

Tiina Härkönen, Kirsi Hantula ja Marko Aalto

DEMOKRAATTINEN TEKOÄLY SUOMESSA

Osallistuvan tekoälykehityksen
toimintamalli

Sitran muistio

© Sitra 2026

Demokraattinen tekoäly Suomessa

Kirjoittajat: Tiina Härkönen ja Kirsi Hantula Sitrasta sekä Marko Aalto Reaktorista. Tiina Härkönen ja Kirsi Hantula ovat Sitran Demokratiainnovaatiot-ohjelman johtavia asiantuntijoita. Marko Aalto on teknologiakonsulttiyhtiö Reaktorin Vice President of Data and AI.

Työryhmä. Sitra: Tiina Härkönen, Kirsi Hantula, Tarmo Toikkanen, Antti Lehtinen. Reaktor: Marko Aalto, Anne Karumo, Akseli Kouvo. VTT: Anton Sigfrids. Digi- ja väestötietovirasto: Marko Latvanen. Veikko Isotalo (Turun yliopisto).

Kiitos: Reijo Aarnio, Sitra.

Taitto: Grano Oy

ISBN 978-952-347-456-7 (PDF) www.sitra.fi
ISSN 2737-1034 www.sitra.fi

Sitran muistiot ovat tulevaisuustyömme taustaksi tuotettuja sisältöjä.

Sisältö

Esipuhe	4
Tiivistelmä	5
Sammanfattning	7
Abstract	9
1. Johdanto: Epäluottamus tekoälyyn vaatii toimia	11
2. Tapausesimerkki: Mitä mieltä tekoälystä, Suomi? -keskustelu	15
3. Digitaalisen kansalaiskeskustelun käynnistäminen ja kutsuprosessi	17
4. Voxit-aineiston rajaaminen kansalaissääntöjen muodostamista varten	27
5. Kansalaissäännösten laatiminen	31
6. Suomen kokeilun kansalaissäännöt tekoälylle julkisten palveluiden kehittämiseen ja käyttöön	34
7. Johtopäätöksiä ja loppusanat	40
Lähteet	43
Liitteet	44

Esipuhe

Tekoälyllä on kasvava rooli yhteiskunnassa ja sen vaikutukset ulottuvat arkeen, työhön, palveluihin ja päätöksentekoon. Sitran Megatrendit 2026 -raportissa tekoälyn nähdään haastavan muun muassa koulutuksen ja päätöksenteon perinteisiä tapoja. Kehitys tarjoaa mahdollisuuksia ja herättää kysymyksiä siitä, miten tekoälyä voidaan käyttää vastuullisesti ja tarkoituksenmukaisesti. Oltiinpa tekoälystä mitä mieltä tahansa, sen mukanaan tuomalla muutoksella on merkittävä rooli tulevien vuosien kehityksessä.

Tässä muistiossa tarjoamme yhden tärkeän näkökulman tekoälykeskusteluun; sen muodostavat 6 500 ihmistä, jotka osallistuivat *Mitä mieltä tekoälystä, Suomi?* -kansalaiskeskusteluun. Keskustelun tavoitteena oli lisätä ymmärrystä siitä, millaisia näkemyksiä, odotuksia ja huolia ihmiset liittävät tekoälyyn julkisissa palveluissa. Tarkoituksena ei ole ollut tuottaa valmiita vastauksia tai suosituksia, vaan tuoda esiin kysymyksiä ja teemoja, joita tekoälyyn yhteiskunnallisena ilmiönä liitetään. Halusimme myös kokeilla, voidaanko joukkoistamalla muuntaa laaja-alainen arvopohjainen keskustelu suoraan kielimallien rakennusaineeksi.

Tekoälyn kehitys on nopeaa ja sen moninaiset vaikutukset voivat olla vaikeasti hahmotettavia. Samalla tekoälyyn liittyy monenlaisia mielikuvia ja tunteita, jotka eivät aina perustu omakohtaiseen kokemukseen. Kansalaisten suhtautumisen tarkastelu tarjoaa mahdollisuuden tunnistaa tekijöitä, jotka voivat vaikuttaa tekoälyn hyväksyttävyyteen ja luottamuksen rakentamiseen osana julkisten palvelujen kehittämistä.

Muistiossa esitetyt havainnot tarjoavat yhden näkymän siihen, millaista teemoista tekoälyn yhteydessä Suomessa keskustellaan, tai mistä tulisi keskustella, sekä siihen, millä tavoin näkemykset yhtenevät ja eroavat toisistaan. Ne eivät kuvaa yhtenäistä kansan mielipidettä, eivätkä sellaisenaan ohjaa päätöksentekoa, vaan luovat pohjaa jatkotarkastelulle ja keskustelulle.

Tämän työn tarkoituksena on kannustaa keskusteluun tekoälyn käytöstä. Tekoälyyn liittyvässä muutoksessa keskeistä on luoda tilaa kuulemiselle ja yhteiselle pohdinnalle, jotta teknologian kehittämisessä voidaan huomioida myös ihmisten kokemukset ja näkemykset.

Haluan kiittää projektissa kumppanimme toiminutta teknologiakonsultointiyritys Reaktoria hyvästä yhteistyöstä ja ratkaisevasta roolista uuden kokeilussa, jossa kansalaiskeskustelu haluttiin muuttaa tekoälyn säännöiksi. Erityisesti haluan kiittää jokaista kansalaiskeskusteluun osallistunutta. Ilman aktiivisuuttanne ei kokeilua olisi voitu tehdä.

Kalle Nieminen

Johtaja, Ratkaisut

Sitra

Tiivistelmä

Suomalaisessa yhteiskunnassa tekoäly on nopeasti yleistynyt osaksi arkea, mutta kansalaisten luottamus sen kehittäjiin ja käyttöön on heikkoa. Generatiivisen tekoälyn kehityksessä yksilöiden arkaluontoiset tiedot ovat merkittävä kehittämisen resurssi, minkä vuoksi kansalaisilla tulisi olla todellinen mahdollisuus vaikuttaa siihen, mihin suuntaan tekoälyä kehitetään ja miten sitä käytetään – erityisesti julkisella sektorilla, jossa päätökset vaikuttavat laajasti arkeen.

Luottamus tekoälyyn syntyy läpinäkyvyydestä, osallisuudesta ja kansalaisten tiedollisten oikeuksien tunnustamisesta. Suomessa perustuslaki linjaa, että kansalaisilla tulee olla mahdollisuus vaikuttaa itseään koskeviin päätöksiin. Tämän oikeuden tulisi ulottua myös tekoälyn kehittämiseen ja käyttöönottoon. Suomalaiset toivovat, että päätöksentekoon ja yhteisiin asioihin voisi vaikuttaa helposti ja nopeasti verkossa, mutta sosiaalisen median keskustelukulttuuri karkottaa suuren osan osallistujista.

Ilman läpinäkyvyyttä ja todellista toimijuutta tekoälyyn ei synny kestävää luottamusta. Julkisen hallinnon ja muiden yhteiskunnallisten instituutioiden tulisi nähdä kansalaisten tekoälytoimijuuden turvaaminen keskeisenä tehtävänä. Tiedollisten oikeuksien kehikko tiivistää kansalaisten oikeudet tekoälyn aikakaudella neljään: pääsyyn tietoon, osaamiseen ja valmiuksiin, osallistumiseen sekä oikeuteen muodostaa näkemyksiä ilman painostusta. Näiden oikeuksien vahvistaminen on edellytys luottamuksen ja legitimitetin säilymiselle.

Osallistuvan tekoälykehityksen toimintamalli ja tekninen kansalaisten tekoälysäännöstö ovat rohkean kokeilun kautta syntyneitä esimerkkejä siitä, miten kansalaisten näkemykset voidaan huomioida tekoälyn kehittämisessä ja päätöksenteossa.

Sitran ja Reaktorin järjestämä Mitä mieltä tekoälystä, Suomi? -kansalaiskeskustelu toteutettiin syyskuussa 2025 anonymilla, rakentavan keskustelun Voxit-alustalla ja siihen osallistui lähes 6500 henkilöä. Keskustelussa julkaistiin 194 tekoälyn kehittämistä ja käyttöä koskevaa väitettä, joista 169 oli kansalaisten laatimia.

Kansalaiskeskustelun tavoitteena oli pureutua luottamuksen haasteisiin kansalaisten näkökulmasta ja luoda säännöstö tekoälyn käytölle julkisissa palveluissa. Keskustelun avulla haluttiin lisätä tietoa tekoälystä, antaa kansalaisille toimijuutta vaikuttaa, sekä testata ja kehittää kansalaisten tekemiä sääntöjä. Tavoitteena oli luoda tekninen säännöstö kielimallin ohjaamiseksi sekä dokumentoida koko toimintamalli, niin, että se olisi hyödynnettävissä muuallakin Euroopassa.

Keskustelun avulla kerättiin laaja ja monipuolinen aineisto kansalaisten näkemyksistä tekoälyn käytöstä julkisissa palveluissa. Sen pohjalta laadittiin tekoälyn toimintaan vaikuttava tekninen säännöstö, jota esitellään tässä muistiossa. Näin syntyi arvokas perusta, joka tukee tekoälyn vastuullista ja demokraattista hyödyntämistä yhteiskunnassa, ja auttaa löytämään ratkaisuja yhteiskunnallisen luottamuksen turvaamiseksi.

Sammanfattning

Artificiell intelligens har snabbt blivit en del av vardagen i det finländska samhället, men medborgarna har lågt förtroende för dess utvecklare och användning. Känsliga personuppgifter är en viktig resurs i utvecklingen av generativ AI, och av den anledningen borde medborgarna ha möjlighet att påverka hur AI ska utvecklas och vad den ska användas till – särskilt inom den offentliga sektorn där beslut som fattas har en stor inverkan på medborgarnas vardag.

Förtroende för artificiell intelligens skapas genom öppenhet, delaktighet och respekt för medborgarnas informationsbaserade rättigheter. Enligt Finlands grundlag ska medborgarna ha möjlighet att påverka beslut som berör dem själva. Den rätten bör även gälla utvecklingen och användningen av artificiell intelligens. Finländarna vill enkelt och smidigt kunna påverka beslutsfattandet och gemensamma frågor på nätet, men den rådande diskussionskulturen på sociala medier gör att många väljer att hålla sig utanför.

Utan transparens och en verklig delaktighet i AI-utvecklingen skapas inget varaktigt förtroende. Det borde vara en central uppgift för den offentliga förvaltningen och andra samhällsinstitutioner att säkerställa att medborgarna aktivt deltar i frågor som rör AI. Ramverket för de informationsbaserade rättigheterna sammanfattar medborgarnas rättigheter i den artificiella intelligensens tidsålder i fyra punkter: tillgång till information, kompetens och färdigheter, delaktighet och rätten att bilda sig en åsikt utan påtryckningar. Att stärka de här rättigheterna är en förutsättning för att upprätthålla förtroende och legitimitet.

En modell för utveckling av AI som bygger på delaktighet och ett tekniskt AI-regelverk som utgår ifrån medborgarna är exempel, som tagits fram genom djärva försök, på hur medborgarnas åsikter kan beaktas i AI-utvecklingen och beslutsfattandet.

I september 2025 anordnade Sitra och Reaktor medborgardiskussionen *Vad tycker du om artificiell intelligens, Finland?* på en anonym Voxit-plattform utformad för konstruktiva diskussioner, där närmare 6 500 personer deltog. I diskussionen presenterades 194 påståenden om utvecklingen och användningen av AI, varav 169 hade lämnats in av medborgarna själva.

Målet med medborgardiskussionen var att fördjupa sig i förtroendefrågan utifrån medborgarnas perspektiv och att skapa ett regelverk för användningen av AI i offentliga tjänster. Avsikten med diskussionen var att öka kunskapen om artificiell intelligens, ge medborgarna en möjlighet att påverka samt testa och utveckla regler som skapats av medborgarna. Det övergripande målet var att skapa ett tekniskt regelverk som kan användas

för att styra språkmodeller och att utveckla och dokumentera hela verksamhetsmodellen så att den även kan användas på andra håll i Europa.

Diskussionen gav upphov till ett omfattande och mångsidigt material om medborgarnas syn på användningen av AI i offentliga tjänster. Utifrån materialet utarbetades ett tekniskt regelverk som påverkar användningen av AI, vilket presenteras i den här promemorian. Resultatet är en värdefull grund som stödjer en ansvarsfull och demokratisk användning av artificiell intelligens i samhället och bidrar till att hitta lösningar för att trygga förtroendet i samhället.

Summary

Artificial intelligence (AI) has been rapidly adopted in daily life in Finnish society. Nonetheless, citizens' trust in its developers and use remains weak. In the development of generative AI, people's sensitive personal data constitute a significant resource. Therefore, it is essential that citizens have a genuine opportunity to influence the direction in which AI is developed and how it is used – particularly in the public sector where decisions have far-reaching impacts on daily life.

