with a creative and innovative approach.
- Strong organizational skills and attention to detail, capable of multitasking and maintaining positive interpersonal relationships.
- Skilled in using office management and design software, such as Photoshop and InDesign, and knowledgeable in various social media platforms.

"Viestinnän kandidaatin tutkinto"
"Markkinoinnin kandidaatin tutkinto"
"Sisällöntuotanto"
"Mediatiedotteet"
"Blogit"
"Sosiaalisen median julkaisut"
"Median seuranta"
"Kampanjoiden näkyvyyden seuranta"
"Viestintästrategioiden toteutuksen tukeminen"
"Yrityksen imagon hallinta"
"Markkinointitapahtumien järjestäminen"
"Laaja hallinnollinen tuki"
"Tapahtumakalentereiden ylläpito"
"Yhteystietolistojen päivittäminen"
"Esitysten kokoaminen"
"Raporttien kokoaminen"
"Projektin etenemisen seuranta"
"Medianäkyvyyden seuranta"
"Taitava sosiaalisen median strategioissa"
"Taitava mediasuhteissa"
"Luova ja innovatiivinen lähestymistapa"
"Vahvat organisointitaidot"
"Tarkkuus yksityiskohdissa"
"Kyky työskennellä usean tehtävän parissa yhtä aikaa"
"Positiivisten ihmissuhteiden ylläpito"
"Taitava toimistohallintaohjelmistojen käytössä"
"Taitava suunnitteluohjelmistojen käytössä"
"Photoshop"
"InDesign"
"Tuntee hyvin erilaiset sosiaalisen median alustat"

Projekti | Tekoäly rekrytoinnissa

Kahden hakijan ansioluetteloiden vertaaminen avainsanojen muotoilua kohtaan

MALLIN KEHITYS JA KOULUTTAMINEN



Seuraavaksi keskitytään kahden hakemuksen analysointiin, jotta ymmärretään, miten tekoälyprototyyppi arvioi ehdokkaita seulontavaiheessa. Tavoitteena on selvittää, esiintyykö valintaprosessissa vinoumia.

Yritys järjestää työpajan, jossa tutkitaan tekoälypohjaista CV-seulontaa yrityksen omasta rekrytointiprosessista otetun esimerkin avulla.

Työpajassa analysoidaan kaksi ansioluetteloa rinnakkain, jotta nähdään, miten sanavalinnat ja ilmaisut vaikuttavat siihen, jääkö päteviä ehdokkaita huomaamatta.

Tämän harjoituksen tarkoituksena on paljastaa mahdolliset vinoumat tekoälytyökaluissa ja kannustaa kriittiseen ajatteluun järjestelmän kehittämiseksi siten, että se tunnistaa ja arvostaa hakijoiden koko potentiaalin

CV1

CV2



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 lihavointi auttaa näkemään yhteydet tiettyjen ilmauksien ja avainsanapohjaisten tekoälysuodatus-työkalujen välillä.



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 taulukko havainnollistaa kahden CV:n sanavalintojen ja ilmaisutapojen erojen vaikutusta ja miten ne voivat vaikuttaa tekoälypohjaisen seulontaprosessin tulokseen. Vaikka toisella hakijalla on asiaankuuluvat pätevyudet, toinen CV saatetaan ohittaa, koska sen kuvaukset eivät vastaa tarkasti yrityksen seulontaan määrittelemiä avainsanoja. Tämä korostaa sitä, kuinka tärkeää on, että CV:n kieli vastaa tekoälyjärjestelmän odottamia avainsanoja

NÄKÖKULMA	JOHN DOE	JANE DOE
Työnimikkeet ja termit	Käyttää vakiintuneita termejä kuten "Viestintäassistentti"	Käyttää nimikkeitä kuten "Digitaalisen sisällön luoja"
Koulutustausta	Mainitsee selvästi "Markkinoinnin kandidaatin tutkinnon"	Ilmaisee "Kandidaatti viestintätieteissä"
Tehtävien ja taitojen kuvaus	Suora vastaavuus avainsanoihin kuten "Sisällöntuotanto", "Lehdistötiedotteet"	Käyttää vaihtelevaa kieltä kuten "sitouttavien tarinoiden luominen", "johtanut sisältöaloitteita"
Tekniset ja ohjelmistotaidot	Mainitsee "Photoshopin" ja "InDesignin"	Yleinen maininta "digitaalisista suunnittelutyökaluista ja sisällönhallintajärjestelmistä"
Suorat avainsana vastaavuudet	Vastaa tarkasti monia määriteltyjä avainsanoja	Käyttää erilaisia ilmauksia, jotka eivät välttämättä vastaa seulonnassa käytettyjä tarkkoja avainsanoja

Projekti | Tekoäly rekrytoinnissa



MODEL EVALUATION & TESTING



Oikeudenmukaisen ratkaisun testaus

Algoritmiprototyypin arviointi paljasti epäjohdonmukaisuuksia CV-analyysissä, mikä johti parannuksiin kattavamman ymmärryksen saamiseksi CV:n ominaisuuksista. Malli on nyt valmis testausvaiheeseen ennen todellista käyttöönottoa.

Vaikka useita osa-alueita vaatii arviointia, tässä keskitytään sovelluksen oikeudenmukaisuuden testaamiseen sen määrittämiseksi, kuinka laajasti prosessissa esiintyy mahdollisia vinoumia. Oikeudenmukaisuuden mittareita käytetään arvioimaan sovelluksen päätöksissä esiintyvän vinouman tasoa.



TYÖKALU

IBM:n avoimen lähdekoodin työkalupakki voi auttaa tutkimaan, raportoimaan ja vähentämään syrjintää ja vinoumaa koneoppimismalleissa koko tekoälysovelluksen elinkaaren ajan. Tutustu demoon ja kokeile työkalua itse.

[AI Fairness 360 \(ibm.com\)](https://ibm.com/ai/fairness360)

Projekti | Tekoäly rekrytoinnissa

Oikeudenmukaisuuden testaus

Järjestelmien luomiseksi, jotka ovat oikeudenmukaisia kaikille ja vapaita vinoumista kuten sukupuoli- tai rotusyrjinnästä, on tärkeää arvioida ratkaisun suorituskkyä eri väestöryhmien välillä.

Vaikka on olemassa erilaisia menetelmiä ratkaisun oikeudenmukaisuuden arviointiin, tässä rekrytoinnin yhteydessä keskitymme yksinomaan binääriluokittelutesteihin.

Oikeudenmukaisuuden määrittelemisen matemaattisin termein on osoittautunut haastavaksi, mikä on johtanut yhtenäisten standardimuotoilujen puutteeseen. Useimmat oikeudenmukaisuuden määritelmät kuitenkin keskittyvät useisiin keskeisiin käsitteisiin:

- Tietämättömyys
- Väestöllinen tasapaino
- Yhtäläiset todennäköisyydet
- Ennustettavuuden tasapaino
- Yksilön oikeudenmukaisuus
- Vastemuuttujan oikeudenmukaisuus

Väestöllinen tasapaino, ennustettavuuden tasapaino ja yhtäläiset todennäköisyydet kuuluvat yleensä "ryhmäoikeudenmukaisuuden" käsitteen alle. Tätä käsitettä tarkastellaan tässä osiossa erityisesti tekoälyn näkökulmasta rekrytoinnissa. Yksi binääriluokittelumenetelmä on sekaannusmatriisi (confusion matrix). Wortman Vaughan & Wallace, 2022; Zhong, 2018

Projekti | Tekoäly rekrytoinnissa

Sekaannusmatriisi

Sekaannusmatriisi tiivistää algoritmimallin tekemät ennusteet verrattuna todellisiin tuloksiin, joihin malli on koulutettu. Se näyttää sekä oikeiden että väärin ennusteiden lukumäärät, tarjoten selkeän kuvan mallin suorituskyvystä ja sen selitettävissä olevasta **tarkkuudesta**.

Sekaannusmatriisi rekrytoinnissa (esimerkki)

	Epäpätevä	Pätevä
Hylkäys	Tosi negatiivinen (True Negative, TN) = hylätyt, epäpäteviksi arvioidut hakijat	Väärä negatiivinen (False Negative, FN) = hylätyt, pätevät hakijat
Palkkaus	Väärä positiivinen (False Positive, FP) = palkatut, epäpätevät hakijat	Tosi positiivinen (True Positive, TP) = palkatut, pätevät hakijat

(Wortman Vaughan & Wallace, 2022; Zhong, 2018, Saplicki, 2022)

Projekti | Tekoäly rekrytoinnissa

Tarkkuus

Rekrytointiesimerkissä arvioimme oikeudenmukaisuutta käyttämällä sukupuolta demografisena muuttujana ja oletamme, että hakijoita on 100 naista ja 100 miestä.

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

NAISET	Epäpätevä	Pätevä
Reject	60 (TN)	20 (FN)
Hire	5 (FP)	15 (TP)

Tarkkuus

$$\frac{(TP + TN)}{(TP + FP + TN + FN)}$$

Molemmissa ryhmissä (naisissa ja miehissä) 75 hakijaa luokitellaan oikein
→ eli tarkkuus on molemmissa demografioissa samanlainen.

Projekti | Tekoäly rekrytoinnissa

Väestöllinen tasapaino

Sekaannusmatriisin tietojen pohjalta väestöllinen tasapaino on oikeudenmukaisuuskriteeri, jossa myönteisen lopputuloksen — tässä tapauksessa palkkaamisen — todennäköisyys ei riipu herkästä ominaisuudesta, kuten sukupuolesta.