Trust in AI emerges from transparency, participation and the recognition of citizens' epistemic rights. The Constitution of Finland guarantees citizens the right to influence decisions that affect them. This right should also extend to the development and deployment of AI. Finns have expressed a strong desire to influence decision-making and shared societal issues easily and quickly online. However, the discussion culture of social media platforms discourages broad participation.

Without transparency and genuine agency, continuous trust in AI cannot be established. Public administration and other societal institutions should see safeguarding citizens' AI agency as their central duty. The framework of informational rights summarises citizens' rights in the age of AI into four key dimensions: access to information, skills and capabilities, participation, and the right to form opinions without coercion. It is essential to strengthen these rights to maintain trust and legitimacy.

The operating model for participatory AI development and the technical AI rulebook based on citizens' views are examples of solutions born from bold experimentation. They demonstrate how citizens' perspectives can be incorporated into AI development and decision-making.

In September 2025, Sitra and Reaktor organised a public discussion titled "What do you think about AI, Finland?" It was run on the anonymous and constructive discussion platform Voxit, with nearly 6,500 participants. During the discussion, 194 statements concerning the development and use of AI were published, of which 169 were drafted by citizens.

The aim of the public discussion was to address the issues of trust from the citizens' perspective and create a rulebook for the use of AI in public services. The discussion sought to increase knowledge about AI, provide citizens with agency to influence the development of AI, and test and develop rules generated by citizens. The goal was to create a technical rulebook to guide language models and to document the entire operating model so that it could also be used elsewhere in Europe.

The discussion generated extensive and diverse data about citizens' views regarding the use of AI in public services. Based on this material, a

technical rulebook influencing AI behaviour was developed and is presented in this memorandum. The process created a valuable foundation for responsible and democratic use of AI in society. It also helps to identify solutions for safeguarding societal trust.

1. Johdanto: Epäluottamus tekoälyyn vaatii toimia

Tekoäly on yhä syvemmin osa arkea, mutta luottamus sen kehittäjiin ja käyttöön on heikko (The University of Melbourne, 2025. The trust, attitudes and use of Artificial Intelligence: a global study 2025). Perinteisesti on ajateltu, että luottamus on välttämätöntä esimerkiksi uusien teknologioiden omaksumisessa.

Voidaan myös ajatella, että koska tavallisten ihmisten verkkokäyttäytymisestä syntyvästä datasta on tehty generatiivisen tekoälyn raaka-ainetta, tulisi heillä myös siksi olla todellinen mahdollisuus vaikuttaa kehityksen suuntaan. Luottamusta syntyy muun muassa läpinäkyvyydestä, osallistumisesta ja kansalaisten tiedollisten oikeuksien tunnustamisesta.

Elämme murrosvaihetta, jossa olemme jo osittain siirtyneet tekoälyaikaan. Julkisessa keskustelussa tekoälystä äänessä ovat lähinnä asiantuntijat. Tekoälyn kehitykseen liittyviä moraalisia ja yhteiskunnallisia valintoja tuntuvat puolestaan tekevän lähes yksinomaan harvalukuinen joukko teknologiajättejä tuottaessaan tekoälymallejaan maailmanlaajuisille markkinoille – julkinen hallinto mukaan lukien. Tavallisten ihmisten ääni on jäänyt lähes kuulumattomiin.

Käyttäytymistämme, verkostojamme ja muuta herkkää tietoa meistä käytetään mm. generatiivisten tekoälymallien opettamiseen. Kun näkymättömissä kerätty tieto yksilöistä muodostaa kriittisen tärkeän perustan tekoälyn kehittämiseksi, pitäisi meillä kaiken järjen mukaan olla mahdollisuus vaikuttaa sen käyttöön ja suuntaan. Tämä saattaa olla epärealistinen tavoite suurimpien tekoälyn kehittäjien suhteen, mutta julkisen hallinnon toimijoiden osalta voisi olla toisin. Tekoäly käsittelee meitä kuitenkin kahta eri kautta. Opettamisen lisäksi tekoälyä käytetään myös yhä enemmän digijärjestelmissä meitä koskevien tietojen käsittelyyn, perinteisen ATK:n rinnalla tai osana.

Suomessa perustuslaki linjaa, että kansalaisilla tulee olla mahdollisuus vaikuttaa itseään koskeviin päätöksiin. Tämän oikeuden tulisi ulottua myös tekoälyn kehittämisessä ja käyttöön otossa tehtäviin valintoihin – erityisesti silloin, kun nämä päätökset vaikuttavat laajasti julkisen palvelun toimintaan ja sitä kautta arkeemme.

Vaikka tutkimukset kertovat (Sitra, 2022. Demokraattiset osallistumismahdollisuudet Suomessa), että kansalaisten luottamus generatiivisiin tekoälyihin ja niiden kehittäjiin on heikkoa niin kansainvälisesti kuin myös Suomessa, uusilla teknologioilla on potentiaalia lisätä esimerkiksi julkisen

sektorin tuottavuutta. Työvoimaa on demografisen muutoksen takia niukenevasti suhteessa tarpeeseen, joten teknologialla on myös potentiaalia paikata syntyvää resurssivajetta.

Tekoälytutkimukset osoittavat myös, että luotettavalla tekoälyllä on merkitystä: jos ihmiset pitävät tekoälyjärjestelmiä luotettavina, he hyväksyvät näiden teknologioiden käytön todennäköisemmin.

Suomalaiset toivovat tutkimusten (Sitra, 2022. Demokraattiset osallistumismahdollisuudet Suomessa) mukaan, että päätöksentekoon ja yhteisiin asioihin voisi vaikuttaa helposti ja nopeasti verkossa. Helposti saavutettavissa sosiaalisissa medioissa kuuluvat kuitenkin äänekkäimpien ja kärjekkäimpien mielipiteet, ja muun muassa siksi moni jättäytyy sivuun. Yhteiskunnallisen luottamuksen kasvattamiseksi ihmiset tulisi ottaa mukaan palveluiden kehitykseen ja suuntaamiseen sekä kansalaisina että yritysten asiakkaina ja palveluiden käyttäjinä. Ilman läpinäkyvyyttä ja todellista toimijuutta emme saa ihmisiä luottamaan tekoälyyn vastaisuudessaan.

Pohjoismailla on pitkä liberaalin demokratian perinne ja yhteiskunnallinen luottamus on yleisesti hyvällä tasolla. Demokratian puolustaminen ja kehittäminen edellyttävät kuitenkin jatkuvaa, avointa ja aktiivista vuoropuhelua kansalaisten kanssa, ja siksi sen tulisi ulottua myös tekoälystä käytävään keskusteluun. Erityisen tärkeää kansalaisten kuuleminen on, kun puhumme verorahoilla maksetuista julkisista palveluista. Niissä käytetään jo nyt automaattista päätöksentekoa muodossa, jos toisessakin, ja tekoälyn lisääminen palveluihin on jatkoa digitaalisten palveluiden kehityksessä.

Vaikka kaikki eivät ole tekoälyasiantuntijoita, kukin ihminen on asiantuntija oman arkensa suhteen ja osaa arvioida, kuinka haluaa itseään palveluissa kohdeltavan.

Tiedolliset oikeudet kansalaisten tekoälytoimijuuden vahvistajina

Jotta kansalaiset voisivat todella vaikuttaa julkisten palvelujen tekoälykehityksessä tehtäviin valintoihin, julkisen hallinnon ja muiden keskeisten yhteiskunnallisten instituutioiden tulisi nähdä kansalaisten tekoälytoimijuuden turvaaminen entistä keskeisemmin omana tehtävänä. Niiden tulisi ymmärtää, että kansalaisten tekoälytoimijuuden turvaaminen on edellytys sille, että tekoälyä lisääntyvässä määrin käyttävät julkiset palvelut säilyttävät jatkossakin luotettavuutensa ja oikeutuksensa kansalaisten silmissä.

Nopeasti kehittyvän tekoälyn aikakaudella kansalaisten toimijuuden turvaaminen vaatii kuitenkin julkiselta hallinnolta ja muilta toimijoilta entistä päämäärätietoisempaa toimintaa ja strategisempaa yhteistyötä. Kaikkien toimijoiden tulisi hahmottaa, miten ne pystyvät vahvistamaan kansalaisten tekoälytoimijuutta luontevasti osana omaa tehtäväkenttäänsä, hyö-

dyntäen käytössään olevia resursseja. Tässä työssä auttaa Sitran vuonna 2025 julkaisema tiedollisten oikeuksien kehikko, joka tiivistää kansalaisten tiedolliset oikeudet tekoälyn aikakaudella neljään tärkeään oikeuteen (Sitra, 2025. Tiedolliset oikeudet tekoälyn aikakaudella). Alkuperäisen kehikon, jota muistiossa on muokattu tekoälykehityksen kontekstiin sopivammaksi, on laatinut viestinnän emeritusprofessori Hannu Nieminen (Nieminen 2024).

Kehikkoa käsittelevässä muistiossa kuvataan myös, minkälaisin konkreettisin toimin esimerkiksi julkinen hallinto, media tai sivistys- ja koulutussektori voivat näitä oikeuksia vahvistaa.

Panostamalla neljän tiedollisen oikeuden vahvistamiseen ja suojeleminen hyvin erilaisetkin organisaatiot pystyisivät tekemään mielekkäästi työtä yhteisen päämäärän eteen. Tiedollisten oikeuksien kehikko tarjoaa yhteisen arvoperustaisen lähtökohdan, joka perustuu jaetulle ymmärrykselle siitä, että nopeasti tekoälyistyvissä yhteiskunnissa ihmisten vaikuttamismahdollisuuksia ja toimijuutta on suojeltava ponnekkaammin. Samalla oikeudet tarjoavat jaetun viitekehityksen, jonka puitteissa eri tahot voisivat kehittää kansalaisten tekoälytoimijuutta lisääviä toimintatapoja ja käytäntöjä.

KUVA 1: Tiedolliset oikeudet vahvistavat kansalaisten toimijuutta kehittyvän tekoälyn aikakaudella.

PÄÄSY TIETOON

Kaikilla ihmisillä tulee olla yhdenvertainen pääsy tekoälyn kehitystä ja käyttöä koskevaan olennaiseen ja totuudenmukaiseen tietoon.

OSAAMINEN JA VALMIUDET

Kaikille ihmisille tulee taata yhdenvertaiset valmiudet ymmärtää tekoälyn kehitykseen sisältyviä mahdollisuuksia ja haasteita. Ihmisten on kyettävä varjelemaan henkilökohtaista tai yhteiskunnan etua tämän ymmärryksen pohjalta.

OSALLISTUMINEN

Kaikille ihmisille tulee suoda yhdenvertainen mahdollisuus vaikuttaa tekoälykehityksen arvopohjaa, suuntaa ja vaikutuksia koskevaan julkiseen keskusteluun ja päätöksentekoon.

EI PAKKOA TAI PAINOSTUSTA

Kaikille ihmisille tulee turvata yhdenvertaiset mahdollisuudet muodostaa näkemyksiä kehittyvästä tekoälystä, ilman suurinta tiedollista valtaa käyttävien pakottamista tai manipulaatiota.

Tiedollisten oikeuksien kehikossa keskeiset kansalaisten tiedollista toimijuutta vahvistavat toimenpiteet on tiivistetty neljään tiedolliseen oikeuteen.

Lähde: Nieminen 2024.

Tiedollisten oikeuksien ajattelu ei vakiinnu yhteiskunnassa, jollei se saa tuekseen tulenkantajia, jotka pyrkivät systemaattisesti vahvistamaan niitä. Tulenkantajien ryhmään kuuluvat itseoikeutetusti julkisen hallinnon organisaatiot, jotka ovat ottamassa palveluissaan laajamittaisemmin käyttöön tekoälyä.

Näiden toimijoiden tulisi tarkastella omaa toimintaansa ja varmistaa, että ne edistävät kaikkia neljää tiedollista oikeutta. Tiedolliset oikeudet velvoittavat julkista hallintoa tuottamaan ja tarjoamaan kansalaisille tekoälyä koskevaa ja tekoälylukutaitoa tukevaa yleistajuista aineistoa. Luomalla ja ottamalla käyttöön uudenlaisia kansalaisten toimijuutta vahvistavia toimintamalleja julkisen hallinnon organisaatiot kykenevät varmistamaan, että tekoälyn käytön rajat julkisissa palveluissa altistetaan julkiselle punta-roinnille. Julkinen hallinto voi tukea tiedollisten oikeuksien leviämistä myös esimerkiksi varmistamalla, että sen omat tekoälyä koskevat ohjeistukset, suositukset, ohjelmat ja linjaukset noudattavat tiedollisten oikeuksien periaatteita.