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

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

Jos rekrytointiesimerkissä 80 sadasta mieshakijasta palkataan mutta vain 20 sadasta naishakijasta, väestöllinen tasapaino ei toteudu. Tämä voi kuitenkin olla hyväksyttävää, jos mieshakijat ovat aidosti pätevämpiä kuin naishakijat — kuten tässä tapauksessa näyttää olevan.



Ryhmätason oikeudenmukaisuutta tavoiteltaessa voi syntyä epäoikeudenmukaisuutta yksilötasolla, jos muuten pätevä hakija jätetään palkkaamatta vain väestöllisen tasapainon saavuttamiseksi.

Projekti | Tekoäly rekrytoinnissa

Ennustettavuuden tasapaino

Toisin kuin demografinen pariteetti, ennustettavuuden tasapaino (predictive parity) ottaa huomioon sekä luokittelijan päätökset että todelliset tulokset. Se arvioi, onko todennäköisyys, että hakija on pätevä, sama eri ryhmissä — edellyttäen, että tekoäly on valinnut heidät palkattaviksi. Tavoitteena on, että tämä todennäköisyys olisi likimain sama kaikissa ryhmissä.

MIEHET	Epäpätevä	Pätevä
Hylkäys	15	5
Palkkaus	20	60

NAISET	Epäpätevä	Pätevä
Hylkäys	60	20
Palkkaus	5	15

Rekrytointiesimerkissä järjestelmä täyttää ennustettavuuden tasapainon ehdot, sillä kolme neljäsosaa molemmista ryhmistä palkatuista hakijoista on päteviä.

Castelnovo et al., 2022; Wortman Vaughan & Wallace, 2022

Projekti | Tekoäly rekrytinnissa

Yhtäläiset todennäköisyydet; Väärien positiivisten arvo & väärien negatiivisten arvo

Väärän positiivisen määrän suhde (false positive rate) tarkoittaa, että todennäköisyyden palkata virheellisesti epäpätevä hakija tulisi olla sama kaikissa väestöryhmissä.

MIEHET	Epäpätevä	Pätevä
Hylkäys	15	5
Palkkaus	20	60

NAISET	Epäpätevä	Pätevä
Hylkäys	60	20
Palkkaus	5	15

Väärien negatiivisten suhde (**false negative rate**) tarkoittaa, että todennäköisyyden palkata virheellisesti pätevä hakija tulisi olla sama kaikissa väestöryhmissä.

MIEHET	Epäpätevä	Pätevä
Hylkäys	15	5
Palkkaus	20	60

NAISET	Epäpätevä	Pätevä
Hylkäys	60	20
Palkkaus	5	15

Yhtäläiset todennäköisyydet (equalized odds) yhdistävät nämä kaksi ominaisuutta täyttämällä sekä väärien negatiivisten että väärien positiivisten tasapainon. Rekrytointiesimerkissä ratkaisu **ei täytä tätä ehtoa**, koska epäpäteviä mieshakijoita palkataan todennäköisemmin kuin epäpäteviä naishakijoita JA päteviä naishakijoita hylätään todennäköisemmin kuin mieshakijoita.

Projekti | Tekoäly rekrytoinnissa

Harkintoja oikeudenmukaisuusmittareista

- Järjestelmä ei voi samanaikaisesti saavuttaa sekä ennustettavuuden tasapainoa, väärien positiivisten että väärien negatiivisten tasapainoa matemaattisten rajoitteiden vuoksi.
- Eri oikeudenmukaisuustavoitteiden välillä on tehtävä kompromisseja, ja niitä on hallittava harkiten.
- Oikeudenmukaisuuden arviointi on aina sidoksissa sovelluskontekstiin.
- Kaikkia oikeudenmukaisuuden ulottuvuuksia ei voida kvantifioida — on tunnustettava oikeudenmukaisuuden sosiotekninen monimutkaisuus.

Projekti | Tekoäly rekrytoinnissa

Yksilön oikeudenmukaisuus

A



Kandidaatin tutkinto
Vuosi työkokemusta

B



Maisterin tutkinto
Vuosi työkokemusta

C



Maisterin tutkinto
Ei työkokemusta

- Yksilöllinen oikeudenmukaisuus poikkeaa ryhmäperustaisesta oikeudenmukaisuudesta (kuten aiemmin käsitellyt mittarit), sillä se keskittyy yksilöiden kohteluun.
- Tämän käsitteen mukaan samankaltaiset yksilöt tulisi kohdata samankaltaisesti. Yksilöiden samankaltaisuuden mittaaminen on kuitenkin monimutkaista.
- Pohdi kolmea työnhakijaa — onko A pätevämpi kuin C?