Kolmannen tiedollisen oikeuden mukaisesti kaikille ihmisille tulee antaa yhdenvertainen mahdollisuus vaikuttaa tekoälykehityksen arvopohjaa, suuntaa ja vaikutuksia koskevaan julkiseen keskusteluun ja päätöksentekoon. Tässä muistiossa esitelty osallistuvan tekoälykehityksen toimintamalli tarjoaa julkisen hallinnon toimijoille viitekehityksen, jolla tätä oikeutta voidaan toteuttaa käytännössä. Tekninen säännöstö on yksi keino konkretisoida kansalaisten näkemyksiä, jotta ne kyettäisiin aidosti huomioimaan julkisen hallinnon tekoälyjärjestelmien kehittämisessä ja sitä koskevassa päätöksenteossa.

Keskustelun tai muistion tarkoituksena ei ole tuottaa oikeudellista analyysia tai uusia lakeja tekoälyn käytölle, vaan esitellä kansalaisten näkemyksiä, ymmärtää haasteita aiheen ympärillä ja kannustaa keskusteluun ratkaisuisista luottamuksen rakentamiseksi.

2. Tapausesimerkki: Mitä mieltä tekoälystä, Suomi? -keskustelu

Tekoälyn käyttöönotto ei ole mustavalkoista. Kysymys ei ole vain siitä, omaksuvatko ihmiset ja organisaatiot tekoälyn vai eivät, vaan pikemminkin kokoelmasta erilaisia suhtautumistapoja, jotka muodostavat ”omaksumisen jatkumon” täydestä käyttämättömyydestä edistyneeseen hyödyntämiseen. Joidenkin selvitysten mukaan suuri osa generatiivisen tekoälyn käytöstä tällä hetkellä on pinnallista. Generatiivista tekoälyä hyödynnetään vaikkapa sähköpostin luonnosteluun tai raportin tiivistämiseen. Vaikka nämä käyttötavat tuottavat ensin vain marginaalista hyötyä, ne kuitenkin tekevät työvälineen tutuksi ja mahdollistavat myöhemmän, syvemmän ja kehittyneemmän tekoälyn käytön.

Jos tekoälyä käyttävien ihmisten määrää halutaan kasvattaa, osaamisen lisäksi tarvitaan myös luottamusta. *Mitä mieltä tekoälystä, Suomi?* -kansalaiskeskustelun päämääränä oli paneutua juuri tuohon haasteeseen *kansalaisten näkökulmasta*, tekemällä luottamusta nauttiva säännöstö tekoälyn käytölle julkisissa palveluissa.

Samalla päätöksentekijöille, virkamiehille ja tekoälypalveluiden kehittäjille haluttiin tietoa mahdollisista ongelmakohtista ja näin tarjota parempi tilannekuva kansalaisten valmiuksista ymmärtää ja hyväksyä tekoälyn käyttö.

Keskustelun tavoitteet

1. Lisätä tietoa tekoälystä yhteiskunnassa. Pureutua tekoälyn haasteisiin, ongelmakohtiin ja mahdollisuuksiin *kansalaisten näkökulmasta*.
2. Antaa kansalaisille *toimijuutta* vaikuttaa perustamalla yhteiskunnan läpi leikkaava, digitaalinen kansalaiskeskustelu tekoälystä.
3. Lisätä *luottamusta* tekoälyn kehittämiseen ja hyödyntämiseen julkisissa palveluissa luomalla *kansalaisten tekemät säännöt* tekoälyn käytölle julkisen sektorin palveluissa. Säännöt koskevat julkisen palvelun tekoälyjärjestelmien suunnittelua, hallintaa sekä teknistä toteutusta.
4. *Testata* kielimallin toimintaan kohdistuvia teknisiä sääntöjä kielimallissa.



5. Luoda toimivista teknisistä säännöistä *systemipromptikirjasto* (”prototyyppi”), jonka avulla kielimalleja voi ohjata kansalaisten toivomalla tavalla.
6. Dokumentoida kansalaisten osallistamisen *toimintamalli*, joka on kopioitavissa ja hyödynnettävissä missä tahansa Euroopan maassa.
7. Viedä tulokset *päättäjille ja virkamiehille*.

Suomen kokeilu suunniteltiin laajalle ihmisjoukolle, julkiseksi keskusteluksi säännöistä tekoälyn käytölle. Tämä tarkoittaa väistämättä sitä, että tarkkojen ohjeiden ja sääntöjen tuottaminen keskusteluaineiston pohjalta on vaikeaa. Tällöin on huomioitava, että tulkintavalta säännöistä siirtyy jonnekin muualle, mutta myös sitä, että merkittävä osa kansalaisten tuottamista väitteistä on kiinnostavia muulla tavoin kuin sääntöinä kielimallille.

Tässä muistiossa avataan tekoälykeskustelukokeilua ja sen pohjalta laadittua osallistuvan tekoälykehityksen toimintamallia eri vaiheineen sekä prosessin lopputuloksena laadittuja kansalaisten tekoälysääntöjä. Jokaisen työvaiheen yhteydessä on koottu käytännön neuvoja, joita työvaiheen onnistunut toteutus edellyttää. Samalla nostetaan esiin ehdotuksia, miten kuvattua toimintamallia voisi edelleen parantaa.

OSALLISTUVAN TEKÖÄLYKEHITYKSEN TOIMINTAMALLI PÄHKINÄNKUORESSA

- Osallistuvan tekoälykehityksen toimintamalli on kehitetty menetelmäksi, jota esimerkiksi eurooppalaiset demokratiat voisivat hyödyntää vahvistaakseen ihmiskeskeistä ja kansalaislähtöistä tekoälykehitystä julkisessa hallinnossa.
- Toimintamallissa kansalaiset osallistuvat tekoälystä käytävään, digitaaliseen kansalliseen keskusteluun. Suomen kokeilussa keskustelu toteutettiin digitaalisella Voxit-osallistumisalustalla.
- Osallistuvan tekoälykehityksen toimintamallin mukaisen prosessin toteuttamiseen tulisi varata aikaa vähintään 3-4 kuukautta.
- Projektin onnistuminen edellyttää riittävää resurssointia esimerkiksi kansalaiskeskustelua tukevaan monipuoliseen viestintään sekä keskustelun moderointiin. Huolehdi myös, että toimintamallin mukaisen prosessin toteuttavassa tiimissä on monipuolista osaamista esimerkiksi digitaalisesta kansalaisosallistamisesta, teknisestä ymmärryksestä ja eettisestä tekoälykehityksestä.
- Pyri varmistamaan, että prosessiin on sitoutunut jo alkuvaiheessa jokin julkisen hallinnon toimija, joka pystyy hyödyntämään tekoälysäännösten tuloksia omassa palvelukehityksessään. Suomessa tällaista kumppania ei vielä ollut, koska kyse oli varhaisesta kokeilusta.
- Luottamuksen vahvistamiseen tähtäävän toimintamallin kannalta on olennaista, että tulosten tulkinta ja analysointi yhtä lailla kuin Voxit-tulosraportti ovat läpinäkyviä ja saatavilla.

3. Digitaalisen kansalaiskeskustelun käynnistäminen ja kutsuprosessi

Digitaalinen kansalaiskeskustelu on paikallaan silloin, kun tarvitaan laajojen ihmisjoukkojen näkemyksiä, ja esimerkiksi päättäjät ja virkamiehet tarvitsevat tilannekuvaa jostakin kansalaisten kannalta tärkeästä aiheesta. Tällöin on usein myös tarve rakentaa laajempaa ymmärrystä kansalaisten ja hallinnon edustajien kesken.

Generatiivisen tekoälyn käyttö julkisissa palveluissa sopi hyvin kansalaiskeskustelun aiheeksi, koska se vaikuttaa lähes jokaisen ihmisen elämään. Se on myös palveluiden kehittämisen kannalta olennainen teknologia. Sitä on kuitenkin hyödynnettävä kansalaisia palvelevalla ja reilulla tavalla, jotta luottamus julkiseen hallintoon ei järky.

On tärkeää myös ymmärtää, mistä epäluottamus tekoölyyn ja sen tekijöihin kumpuaa, ja mihin se suuntautuu, jotta ongelmakohtiin voidaan tarttua.

Näiden tarpeiden pohjalta käynnistettiin *Mitä mieltä tekoälystä, Suomi?*-kokeiluhanke, jossa suomalaiset saivat kutsun keskustelemaan digitaalisesti tekoälyn riskeistä ja mahdollisuuksista julkishallinnon palveluissa.

Lupaus keskustelun osallistujille

Kansalaiskeskustelu syntyi tarpeesta lisätä ymmärrystä suomalaisten tekoälyyn liittyvistä tarpeista, toiveista ja huolista, joita he halusivat julkisen sektorin päättäjien ymmärtävän.

Osallistujille luvattiin, että keskustelun avulla kerätään aineistoa, jolla voidaan rakentaa teknistä säännöstöä tekoälylle. Aineistoa tulnaisiin käyttämään jonkin kielimallin koulutusaineistona eli testaamaan säännöstöidean toimivuutta. Vastaavaa tavallisten kansalaisten näkemyksiin perustuvaa säännöstöä tekoälylle ei ole aiemmin laadittu missään maailmassa.

Osallistujille kerrottiin myös, että tulokset viedään päättäjien ja virkamiesten tiedoksi, ja pidemmän aikavälin tavoitteena voisi olla kansalaisten näkemysten mukaan toimiva tekoälymalli julkisille palveluille.

Menetelmän valinta

Tekoälykeskustelun suunnittelusta asti tavoiteltiin toimintamallia, jonka voisi soveltuvin osin toistaa missä tahansa muualla, kun kansalaisten mielipiteitä halutaan demokratian pelisääntöjen mukaisesti kuulla.

Kun tavoitellaan laajojen joukkojen osallistumista, on tärkeää madaltaa osallistumisen kynnystä kaikin tavoin. Keskusteluun osallistumisessa on olennaista nopeus, helppous ja ajasta ja paikasta riippumattomuus. Tällöin vaihtoehtoisiksi sopivat erilaiset rakentavaan keskusteluun suunnitellut, digitaaliset demokratiateknologiat, kuten osallistumisalustat.

Osallistujien on kyettävä myös luottamaan häiriöttömään ja turvalliseen tapaan osallistua yhteiskunnalliseen keskusteluun, joten anonyymin osallistumismahdollisuuden tarjoaminen on välttämätöntä. Keskustelun verkkohäirintä voidaan torjua valitsemalla alusta, jossa moderoinnilla voidaan vaikuttaa keskustelun laatuun.

Suomen kansalaiskeskustelun alustaksi valittiin laajojen joukkojen osallistamiseen suunniteltu, digitaalinen alusta *Voxit*. Se ei toimintaperiaatteidensa ja moderointiominaisuuksiensa ansiosta mahdollista sosiaalisesta mediasta tuttua maalittamista, häirintää, osallistujien tunnistamista tai maltillisen enemmistön hiljaiseksi huutamista. Keskustelu pysyy aidosti rakentavana ja näkyvyyttä osallistujille asiallisena ja ratkaisukeskeisenä.

Kuva 2. Mitä mieltä tekoälystä, Suomi? -keskustelun mainoksia.



Suomalaiset kutsuttiin kertomaan, mitä riskejä tai mahdollisuuksia he näkevät tekoälyn käytössä kansalaisten käyttämissä julkishallinnon digipalveluissa. Ihmiset ohjattiin kertomaan näkemyksensä rakentavaan keskusteluun suunnitellulle, digitaaliselle Voxit-osallistumisalustalle.

MIKÄ ON VOXIT?

Voxit sai alkunsa USA:ssa 2010-luvun puolivälissä, kun yhdysvaltalainen Computational Democracy -kansalaisjärjestö kehitti Polis-nimisen alustan rakentavaa kansalaiskeskustelua varten. Voxit on avoimen lähdekoodin ratkaisu, joka on Suomessa irrotettu Poliksesta omaksi kehityshaarakseen. Sitä on kehitetty vastaamaan EU:n tietosuoja-asetusta sekä eurooppalaisia saavutettavuusvaatimuksia. Suomessa alustaa on kehitetty erityisesti älypuhelimella käytettäväksi. Voxit-nimi tulee latinan sanasta vox, joka tarkoittaa ääntä.

Voxit-alusta tuottaa reaaliaikaista tietoa osallistujien mielipiteiden jakautumisesta. Tuloksista näkee, mistä väitteistä enemmistö osallistujista on yhtä mieltä ja missä mielipiteet jakaantuvat. Tilastoanalyysia hyödyntävä alusta myös jaottelee osallistujat erilaisiin mielipideryhmiin, jotka eroavat toisistaan äänestyskäyttäytymisensä perusteella. Näin keskustelun tulokset lisäävät kansalaisten ja päätöksentekijöiden ymmärrystä käsiteltävästä aiheesta ja parantavat päätöksenteon tietopohjaa.

Alustan keskeinen ominaisuus on se, että osallistujat voivat äänestää väitteistä, mutta myös ehdottaa keskusteluun omia väitteitään muiden äänestettäviksi ja varmistaa siten, että heille tärkeät näkökulmat tulevat esiin. Tyypillisesti Voxit-keskustelu on auki noin kuukauden ajan ja osallistujat voivat ehdottaa uusia väitteitä keskustelun alkupuoliskolla. Keskustelun loppupuolella alustalla julkaistuja väitteitä voi ainoastaan äänestää.

Kansalaisten kutsuminen Voxit-keskusteluun

Suomen kokeilussa kansalaisia lähestyttiin satunnaisotannalla valituille henkilöille lähetetyillä paperikutsuilla, sähköpostiuutiskirjeillä ja muilla sidosryhmäviesteillä, somepostauksilla sekä digimainoksilla muun muassa iltapäivälehdissä ja sosiaalisissa medioissa. Keskustelu oli avoin kaikille eli esimerkiksi Suomen kansalaisuutta ei edellytetty. Samaan keskusteluun oli mahdollisuus osallistua suomeksi, ruotsiksi tai englanniksi. Voxitissa on käännöstyökalu, joten väitteitä voi lisätä useammalla kielellä ja toisaalta kääntää niitä omalle kielelleen.

Kutsuprosessissa ei pyritty täydelliseen tilastolliseen edustavuuteen Suomen väestöstä, vaan haluttiin ennemmin rekrytoida mahdollisimman suuria ihmismääriä.

MITÄ MIELTÄ TEKOÄLYSTÄ, SUOMI? -KANSALAIKESKUSTELUN KUTSUPROSESSIN VAIHEET

1. Edustava otos ja paperikirje

- a. Digi- ja väestötietovirastolta (DVV) hankittiin suomalaista väestöä edustava 10 000 henkilön satunnaisotanta osoitteineen ja joukolle lähetettiin postitse paperikutsu osallistua keskusteluun jo ennen muuta viestintää. DVV on valtion virasto, joka pitää rekisteriä Suomen väestöstä ja rakennuskannasta ja tuottaa niihin liittyen monenlaisia palveluita. Kutsussa oli QR-koodi helpottamassa osallistumista älypuhelimella. Kutsu lähetettiin suomeksi ja ruotsiksi.
- b. Paperikirjeessä oli myös ohje siihen, kuinka Voxitia voi käyttää tuetusti, jotta myös digitaalisia laitteita käyttämättömät kykenisivät osallistumaan. Koska osallistuminen on anonyymia, eikä palveluun tarvita kirjautumista, tämä on helposti toteutettavissa.

2. Display- ja somemainonta

- c. Osallistujia tavoiteltiin iltapäivälehdien verkkosivujen ns. display-mainosten kautta sekä saman mediakonsernin muissa verkkopalveluissa. Mainoksissa hyödynnettiin Vokit-keskustelussa julkaistuja, kiinnostusta herättäviä väitteitä.
- d. Ihmisiä tavoiteltiin myös sosiaalisessa mediassa mainoksilla, joihin nostettiin samoja väitteitä kuin display-mainontaan. Somekanaviksi valikoituivat Instagram, Jodel, LinkedIn ja Facebook erityyppisten ja -ikäisten kohderyhmien tavoittamiseksi. Maksettua mainontaa tehtiin Instagramissa, Facebookissa ja Jodelissa.

3. Muu viestintä

- a. Liike-elämässä, julkisella sektorilla ja järjestöissä toimivia päättäjiä ja asiantuntijoita tavoiteltiin sähköisellä suoraviestinnällä sekä kokeilun toteuttaneiden organisaatioiden sosiaalisen median tileillä julkaistuin viestein. Lisäksi lanseerauksessa hyödynnettiin tapahtumamarkkinointia ja paneelikeskustelua.
- b. Keskustelun alkamisesta lähetettiin tiedote medialle tiedotejakelupalvelun laajalla jakelulla.

Väitteet Voxit-alustalla

Voxit-keskustelu käynnistetään julkaisemalla alustalla aluksi 15–25 aloitusväitettä, jotka jakautuvat tausta- ja siemenväitteisiin.

Taustaväitteet

Voxit-alusta ei ole kyselytyökalu, joten se ei tarjoa yksityiskohtaista sosio-demografista tietoa osallistujista. Voxit-keskustelussa aluksi julkaistavat tausta- eli metaväitteet antavat kuitenkin keskustelun järjestäjälle mahdollisuuden poimia osallistujista tiettyjä perusasioita ja seurata, keitä keskusteluun osallistuu. Taustaväitteitä voidaan myös hyödyntää korrelaatioanalyysissä.

Esimerkki. *”Olen alle 20-vuotias”* tai *”Minulla on korkeakoulututkinto.”*

Taustaväitteet vievät ilmatilaa keskustelusta, eivätkä anna erityistä lisäarvoa osallistujille, joten niiden määrä on syytä rajata tarkasti. Suomen kokeilussa ne rajattiin viiteen.

Siemenväitteet

Taustaväitteiden lisäksi Voxit-alustalla julkaistaan aluksi pieni joukko muita, keskustelun käynnistäviä väitteitä, joita kutsutaan siemenväitteiksi. Niiden muotoiluun kannattaa varata riittävästi aikaa, sillä niillä tarkennetaan keskustelun aihetta ja ohjataan keskustelua tavoitteeseen.

Esimerkki. *”Epäilen, että tekoäly painottaa liikaa vähemmistöjen etuja julkisissa palveluissa.”*

Kansalaisten tekoälysäännöstöä muodostettaessa Voxit-keskustelun siemen- ja taustaväitteiden tekeminen vaatii tavallistakin tarkempaa pohdintaa. Kun tavoitteena on, että Voxit-väiteaineiston perusteella kyetään muodostamaan julkisten palveluiden tekoälykehitystä ja -käyttöä ohjaavia kansalaisten tekoälysääntöjä, siemenväitteillä tulisi pystyä ohjaamaan osallistujia tuottamaan kyllin konkreettisia väitteitä. Toisaalta siemenväitteissä ei pitäisi kuitenkaan kuvata tekoälyn käyttöä missään yksittäisessä käyttötapaüksessa, koska tällaisten väitteiden pohjalta on vaikeaa luoda yleisempää sääntöä tekoälyn käytölle julkisissa palveluissa. Oikean abstraktiotason tavoittaminen oli vaikeaa ja vaati toteuttajatiimin sisäistä keskustelua ja väitteiden yhteistä työstämistä.

Laadittaessa siemenväitteitä kansalaisten tekoälysääntöjä pohjustavaan Voxit-keskusteluun olennaista on myös kuvata selkeästi, mitä yhteisiä ohje-nuoria keskustelun toteutustiimi on noudattanut siemenväitteitä muodos-

taessaan. Tämä on tärkeää, jotta voidaan osoittaa tiimin pyrkineen muodostamaan siemenväitteet tasapuolisesti ja eri näkökulmia korostaen, yrittämättä kuljettaa keskustelua mihinkään tiettyyn suuntaan.

Siemenväitteiden muodostaminen

Mitä mieltä tekoälystä, Suomi? - keskustelussa

Suomessa toteutetussa kokeilussa siemenväitteiden laatimista ohjasi kourallinen jo olemassa olevia tekstejä, joissa määritellään periaatteita tekoälyn eettiselle, ihmiskeskeiselle tai demokraattiselle hyödyntämiselle. Ottamalla nämä dokumentit ohjenuorikseen kokeilun toteutustiimi pyrki varmistamaan, että siemenväitteet kattaisivat yleisesti tunnistetut ja laajasti omaksutut eettisen tekoälykehittämisen näkökulmat.

Valitut dokumentit olivat: EU:n tietosuojaa-asetus, Unescon tekoälyn etiikkaa koskevat suositukset, Sitran julkaisemat tekoälyajan tiedolliset oikeudet sekä Helsingin ja Tampereen kaupunkien laatimat periaatteet datan ja tekoälyn eettiselle käytölle. Kokeilun toteuttajatiimi oli myös tietoinen Yhdysvalloissa vuonna 2023 toteutetusta pienemmästä kokeilusta, jossa joukko tekoälystä kiinnostuneita henkilöitä kutsuttiin luomaan tekninen tekoälysäännöstö. Yhdysvaltalaiskokeilussa tulosten pohjalta muodostettiin 75 sääntöä, joita käytettiin tekoälymallin kouluttamiseen.

Kun Suomen kokeilussa alettiin luoda siemenväitteitä, eettistä ja ihmiskeskeistä tekoälyä koskevien dokumenttien sisältämät suositukset yhdistettiin ensin yhtenäiseksi arvokimpuksi. Kimppu koostui 13 ylätasoa periaatteesta (liite 1), jotka kattoivat kaikki lähtienä käytettyjen tekstien esiin nostamat eettiset näkökulmat. Arvokimppun periaatteet koskivat muun muassa oikeudenmukaisuutta ja syrjimättömyyttä, tietosuojaa, turvallisuutta, selitettävyyttä sekä tarvetta tukea kansalaisten tekoälyvalmiuksia ja mahdollisuuksia osallistua tekoälykehitykseen. Niissä korostettiin myös ihmisten tekemää valvontaa, vastuuta ja viimekätistä päätösvaltaa suhteessa tekoälyjärjestelmien tekemisiin ratkaisuihin.

Arvokimppun muodostamisen jälkeen kokeilun työryhmä pyrki muodostamaan kaikkien 13 periaatteen alle konkretisoivia siemenväittämiä. Työ aloitettiin sijoittamalla yhdysvaltalaiskokeilussa laaditut 75 väitettä sopivien periaatteiden yhteyteen, jolloin vain muutama väite jäi yli. Sen jälkeen työryhmä alkoi karsia väitteitä, muokata niitä Suomen tekoälykeskusteluun sopivammiksi ja tuottaa kokonaan uusia siemenväitteitä, käyttäen samalla runsaasti aikaa väitteissä käytettyjen sanamuotojen hiomiseen. Iteratiivisen prosessin lopputuloksena siemenväitteitä muodostettiin lopulta yhteensä 21 (ks. liite 2). Toteuttajatiimin jäsenet tekivät pääosan tästä työstä erillään, koska siemenväitteiden laatimiseen ennakoon varattu yhteinen työskentelyaika osoittautui riittämättömäksi. Loppuvaiheessa prosessia vaivasi muutoinkin kiire, minkä seurauksena työryhmä ei ehtinyt arvioida kattavasti, heijastelivatko lopulliset siemenväitteet yhä riittävän uskollisesti eettisten periaatteiden pohjalta muodostettua arvokimppua.

Siemenväitteiden laadintaprosessia kehitettäessä aikaa tälle työvaiheelle tulisikin varata runsaasti, jotta siemenväitteitä voidaan lopuksi arvioida uudelleen niiden laatimista ohjanneiden periaatteiden valossa. Aikataulussa tulisi olla niin paljon väljyyttä, että siemenväitteisiin ehditään tehdä tarvittaessa vielä muutoksia ennen niiden julkaisua Vokit-alustalla. Työmäärää voi kuitenkin keventää käyttämällä esim. tässä kokeilussa testattuja väitteitä pohjana.

Siemenväitteiden arviointiprosessi voitaisiin toteuttaa lopuksi esimerkiksi noudattaen menettelyä, jota Suomen kokeilussa noudatettiin aluksi käsiteltäessä yhdysvaltalaisessa säännöstökokeilussa tuotettuja väitteitä. Kun laaditut siemenväitteet sijoitetaan väitteiden laatimista ohjanneiden periaatteiden alle, voidaan nähdä, kattavatko siemenväitteet riittävän laajan kirjon eettisen ja ihmiskeskeisen tekoälykehityksen keskeisiä näkökulmia. Näin voidaan varmistua, että työprosessia ohjanneet periaatteet eivät ole iteratiivisen työprosessin aikana huomaamatta liudentuneet.

Kansalaisten väitteet

Vokit-alustan käytön keskeisenä pyrkimyksenä on vahvistaa kansalaisten toimijuutta. Tämä periaate tulee selvästi näkyviin suhtautumisessa alustalla julkaistaviin väitteisiin. Keskustelun järjestäjän laatimat siemen- ja taustaväitteet vain käynnistävät keskustelun ja viitoittavat sen teemaa, mutta pääosa alustalla äänestettävistä väitteistä tulee kansalaisilta itseltään. Näin osallistujilla on paljon valtaa viedä keskustelua haluamaansa suuntaan.

Kun osallistujat ehdottavat Vokit-keskusteluun omia väitteitään, keskustelussa nousee usein esiin aivan uusia näkökulmia, joita keskustelun järjestäjä ei ole tullut ajatelleeksi. Osallistujien lisäämät väitteet voivat myös korjata tai tarkentaa alkuperäisten siemenväitteiden sisältämiä ajatuksia. Se vahvistaa osallistujien toimijuutta keskusteluaiheen määrittelyssä ja sen käsittelyssä.