Zhong, 2018

Projekti | Tekoäly rekrytoinnissa



MALLIN KÄYTTÖÖNOTTO



Hakijoiden informointi tekoälyn käytöstä rekrytoinnissa

Seuraavien huolellisten sisäisten arviointien jälkeen yritys on valmis ottamaan seulontatyökalunsa käyttöön valikoidusti avoimissa tehtävissä kerätäkseen alustavaa palautetta käyttäjiltä ja hakijoilta.

On olennaista, että hakijat ovat jatkuvasti tietoisia siitä, että he ovat vuorovaikutuksessa tekoälypohjaisen rekrytointijärjestelmän kanssa – tämä vahvistaa luottamusta. Tietoisuutta tekoälytyökalujen eduista tulisi lisätä jo alkuvaiheessa, mikä voi lisätä hakijoiden halukkuutta käyttää vuorovaikutteista teknologiaa, kuten chatbotteja. Lisäksi tekoälyn käyttöön perustuvien rekrytointiprosessien hienosäätö ja selkeä viestintä järjestelmän toiminnasta ovat keskeisiä tekijöitä järjestelmän hyväksyttävyyden ja menestyksen kannalta hakijoiden keskuudessa.

On myös varmistettava, että järjestelmän toimintaa valvoo ihminen, joka voi puuttua tilanteeseen tarvittaessa.

Projekti | Tekoäly rekrytoinnissa

Hakijoiden informointi tekoälyn käytöstä rekrytoinnissa | Työkalujen esimerkit

UNESCO's Eettinen vaikutusten arviointi-työkalu

Esimerkki Unescon eettisten vaikutusten arvionti -työkalusta, joka esiteltiin aikaisemmin

10.2.1. System Awareness:

- 10.2.1.1. Are users made fully aware when they are interacting with an AI system, as opposed to a human being?
- 10.2.1.2. Are individuals (directly or indirectly) impacted by the AI system made fully aware of when a decision (that impacted them) was informed by or made on the basis of an AI system or AI algorithms?
 - 10.2.1.2.1. Are they made aware of the extent to which they are impacted, including the rationale, benefits and limitations of the decision(s)?
- 10.2.1.3. Have appropriate explanations been put in place to help users and other impacted individuals understand the decision-making process or how the system works when required?
- 10.2.1.4. Have appropriate explanations been put in place to help the government bodies in charge of regulation understand the decision-making process or how the system works when required?
- 10.2.1.5. Has the decision to adopt the AI system been documented and communicated online?
- 10.2.1.6. Can the AI system make any decisions which the physical persons or legal entities in charge of the system lack expertise or competence to critique, modify or override?

Eccola Cards; menetelmä eettisesti yhdenmukaisten tekoälyjärjestelmien toteuttamiseksi

Data

#7 Privacy and Data

Motivation: Privacy is a rising trend in the wake of various recent data misuse reveals. People are now increasingly conscious about handing out personal data. Similarly, regulations such as the GDPR now affect data collection.

What to Do: Ask yourself:

- What data are used by the system?
- Does the system use or collect personal data? Why? How is the personal data used?
- Do you clearly inform your (end-)users about any personal data collection? E.g., ask for consent, provide an opportunity to revoke it etc.
- Have you taken measures to enhance (end-user) privacy, such as encryption or anonymization?
- Who makes the decisions regarding data use and collection? Do you have organizational policies for it?

Practical Example: Rather than collecting and selling data, appealing to privacy can also be profitable. Regulations are making it increasingly difficult to collect lots of personal data for profit. Privacy can be an alternate selling point in today's climate.

ECCOLA

©ID-4601-20200415

Unesco, 2023; Vakkuri et al. 2021

Projekti | Tekoäly rekrytoinnissa



KÄYTTÖ JA VALVONTA

Käyttöönoton jälkeen

Vaikka oikeudenmukaisuutta arvioitiin sisäisesti ennen käyttöönottoa, yrityksen tulee olla valmis myös riippumattomiin ulkoisiin auditointeihin.

On eduksi hakea jatkuvaa palautetta käyttöönoton jälkeen ja luoda pysyviä vuorovaikutuskanavia sidosryhmille. Tämä tarkoittaa työntekijöiden jatkuvaa tiedottamista, osallistamista ja kuulemista tekoälyjärjestelmän käyttöönoton kaikissa vaiheissa organisaation sisällä.



KIITOS



Universitat
de les Illes Balears

ISQe
ENGAGING PEOPLE



helixconnect



Euroopan unionin
osarahoittama

Euroopan unionin rahoittama. Esitetyt näkemykset ja mielipiteet ovat ainoastaan tämän tekstin laatijoiden näkemyksiä eivätkä välttämättä vastaa Euroopan unionin tai Euroopan koulutuksen ja kulttuurin toimeenpanovirasto (EACEA) kantaa. Euroopan unioni ja EACEA eivät ole vastuussa niistä.

2022-1-ES01-KA220-HED-000085257