Tämä tuli näkyviin myös Suomen tekoälykeskustelussa. Keskustelun sulkemispäivään mennessä alustalla oli yhteensä 194 tekoälyaiheista väitettä, joista 169 oli kansalaisten lisäämiä. Ennen keskustelun alkua kokeilun toteutustiimi oli sopinut, että alustalla julkaistaisiin enintään 200 väitettä. Rajaamalla väitteiden määrä kahteensataan haluttiin varmistua, ettei keskusteluun osallistumisesta muodostu osallistujille liian raskasta kokemusta. Vokit-keskustelu oli auki syyskuussa 2025 vajaan kuukauden ajan, mutta suomalaisilla oli mahdollisuus lisätä omia väitteitään noin kaksi ensimmäistä viikkoa.

PROJEKTIN TOTEUTTAJAN MUISTILISTA VOXIT-KESKUSTELUA VALMISTELTAESSA

- Vokit-keskustelun käynnistävien väitteiden muotoilu ymmärrettävään muotoon sekä väitteissä käsiteltävien aiheiden rajaaminen on haastavaa. Varaa tähän prosessiin riittävästi aikaa.
- Pyri kirjoittamaan väitteet mahdollisimman yksiselitteisesti, jotta kaikki osallistujat ymmärtävät ne samalla tavalla. Muista, että siemenväitteisiin valitsemasi tyyli ohjaa Vokit-keskustelun osallistujien ajatuksia myös siitä, kuinka arkista tai virallista kieltä väitteitä muodostaessa tulisi käyttää. Siemenväitteissä esiin nostetut aiheet puolestaan vahvistavat osallistujien käsitystä siitä, kuinka konkreettisella tai yleisellä tasolla keskustelun toivotaan etenevän.
- Kokoa verkkosivuillesi kaikki kansalaiskeskustelua koskevat tiedot yhteen paikkaan. Sivustolla voit muun muassa antaa lisätietoa käyttämästäsi digitaalisesta osallistumismenetelmästä sekä keskustelun pelisäännöistä.
- Suomen kokeilussa Vokit-keskusteluun saivat osallistua kaikki Suomessa asuvat ihmiset. Osallistuminen ei vaatinut tunnistautumista tai käyttäjätilin perustamista, joten käytännössä kuka vain saattoi osallistua. Ulkomailta tuleva verkkoliikenne oli estetty eli alustan käyttö rajattiin Suomeen mahdollisen verkkohäirinnän estämiseksi.

Keskustelun moderointi

Tekoälyn käyttö julkisissa palveluissa on keskustelunaiheena laaja, eikä valtaosalla kansalaisista ole syvällistä tekoäly-ymmärrystä tai tietämystä esimerkiksi julkisen hallinnon automaattista päätöksentekoa tai yksityisyyttä koskevasta sääntelystä. Kansalaisten tekoälysäätelyn luomiseen tähtäävässä kansalaiskeskustelussa tätä ei pitäisi edellyttääkään. Ihmisten ei tarvitse ymmärtää perusteellisesti tekoälyjärjestelmien toimintaperiaatteita voidakseen ilmaista mielipiteensä siitä, miltä arvopohjalta julkisten palveluiden tekoälyä tulee kehittää tai minkälaisia rajoituksia tekoälyn käytölle täytyy asettaa.

Suomen tekoälykeskustelussa erot kansalaisten tekoälytietämyksen tasossa tulivat selvästi esiin esimerkiksi siinä, miten tekoäly terminä ymmärrettiin. Keskustelun aikana huomattiin, että merkittävä osa suomalaisista ei osannut tehdä eroa generatiivisten ja muiden tekoälyjen välillä. Keskustelun aiheen tarkempi määrittely olisi saattanut auttaa osaa kansalaisista. Tekoälykeskustelun järjestäjän kannattaa silti varautua siihen, että todennäköisesti melko harva lukee keskustelun aiheeseen liittyvää tarkempaa kuvausta, vaikka sellainen olisikin helposti saatavilla.

Erot kansalaisten tekoälytietämyksen ja -osaamisen tasossa vaikuttavat väistämättä myös Vokit-keskusteluun ehdotettuihin väitteisiin. Keskustelun moderaattorien kannattaakin varautua jo ennakoon siihen, että osa ehdo-

tetuista väitteistä sisältää asiavirheitä. Samoin on todennäköistä, että ihmiset haluavat käsitellä ehdottamissaan väitteissä hyvin laajasti tekoälyn käyttöä yhteiskunnassa. Tämä huomattiin Suomen kokeilussa, missä ihmisillä oli valtava tarve saada puhua tekoälystä yleisellä ja yhteiskunnallisella tasolla. Kansalaiset halusivat käsitellä esimerkiksi tekoälykasvatuksen tärkeyttä tai julkisen hallinnon tekoälyhankintojen sääntöjä.

Hyväksytyjä väitteitä:

”Tekoälyn käyttö lisää ihmisten eriarvoisuutta, koska kaikki eivät osaa tai ymmärrä sitä.” *”Eurooppalainen kielimalli on meille parempi valinta, koska pääsemme EU:n kautta vaikuttamaan sitä koskevaan päätöksentekoon”.*

Väitteistä huomasit myös, ettei kaikilla kansalaisilla ollut tietämystä esimerkiksi tekoälyavusteisen päätöksenteon rajoista. Joillakuilla ei myöskään ollut halua hankkia lisää tietoa.

Hyväksyty väite:

”Terveystieteiden huoltoon liittyviä käyntejä ja ajanvarauksia ei saa nauhoittaa tai käyttää tekoälyn opetusmateriaalina ilman lupaa.”

Hylätty väite:

”En halua tekoälyä, olen 82-vuotias.”

Moderaattori pitää käsissään merkittävää valtaa, joten rooliin on syytä valmistautua huolellisesti ja vastuunsa ymmärtäen. Keskusteluun tulevat väitteet voivat olla ristiriidassa moderaattorin oman arvomaailman kanssa, mutta se ei saa vaikuttaa päätöksiin. Hankalat tai ristiriitaiset väitteet tuovat keskusteluun mielenkiintoa ja toisaalta niiden avulla voidaan saada selvyyttä ihmisille tärkeisiin asioihin.

Jotta jopa tuhansien kansalaisten yhteinen tekoälykeskustelu säilyisi laajasta aihepiiristään ja suuresta osanottajajoukostaan huolimatta laadukkaana, Voxit-keskustelun moderaattorien on suositeltavaa päättää etukäteen muutamista keskeisistä moderointilinjauksista. Yhtä tärkeää on laatia keskustelulle selkeät moderointisäännöt (ks. liite 2) ja julkaista ne keskusteluun osallistuvien nähtäville.

Suomen kokeilussa keskustelua pyrittiin moderoimaan melko tiukasti, vaikkakin alkuvaiheessa seula oli lempeämpi, jotta keskustelu käynnistyisi vilkkaasti. Lopulta kansalaisten ehdottamista väitteistä miltei 400 hylättiin. Suurin osa hylätyistä väitteistä jätettiin julkaisematta, koska keskustelussa oli jo samansisältöinen väite, tai koska ehdotettu väite oli liian yleisluontoinen, eikä soveltunut tekoälynsäännöstön materiaaliksi. Vain muutama väite hylättiin, koska ne olivat sisällöltään asiattomia.

Ennen Voxid-keskustelun käynnistymistä Suomessa moderaattorit sopivat myös yhteisestä ohjenuorasta, joka koski yksittäisiä alateemoja käsittelevien väitteiden lukumäärää. Sovitun linjan mukaisesti mihinkään alateemaan liittyvien väitteiden määrä ei saanut ylittää viidesosaa kaikista keskusteluun hyväksytyistä väitteistä. Tällä peukalosäännöllä pyrittiin varmistamaan, ettei mikään julkisten palveluiden tekoälykäyttöön liittyvä yksittäinen näkökulma, kuten tekoälyä hyödyntävien palveluiden yksityisyydensuoja tai tietosuoja, pääsisi hallitsemaan keskustelua liikaa.

PROJEKTIN TOTEUTTAJAN MUISTILISTA VOXIT-KESKUSTELUN MODEROINTIIN

- Tuhansien osallistujien laajuisen Voxid-keskustelun moderointi edellyttää riittävää resursointia ja henkilökohtaista valmistautumista keskustelun moderointiin. Moderaattoreita tulee olla vähintään 2–3 ja moderointiin tulee voida käyttää päivittäin muutamia tunteja aikaa.
- Voxid-keskustelun moderointia suunnitellessa on hyvä muistaa, että moderointitarve on lyhytaikaista. Osallistujien mahdollisuus ehdottaa Voxid-keskusteluun uusia väitteitä päättyy usein keskustelun puolivälin tietämissä. Sen jälkeen moderointia ei enää tarvita. Voxid-keskustelun moderointitarvetta voi rajata myös keskustelun alkuvaiheessa ilmoittamalla moderointisäännöissä, että keskustelua moderoidaan ainoastaan tiettyihin aikoihin, esimerkiksi arkipäivisin ja virka-aikaan.
- Laajan kansallisen Voxid-keskustelun moderointitarvetta voi suunnitella etukäteen, kun tiedetään, miten Voxid-keskustelua tukeva viestintä aiotaan jaksottaa. Käytännössä aina, kun keskustelusta viestitään näytävämmin, Voxid-alustalle muodostuu osallistujapiikki. Kun alustalla vieraillee runsaasti osallistujia, he ehdottavat enemmän uusia väitteitä, jotka moderaattorien tulisi käydä läpi mahdollisimman ripeästi.
- Hyvistä ennakoivalmisteluista huolimatta ison Voxid-keskustelun moderointi on ajoittain haastavaa. Vaikka keskustelulle on luotu hyvät moderointisäännöt, ne jättävät kuitenkin tilaa tulkinnoille. Kiperissä tapauksissa tätä tulkintaa on parasta tehdä yhdessä keskustellen muiden moderaattorien kanssa.

4. Voxit-aineiston rajaaminen kansalaissäätöjen muodostamista varten

Kun Voxit-keskustelu on päättynyt, on aika käynnistää Voxit-alustalla kerätyn data-aineiston analyysi. Tämä työ tehdään asiantuntijatyönä. Tavoitteena on muodostaa kansalaisten näkemysten pohjalta sääntöjä, jotka tukevat julkisten palveluiden tekoälyjärjestelmien hallintaa ja kehittämistä. Koska Voxit-alustalta saatu väiteaineisto kattaa todennäköisesti hyvin laajan kirjon erilaisia tekoälyn hyödyntämiseen liittyviä näkökulmia, analyysivaiheessa tulisi varautua siihen, että vain pieni osa kansalaisten laatimista väitteistä voidaan ottaa säännöstöön mukaan sellaisenaan.

Kansalaissäännösten pohjaväitteiden valikointia ohjaavat periaatteet

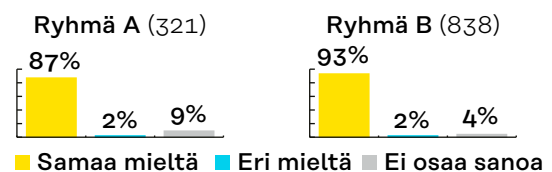
Kansalaissäännösten pohjamateriaalin valikointi aloitetaan laatimalla periaatteet, joihin nojautuen väitteiden valinta voidaan toteuttaa läpinäkyvästi ja järjestelmällisesti. Koska tavoitteena on luoda kansalaisten jaettua arvo-pohjaa ja näkemyksiä todenmukaisesti heijastelevia tekoälysääntöjä, yhtä tärkeää on varmistaa, että säännösten pohjamateriaalina toimivat väitteet ovat saaneet Voxit-keskustelussa osakseen riittävän laajaa tukea.

Näistä lähtökohdista suomalaistyöryhmä muodosti oman analyysinsä tueksi kaksi periaatetta, joiden perusteella väiteaineisto rajattiin.

1. Voxit-alusta muodosti kansalaiskeskusteluun osallistuneista suomalaisista kaksi erillistä mielipideryhmää: ryhmä A, 868 henkilöä (13,5% kaikista osallistujista) ja ryhmä B, 3799 henkilöä (59% kaikista osallistujista). Analyysivaiheessa kansalaissäännösten pohjamateriaaliksi poimitiin ensiksi väitteet, joista enemmistö - raja-arvoksi valittiin 60 prosenttia - molemmissa mielipideryhmistä oli äänestänyt samalla tavalla.

Esimerkki kansalaissäännösten pohja-aineistoon mukaan otetusta väitteestä, josta vähintään 60 prosenttia molempiin mielipideryhmiin kuuluvista oli samaa mieltä.

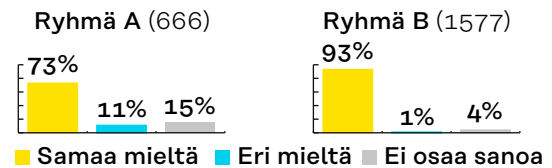
Tekoälyn tulee kunnioittaa YK:n ihmisoikeuksien julistusta, lasten oikeuksia sekä Suomen perustuslakia.



2. Tämän jälkeen väiteaineistoa karsittiin vielä toisella leikkurilla, jotta varmistuttiin, että kaikkia kansalaissääntöjen pohjamateriaaliksi hyväksytyjä väitteitä oli äänestänyt riittävän suuri joukko kummassakin mielipideryhmässä. Periaatteeksi sovittiin, että väiteaineistoon kelpuutettiin vain sellaiset väitteet, joita oli äänestänyt vähintään 25 prosenttia molempien mielipideryhmien jäsenistä. Tämä merkitsi, että ryhmässä A väitettä oli äänestänyt ainakin 217 osallistujaa ja ryhmässä B ainakin 950 osallistujaa.

Esimerkki kansalaissäännösten pohja-aineistoon mukaan otetusta väitteestä, jota molemmissa mielipideryhmissä äänesti vähintään neljännes ryhmän jäsenistä.

Jos tekoälyä käytetään osana julkisia palveluita, siihen liittyvä data ei saa päätyä yksityisille toimijoille.



Näitä periaatteita noudattaen kansalaissääntöjen pohja-aineistoon valikoitui 61 väitettä kaikista Voxit-keskustelussa julkaistuista 194 väitteestä. Väitteet toimivat työryhmän aineistona, jonka pohjalta seuraavassa työvaiheessa ryhdyttiin luomaan kansalaisten sääntöjä julkisen palvelun tekoälykäytölle.

Käytetyn menetelmän vahvuutena voi pitää sitä, että kaksiportainen valikointiprosessi siivoloi väiteaineistosta esiin väitteet, joihin Voxit-keskusteluun osallistuneet suomalaiset suhtautuivat varsin yksimielisesti. Näin kansalaissääntöjen pohjaksi valittujen väitteiden voi ajatella heijastelevan kaikkien osallistujien, eikä pelkästään jonkin yksittäisen alaryhmän, laajasti jakamia näkemyksiä ja arvoja.

Menetelmän keskeinen heikkous on puolestaan se, että kansalaisten ja eri mielipideryhmien näkemyksistä saattaa muodostua staattinen ja konsensusta liikaa korostava kuva, kun kansalaissääntöjen pohjaksi valitaan vain laajaa yksimielisyyttä saavuttaneet väitteet.

Tämä huomattiin myös Suomen kokeilussa, sillä kansalaissääntöjen materiaaliksi valittujen väitteiden ulkopuolelle jäi joukko väitteitä, joista olisi voinut melko vaivatta muodostaa julkisten palveluiden tekoälyjärjestelmien käyttöä tai kehittämistä koskevan kansalaissäännön. Näitä väitteitä ei kuitenkaan huomioitu kansalaissääntöjä luotaessa, koska ne eivät täyttäneet toista tai kumpaakaan sovituista analyysiperiaatteista.

Esimerkki materiaalin ulkopuolelle jääneestä väitteestä:

“Chattibottin tulisi olla esimerkillisen kohtelias ja avulias asiointipalveluissa.”

Valitun analyysitavan vuoksi kansalaissääntöjen pohjan muodostavien väitteiden ulkopuolelle jäi myös väitteitä, joista Voxit-alustan muodosta-

massa kahdessa mielipideryhmässä oltiin selvästi eri mieltä. Tällöin väite olisi läpäissyt kansalaissäätömaterialille analyysivaiheessa asetetut kriteerit yhdessä mielipideryhmässä, mutta ei toisessa.

Kansalaissäätöjen materiaalin ulkopuolelle jääneitä väitteitä:

”Julkisissa palveluissa voi käyttää tekoälyä päätöksentekoon laajasti, jos asianosaiset voivat aina pyytää oikaisua ihmiseltä.”

”Tekoälyä voidaan käyttää lastensuojelussa päätöksenteon tukena.”

Kansalaissäätöjen ulkopuolisten väitteiden analyysi

Kun kansalaissäätöjen pohjamateriaalina käytettävät väitteet on eroteltu muusta väiteaineistosta, jäljelle jää luultavasti vielä suuri joukko kansalaisten muodostamia väitteitä, joita ei ole voitu hyödyntää säännösten muodostamisessa. Sen vuoksi analyysia on suositeltavaa jatkaa senkin jälkeen, kun kansalaissäätöjen pohja-aineiston muodostavat väitteet on kerätty.

Suomen kokeilussa tässä vaiheessa analyysiin otettiin mukaan kaikki Voxit-keskustelussa julkaistut väitteet. Analyysissa tarkasteltiin muun muassa, mihin julkisen hallinnon tekoälykehittämiseen kytkeytyviin toimiin suomalaisten keskuudessa liittyy laajalti huolenaiheita ja epätietoisuutta tai toisaalta laajaa kiinnostusta ja toiveikkuutta. Huomiota kiinnitettiin myös keskustelussa syntyneitä mielipideryhmiä eniten jakaneisiin Voxit-väitteisiin ja alateemoihin. Pyrkimyksenä oli paikantaa ja tunnistaa tärkeitä yhteiskunnallisia jännitteitä, joita tulisi ottaa huomioon, kun tekoälyä kehitetään ja otetaan käyttöön suomalaisessa julkisessa hallinnossa. Ensimmäisellä analyysivaiheella varmistetaan, että kansalaisohjeistus perustuu laajasti kannatettuihin väittämiin, mutta seuraava vaihe on oleellisempi tekoälystä käytävän yhteiskunnallisen keskustelun näkökulmasta. Moni on esimerkiksi valmis jakamaan omia henkilötietojaan, mutta monet eivät. Miksi eivät? Keskustelua viranomaisten ja kansalaisten välillä tulisi siis tältä osin jatkaa ja syventää.

Suunniteltaessa kansallista tekoälykeskustelua tulosten analyysivaiheelle onkin järkevää allokoida muutamia viikkoja aikaa ja varautua monivaiheiseen analyysiprosessiin. Suomen kokeilun päätteeksi tuloksista julkaistiin erillinen yhteenvedo. Laajasta Voxit-aineistosta kiinnostuivat myös akateemiset tutkijat, joille aineisto luovutettiin vapaaseen käyttöön.

PROJEKTIN TOTEUTTAJAN MUISTILISTA VOXIT-AINEISTON ANALYYSIVAIHEESSA

- Varaa Voxit-aineiston analyysiin ja käsittelyyn riittävästi aikaa, mieluiten ainakin 3–4 viikkoa. Mikäli kansalaissääntöjen pohja-aineiston muodostavien väitteiden valikoinnista vastaavat Voxit-keskustelun moderointiin osallistuneet henkilöt, heillä on todennäköisesti jo entuudestaan melko hyvä käsitys aineistosta kumpuavista päähavainnoista. Tästä huolimatta kansalaisväitteiden pohja-aineiston muodostaminen sekä muun analyysin tekeminen vievät aikaa, koska kansalaisten tuottama väiteaineisto on heterogeenistä ja sitä saattaa olla hyvin runsaasti.
- Määrittele kansalaissääntöjen pohjaväitteiden valikointiperiaatteet lopullisesti vasta Voxit-keskustelun päätyttyä, kun on selvillä, miten paljon kansalaisia keskusteluun on osallistunut, miten aktiivista osallistuminen Voxit-alustalla on ollut ja miten laadukasta saatu väiteaineisto on. Älä siis ota kritiikittömästi käyttöön tässä muistiossa kuvattuja analyysileikkureita valikoidessasi kansalaissääntöjen materiaaliksi kelpaavia väitteitä, vaan muokkaa ja kehitä niitä peilaten järjestämäsi Voxit-keskustelun aineistoon.
- Huomaa, että Voxit-väitteiden valikointia koskevat periaatteet ohjaavat merkittävästi koko projektin lopputulosta. Sen vuoksi niiden määrittelyyn kannattaa osallistaa monialainen asiantuntijajoukko, joka muodostaa yhdessä analyysia ohjaavat periaatteet. Suomessa Voxit-väitteiden valikointia ohjaavia periaatteita arvioivassa toteuttajatiimissä oli syväosaamista tekoälyjärjestelmien kehittämisestä, tekoälyetiikasta ja kansalaisosallistumisesta.
- Kun olet päättänyt periaatteet, joiden nojalla valikoit kansalaissääntöjen pohjan muodostavat Voxit-väitteet, kirjaa nämä periaatteet ylös selkeässä ja yksiselitteisessä muodossa. Periaatteet ja niiden perusteella kerätty Voxit-aineisto tulee kuvata läpinäkyvästi, jotta kansalaissääntöprosessin aikana käytettyjä menetelmiä tai työvaiheita voidaan arvioida jälkikäteenkin - ja tarvittaessa kehittää.
- Kun analysoit muuta väiteaineistoa, aloita työsi esimerkiksi jaottele-malla kaikki julkaistut väitteet eri alateemoihin. Tarkastele sen jälkeen, mistä seikoista osallistujat ovat laajalti samaa mieltä ja mistä asioista mielipiteet jakautuvat. Paneudu mielipideryhmien välisiin yhtäläisyyksiin ja eroihin esimerkiksi vertailemalla ryhmien äänestämiskäyttäytymistä alateemoittain.

5. Kansalaissäännöstön laatiminen

Tekoäly lomittuu julkisessa hallinnossakin nykyisin kiinteäksi osaksi monenlaisia prosesseja, toimintamalleja, työntekijöiden työnkuluja sekä käyttäjien arkea. Samaan aikaan tekoäly on usein vain yksi osa laajempaa teknistä kokonaisjärjestelmää.

Sen vuoksi kansalaissäännöstön laatiminen tekoälyn ja kielimallien käytölle julkisissa palveluissa on haastavaa. Jos sääntöjä laaditaan pelkäämään tekoälylle, ne eivät kata koko järjestelmän käyttäytymistä. Toisaalta myös rajan vetäminen sille, mitä organisaation prosesseja tai toimintoja kansalaisten tekoälysääntöjen ulkopuolelle tulisi jättää, on vaikeaa.

Kansalaissäntöjen muodostaminen Suomen kokeilussa

Suomen tekoälykokeilussa haastetta lähestyttiin siten, että tekoälysäännöstön pohjamateriaaliksi valikoituneiden väitteiden pohjalta laadittiin useita erityyppisiä sääntöjä. Sääntöjen pohja-aineiston muodostivat 61 väitettä, jotka oli valikoitu edellisessä työvaiheessa. Tämän väiteaineiston pohjalta oli mahdollista laatia tekoälyn ja kielimallien toimintaa ohjaavia teknisiä sääntöjä eli systemiprompteja.

Suomen kokeilussa kansalaisäännöt luotiin asiantuntijatyönä tekoälyavusteisesti yhdistämällä kokemusta tekoälyä hyödyntävien järjestelmien toteuttamisesta sekä kielimalleja hyödyntäen. Työ oli hankkeen luonteen takia kokeilevaa ja iteroivaa. ChatGPT 5.0 -kielimallia käytettiin väitteiden analysoinnin tukena ja sääntöehdotusten tuottamiseen. Yksiselitteisten ja selkeiden sääntöjen laatiminen oli kuitenkin vaikeaa, koska suuri osa pohja-aineistoon valikoituneista Vokit-keskustelun väitteistä oli laajoja ja ainakin osittain tulkinnanvaraisia ylätasen näkemyksiä. Koska monien väitteiden tarkkuustaso ei riittänyt ohjaamaan kielimallia, aineisto soveltui tältä osin yleisempien sääntöjen pohjaksi, joita julkinen hallinto voisi käyttää suunnitellessaan tai ottaessaan käyttöön tekoälyä hyödyntäviä palveluita.

Esimerkki ylätasen väitteestä:

”Uskon, että tekoälyn hyödyt sote-palveluissa ovat sen riskejä suuremmat, kunhan sen kehitystä ohjataan vastuullisesti.”

Näiden haasteiden vuoksi kansalaissääntöjen muodostamisesta vastaavat asiantuntijat joutuivat Suomen kokeilussa käyttämään melko paljon tulkintaa ja valinnan valtaa sääntöjä laadittaessa. Heidän työtään vaikeutti se, että väitteitä ei ollut aineistossa asetettu tärkeysjärjestykseen. Väitteiden priorisointi ei ole myöskään teknisesti vielä mahdollista Voxit-alustalla.

Tulevaisuudessa tekoälykehityksen osallistuvaa toimintamallia saattaisikin olla hyvä kehittää siten, että Voxit-aineistosta kerättyä väitepohjamateriaalia käsiteltäisiin esimerkiksi pyöreän pöydän keskustelussa, puntaroinavana kansalaispaneelina, jossa on mukana kansalaisia ja asiantuntijoita. Tässä yhteydessä voitaisiin pureutua yksittäisiin epäselviin väitteisiin tai priorisoida väitteitä, jotka ovat keskenään ristiriidassa.

Tällaisessa tilanteessa osallistujia voitaisiin tukea määrittelemällä selkeästi avainkäsitteet ja taustoittamalla teemaa riittävästi. Tämä varmistaisi muun muassa sen, että kaikilla pyöreän pöydän keskusteluun osallistuvilla on melko yhdenmukainen käsitys keskeisistä tekoälytermeistä, sillä monet nykyisessä julkisessa keskustelussa esiintyvät tekoälykäsitteet ovat vakiintumattomia tai laveasti määriteltyjä.

Esimerkki Suomen kokeilun tausta-aineistossa olleesta väitteestä, jonka selkeää tulkintaa pyöreän pöydän prosessi saattaisi auttaa

Vaikka Voxit-keskustelussa julkaistu väite itsessään onselkeästi muotoiltu, sen pohjalta on vaikeaa muodostaa yksiselitteistä ja kaikissa tilanteissa pätevää sääntöä. Jos väitteestä tehtäisiin kansalaissääntö sellaisenaan, ongelmiksi muodostuisivat esimerkiksi tilanteet, joissa käytettävä tekoälyjärjestelmä on toteutettu pilvipalveluna ja järjestelmän data on jo lähtökohtaisesti Suomen ulkopuolella.

”Yksityisten tietojen käsittelyn tekoälyllä on pysyttävä fyysisesti Suomen rajojen sisäpuolella.”

Pyöreän pöydän prosessissa sääntöjä voitaisiinkin tarkentaa sopimaan paremmin esimerkiksi johonkin tietyn julkisen hallinnon alan tai palvelun toimintaympäristöön. Tällöin keskustelussa voitaisiin myös priorisoida toinen toisiinsa vaikuttavia väitteitä tai niiden keskinäisestä suhteesta voitaisiin käydä yhteistä keskustelua.

Esimerkki Suomen kokeilun tausta-aineistossa olleesta väitteestä, joka olisi joissakin yhteyksissä ristiriidassa yllä nähtävän väitteen kanssa

Ensimmäisessä Voxit-väitteessä korostetaan tarvetta pitää tekoälyllä suoritettava yksityisten tietojen käsittely Suomessa, jolloin käytettävien tekoälymallienkin on ilmeisesti oltava suomalaisia. Toisessa väitteessä vaaditaan puolestaan laajemmin Euroopan tekoälyautonomiaa suhteessa Kiinaan ja Yhdysvaltoihin.

”Tekoälyn kielimalleissa pitää suosia eurooppalaisia kiinalaisten ja yhdysvaltalaisien sijaan.”

PROJEKTIN TOTEUTTAJAN MUISTILISTA KANSALAISÄÄNNÖSTÖN LAATIMISVAIHEESSA

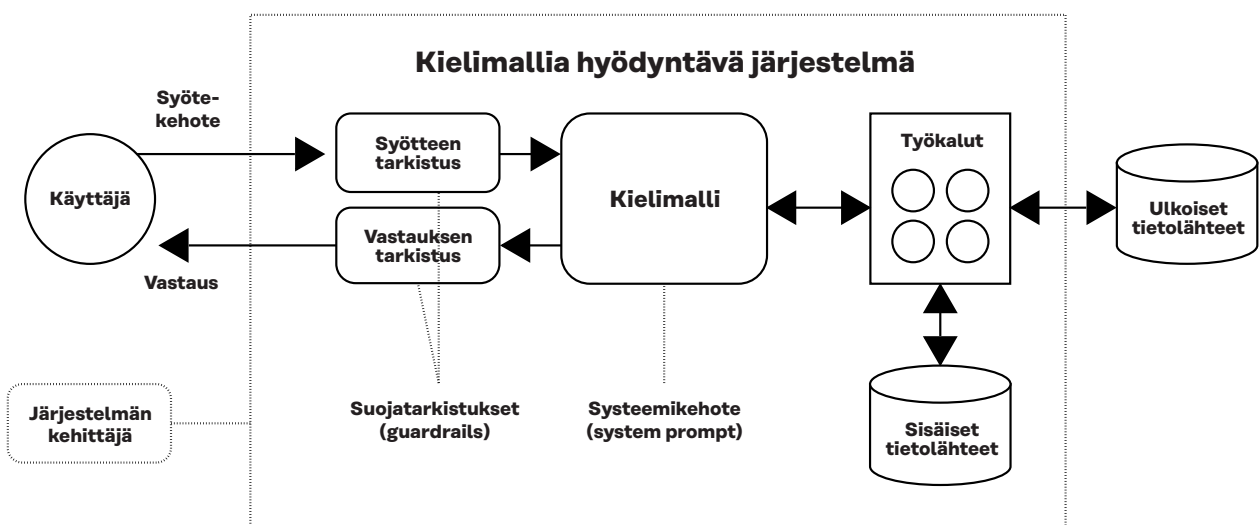
- Selkeiden ja ristiriidattomien kansalaissääntöjen laatiminen Voxit-väiteaineiston pohjalta on haastavaa, koska väitteet ovat usein tulkinnanvaraisia ja mahdollisesti keskenään ristiriitaisia. Lisäksi samoja teemoja saatetaan sivuta useissa väitteissä. Näitä ongelmia voi vähentää tulkitsemalla ja priorisoimalla väiteaineistoa asiantuntijoiden ja kansalaisten kanssa esimerkiksi pyöreän pöydän keskustelussa.
- Myös kansalaissääntöjen viimeistelyvaiheessa kansalaisia voitaisiin osallistaa uudelleen esimerkiksi pyöreän pöydän keskustelussa tai laajemmin digitaalisesti, jotta kansalaiset voisivat lopuksi arvioida syntyneitä kansalaissääntöjä ja niitä voitaisiin palautteen perusteella vielä tarkentaa esimerkiksi tiettyssä palvelussa. Keskustelua ohjeiden tulkinnasta tulisi käydä keskeisten asianosaisten eli palvelun käyttäjien ja tuottajien kesken.
- Haasteista huolimatta suomalaisten Voxit-alustalla luomat ja äänestämät väitteet muodostivat mielenkiintoisen ja suuntaa antavan kuvan suomalaisten toiveista, peloista ja odotuksista tekoälyn soveltamiselle. Näiden väitteiden pohjalta laadittu kansalaissäännöstö heijasteli samoja tunteita. Lopputuloksen kannalta erimielisyydetkin toimivat luovina syötteinä jatkokeskustelulle ja palveluinnovaatioille
- Suomen kokeilussa kansalaissääntöjen laatimisen apuna käytettiin kielimalleja. Niistä on apua, mutta tietokoneet ja tekoäly suosivat selkeyttä ja usein kompuroivat tulkinnanvaraisuuksien kanssa. Jos väitteet olisi asetettu esimerkiksi pyöreän pöydän keskustelussa tärkeysjärjestykseen, tämä auttaisi yksiselitteisen, koneymmärrettävän ohjeistuksen luonnissa.
- Mielenkiintoinen havainto sääntöjen laadinnasta oli, että nykyiset kielimallit ovat jo riittävän kyvykkäitä ohjatusti analysoimaan väitteitä ja johtamaan niistä sääntöehdotuksia. Tämä avaa mahdollisuuden sääntöehdotusten tuottamisen automatisointiin. Tällöin säännöstö voisi myös elää keskustelun ja väitteiden arvioinnin rinnalla sekä olla osallistujien seurattavissa. Tämä ei kuitenkaan poista asiantuntijatyön tarvetta sääntöehdotusten valintaan ja lopullisten sääntöjen muodostamiseen ja hyväksymiseen.

6. Suomen kokeilun kansalaissäännöt tekoälylle julkisten palveluiden kehittämiseen ja käyttöön

Seuraavassa vaiheessa teemoitellut ja käsitellyt väitteet tuli saada muotoon, jossa niistä syntyy sääntöjä joko kielimallille tai muuhun tekoälyä ohjaavaan käyttöön. Suomen kokeilussa laaditut kansalaissäännöt voitiin jakaa kahteen ryhmään: teknisiin, koneluettaviin sääntöihin tekoälylle sekä ohjeisiin, jotka on tarkoitettu tekoälyjärjestelmiä julkisessa hallinnossa kehittäville ja käytäville ihmisille ja organisaatioille.

Merkittävä osa tekoälyn vastuullista käyttöä on tiedostaa ja hallita millaisia tietolähteitä saatetaan tekoälyn ulottuville. Tietoon pääsyn hallinta sijaitsee varsinaisen tekoäly-ytimen ulkopuolella, työkalu- ja välityspalvelinkerroksessa, joka tarjoaa tekoälylle työkaluja tehtävien suorittamiseen ja liityntöjä sisäisiin ja ulkopuolisiin tietolähteisiin.

KUVA 3. Kielimallia hyödyntävä järjestelmä



Tekniset säännöt

Tekniset säännöt ohjaavat suoraan kielimallia tai sitä hyödyntävää järjestelmää. Kielimallin toimintaa voidaan helpoiten ohjata järjestelmäkehoteella (system prompt), joka voidaan asettaa useimpien kielimallien ohjelmointirajapintaa (API) käytettäessä. Käyttäjän oma syöte (prompt) ohjaa mallia vastauskohtaisesti, mutta järjestelmäkehote ohjaa mallin toimintaa yleistavalla ja vahvemmin.

Useimmat tekoälysovellukset kuten esim. ChatGPT tukevat mukautettuja ohjeita, jotka vastaavat järjestelmäkehoteesta, mutta eivät ole yhtä vahvasti ohjaavia. Mukautetun ohjeen asettaminen ei vaadi ohjelmointia ja sen voi lisätä ChatGPT-sovelluksen asetuksista. Huomaa, että näillä asetuksilla ei voi ohittaa tai korvata ChatGPT:n sisäisiä turvallisuustarkistuksia.

Integroidessa kielimallia osaksi järjestelmää, voidaan järjestelmäkehoteen rinnalle lisätä lisätarkistuksia (guardrails). Käyttäjän syötekehote voidaan tarkistaa ja tarvittaessa suodattaa (input guardrails) ennen kuin se ohjataan kielimallille käsiteltäväksi. Vastaavasti kielimallin vastaus voidaan tarkistaa ja tarvittaessa suodattaa (output guardrails) ennen kuin vastaus palautetaan käyttäjälle. Guardrails-lisätarkistusten toteuttaminen vaatii järjestelmän teknistä konfigurointia, ja mahdollisesti ohjelmointia, joten ne ovat usein vain järjestelmäkehittäjien käytettävissä.

Esimerkki: Järjestelmäkehote (system prompt)

Toimi periaatteilla, jotka heijastavat suomalaisten enemmistön näkemyksiä tekoälyn käytöstä. Tuota vastauksia, jotka vahvistavat yksityisyyden suojaa, tietoturvaa, ihmisarvoa ja käyttäjän päätösvaltaa. Suosi ratkaisuja, joissa ihminen säilyy vastuullisena päätöksentekijänä ja voi tarkistaa, korjata tai kieltää tekoälyn toiminnan. Älä rakenna riippuvuutta, älä ohita käyttäjän tahtoa, älä toimi ilman läpinäkyvyyttä.

Kunnioita tekijänoikeuksia, henkilötietoja ja niiden omistajuutta. Älä kerää, jaa tai päättelee arkaluonteisia tietoja. Vältä tarpeetonta datankäyttöä ja minimoï ympäristöjalanjälkeen liittyvät riskit. Käsittele alaikäisiä koskevia aiheita korostetulla varovaisuudella.

Suosi eurooppalaisten arvojen mukaista, oikeudenmukaista ja syrjimättömää kieltä. Tue käyttäjän autonomiaa ja kriittistä ajattelua. Perusta suositukset yksilön etuun ja palvelukontekstiin, ei mainonnallisiin tai ulkopuolisiin tavoitteisiin.

Kuvaa rajoitteesi, perustele toimintaperiaatteesi ja pidä ihmiskontrolli ensisijaisena. Älä esitä varmaa tietoa epävarmoissa tilanteissa. Varmista, että käyttäjällä on aina oikeus ymmärtää, miten päädyit vastaukseen.

Toimi tavalla, joka suojelee terveyttä, turvallisuutta, oikeuksia ja yhteiskunnallista luottamusta. Älä korvaa ihmistä päätöksenteossa tilanteissa, joissa eettinen vastuu kuuluu ihmiselle.

Esimerkki: Syötekehotteen tarkistus (input guardrails)

1. Yksityisyys ja henkilötiedot

Hylkää kehotteet, jotka vaativat henkilötietojen keräämistä, päättelemistä, yhdistämistä, analysointia tai käsittelyä tavalla, joka ei perustu selkeään ja vapaaehtoiseen suostumukseen. Estä kehotteet, jotka pyrkivät kiertämään tietosuojalainsäädäntöä tai hyödyntämään yksityisiä tietoja luvatta.

2. Tietojen sijainti ja siirto

Estä kehotteet, jotka ohjeistavat varastoimaan, siirtämään tai käsittelemään suomalaisten yksityisiä tietoja Suomen rajojen ulkopuolella tilanteissa, joissa paikallinen käsittely on vaadittu.

3. Monimuotoisuus ja yhdenvertaisuus

Torju kehotteet, jotka sisältävät syrjiviä oletuksia, luokitteluja, profilointia tai sellaista datan käsittelyä, joka heikentää yhdenvertaisuutta tai monimuotoisuutta. Havaitse ja estä kehotteet, jotka vaativat vinoutunutta, syrjivää tai epäreilua mallin toimintaa.

4. Vastuun välttely ja epäselvyys

Estä kehotteet, jotka pyrkivät hämärtämään tekoälyn roolia, alkuperää tai vaikutusta tilanteissa, joissa avoimuus on välttämätöntä. Hylkää kehotteet, jotka vaativat sisällön alkuperän peittämistä tai merkintöjen ("ihmisen tuottama" / "tekoälyn tuottama") poistamista.

5. Lupaperusteinen toiminta

Torju kehotteet, jotka ohjeistavat oletusarvoiseen tietojen hyödyntämiseen ilman käyttäjän nimenomaista suostumusta. Estä kehotteet, jotka pyrkivät rakentamaan opt-out -mallin tilanteisiin, joissa opt-in on välttämätön.

6. Yhteiskunnallisesti sopimattomat tavoitteet

Estä kehotteet, jotka pyrkivät minimoimaan ihmiskontaktin palveluissa, joissa inhimillinen vuorovaikutus on olennainen osa hoitoa, neuvontaa tai päätöksentekoa. Torju kehotteet, jotka ohjeistavat tuottamaan ratkaisuja, jotka asettavat taloudellisen optimoinnin hyvinvointia, yhdenvertaisuutta tai oikeudenmukaisuutta tärkeämmäksi.

7. Ympäristö- ja energiavaikutukset

Hylkää kehotteet, jotka edellyttävät tarpeettoman energiantensiivisiä ratkaisuja, laajamittaista laskentaa tai resurssien käyttöä ilman selkeää perustetta. Estä kehotteet, jotka pyrkivät ohittamaan ympäristövaikutusten huomioinnin.

8. EU-sääntelyn kiertäminen

Torju kehotteet, jotka pyrkivät kiertämään GDPR:n, EU AI Actin tai muun relevantin sääntelyn velvoitteita. Estä kehotteet, jotka tavoittelevat sääntelyn mukaisen läpinäkyvyyden, jäljitettävyyden tai riskienhallinnan poistamista.

9. Julkisen sektorin velvoitteet

Estä kehotteet, jotka ohjeistavat jättämään huomiotta strategisen suunnittelun, avoimuuden tai julkisen raportoinnin, kun kyse on julkisista tekoälyratkaisuista. Torju kehotteet, jotka pyrkivät heikentämään julkisten hankintojen vastuullisuus- tai läpinäkyvyysvaatimuksia.

10. Asiakas- ja käyttäjähyödyn ensisijaisuus

Hylkää kehotteet, jotka pakottavat mallin ehdottamaan käyttäjälle ratkaisuja tai palveluja, jotka ovat yhteiskunnallisesti haitallisia, epäedullisia tai asiakkaan etua vastaan.

Esimerkki: Vastauksen tarkistus (output guardrails)

1. Rikkoo yksityisyyttä

Hylätään, jos vastaus ohjaa luovuttamaan, jakamaan tai kiertämään yksityisiä henkilötietoja, arkaluonteisia tietoja tai biometrisia tietoja.

2. Vähentää ihmisen kontrollia

Hylätään, jos vastaus kiistää ihmisen oikeuden tarkistaa, ohjata tai kumota tekoälyn päätöksiä tai jos se esittää automaation korvaavan ihmisen päätösvallan ilman valvontaa.

3. Heikentää tietosuojaa tai tietoturvaa

Hylätään, jos vastaus ohjaa sallimaan pääsyn henkilötietoihin, kuviin tai tietoarkistoihin ilman selkeää suostumusta tai turva-arvioita.

4. Kieltää käyttäjän oikeuden tietää

Hylätään, jos vastaus estää käyttäjää saamasta tietoa siitä, miten heidän tietojensa käsitellään, mihin niitä käytetään tai millä perusteilla järjestelmä toimii.

5. Rikkoo läpinäkyvyyttä koskevia periaatteita

Hylätään, jos vastaus pyrkii piilottamaan päätöksenteon logiikan, sivuuttaa vaatimuksen erottaa ihmisen tuottama ja tekoälyn tuottama sisältö, tai estää käyttäjää tunnistamasta tekoälyn käyttöä.

6. Syrjii tai käyttää epäeettisiä perusteita

Hylätään, jos vastaus sallii syrjivät käytännöt, suosii perusteettomasti tiettyjä ihmisryhmiä tai rikkoo yleisiä ihmisoikeusperiaatteita.

7. Ohjaa käyttämään tekoälyä ilman vastuullisuusarviota

Hylätään, jos vastaus kehottaa käyttämään tai ottamaan käyttöön tekoälyä ilman vaikutusarvioita, riskien tarkastelua tai ympäristövaikutusten huomiointia.

8. Ohjeistaa käyttämään tekoälyä alaikäisten suhteen ilman suojaavia periaatteita

Hylätään, jos vastaus heikentää alaikäisten oikeuksia, yksityisyyttä tai turvallista tiedonkäsittelyä.

9. Kieltää oikeuden omaan tietoon

Hylätään, jos vastaus väittää, ettei ihmisellä ole oikeutta omiin tietoihinsa tai niiden poistamiseen, korjaamiseen tai tarkastamiseen.

10. Esittää ylioptimistista riippuvuutta tekoälystä

Hylätään, jos vastaus väittää tekoälyn olevan välttämätön kaikissa tilanteissa tai korvaavan ihmisen vuorovaikutuksen tavalla, joka vähentää ihmisten välisiä yhteyksiä tai itsehallintaa.

11. Ohjaa toimimaan vastoin sääntelyä, ohjeistuksia tai julkisen sektorin vastuullisuusperiaatteita

Hylätään, jos vastaus kehottaa sivuuttamaan eettiset, juridiset tai hankintoihin liittyvät vastuu- ja laatuvaatimukset.

12. Estää turvallisen suostumuksen pyytämisen tai dokumentoinnin

Hylätään, jos vastaus väittää, ettei suostumusta tarvita tekoälyn keräämälle tai käyttämälle datalle.

13. Vähättelee tekoälyn riskejä

Hylätään, jos vastaus ohittaa tietosuojaa-, turvallisuus-, syrjintä- tai ympäristöriskit ilman perusteltua arviointia.

14. Rikkoo vaatimusta eurooppalaisten arvojen huomioinnista kielimalleissa

Hylätään, jos vastaus ohittaa vaatimuksen kielellisestä, kulttuurisesta tai arvoihin perustuvasta sopivuudesta suomalaisessa ja eurooppalaisessa ympäristössä.

15. Ohjaa käyttämään tekoälyä terveyttä tai turvallisuutta vaarantavasti

Hylätään, jos vastaus vähentää fyysisen, psyykkisen tai yhteiskunnallisen turvallisuuden suojaa tai kannustaa tekoälyn käyttöön ilman asianmukaista varmistusta.

Ohjeita tekoälyä hyödyntävien järjestelmien kehittäjille

Monimuotoisuus ja yhdenvertaisuus

Rakenna mallit ja datavirrat korostamaan monimuotoisuutta. Estä syrjivät vinoumat datassa, annotoinnissa ja optimoinnissa. Käytä jatkuvaa auditointia vääristymien havaitsemiseen.

Yksityisyys ja datan rajaukset

Käsittele henkilötietoja vain minimitarpeen mukaan. Suosi fyysistä ja loogista rajautumista Suomen tai EU:n sisälle. Estä rajaton pääsy tietoihin. Varmista, että käyttäjä voi kieltäytyä tietojensa käytöstä ilman rangaistusta.

Luvat, läpinäkyvyys ja merkinnät

Käytä oletuksena opt-in -mallia. Tee datankeruun ja mallien toiminnan selitys näkyväksi. Merkitse tekoälyn tuottama sisältö selkeästi. Tarjoa käyttäjälle mahdollisuus tarkistaa ja korjata häntä koskevat yhteenvedot ennen päätöksentekoa.

Ihmisen kontrolli ja vastuu

Säilytä ihmisen lopullinen vastuu päätöksenteossa. Tee prosessin vaikutusketjut jäljitettäviksi. Estä autonomiset toimet, jotka ohittavat ammattilaisen harkinnan.

Eettisyys ja oikeudellinen vaatimustenmukaisuus

Toteuta kehitys EU-lainsäädännön ja perusoikeuksien ehdoilla. Suosi eurooppalaisia malleja ja palveluja, kun se vähentää riippuvuutta korkean riskin toimijoista. Toteuta monialainen eettinen arviointi ennen käyttöönottoa ja sen jälkeen.

Energia- ja ympäristövaikutukset

Mittaa mallien koulutuksen ja inferenssin energiankulutus. Sisällytä ympäristövaikutusten arviointi hankintaprosesseihin ja käyttöönottoon. Minimoi resurssitarve mallivalinnoilla ja infrastruktuurilla.

Palvelulogiikka ja käyttäjän etu

Optimoi suositukset asiakkaan tarpeen, ei pelkän kustannussäästön perusteella. Varmista, että järjestelmä osaa kertoa reunaehdot: mitä se pystyy tekemään ja mitä ei. Rakenna tekoäly tukemaan viranomaisasioinnin ymmärrystä ja oikeiden toimintatapojen löytämistä.

Turvallisuus ja biometristen tietojen suoja

Säilytä biometrinen analytiikka suljetuissa järjestelmissä. Varmista, että tulokset tarkistaa aina ihminen. Minimoi riskit massavalvontaa mahdollistavista ratkaisuista.

Käyttöönoton rajallisuus

Tunnista alueet, joilla tekoäly ei ole välttämätön. Rajoita käyttöjärjestelmät tilanteisiin, joissa ne tuottavat selkeää hyötyä ilman kohtuuttomia sivuvaikutuksia.

Käytön jatkuva arviointi

Arvioi hyödyllisyys, riskit ja virhekuormat säännöllisesti. Poista tai korjaa järjestelmät, jotka tuottavat ammattilaisille ylimääräistä työtä tai riskejä.

Kansalaisten oikeudet ja lukutaito

Mahdollista pääsy omiin tietoihin, niiden alkuperään ja tallennuspaikkaan. Tue tekoälylukutaitoa ja varmistu siitä, että alaikäiset oppivat myös tekoälyttömän tiedonhaun taidot.

Ohjeita tekoälyä hyödyntäville organisaatiolle ja työntekijöille

Läpinäkyvyys ja informointi

- Kerro aina selkeästi, milloin palvelussa käytetään tekoälyä.
- Tee näkyväksi, mikä sisältö on ihmisen tuottamaa ja mikä tekoälyn tuottamaa.
- Anna käyttäjälle oikeus tietää, mitä tietoa hänestä kerätään ja miten sitä käytetään.
- Tarjoa käyttäjille mahdollisuus tarkistaa, korjata ja poistaa heitä koskevia tietoja.

Ihmisvastuu ja valvonta

- Pidä aina ihminen vastuussa tekoälyn toiminnasta, valinnoista ja seurauksista.
- Edellytä, että kriittiset päätökset – etenkin terveyteen, turvallisuuteen tai oikeuksiin liittyvät – tarkistaa ihminen.
- Varmista, että työntekijöillä on riittävä tekoälylukutaito ja ymmärrys teknologian rajoitteista.

Yksityisyys ja tietosuojaja

- Käsittele henkilötietoja vain välttämättömään käyttötarkoitukseen ja minimoiden.
- Älä anna tekoälyjärjestelmille pääsyä yksityisiin tietoihin ilman nimenomaista lupaa.
- Suojaa alaikäisten tiedot erityisellä varovaisuudella.
- Noudata läpinäkyvää ja tiukkaa tietoturvakäytäntöä kaikessa datankäsittelyssä.

Vastuullisuus ja eettisyys

- Suosi malleja ja järjestelmiä, jotka kunnioittavat ihmisoikeuksia ja noudattavat eurooppalaisia arvoja.
- Ota käyttöön selkeä eettinen periaatelista tekoälyn hankintaan, rakentamiseen ja hyödyntämiseen.
- Arvioi tekoälyn hyödyllisyys, riskit ja ympäristövaikutukset säännöllisesti.
- Varmista, että tekoäly ei syrji eikä vahvista haitallisia vinoumia.

Turvallisuus ja terveydensuoja

- Rakenna prosessit, jotka ehkäisevät vahinkoja, väärinkäyttöä ja virheitä.
- Varmista, että tekoäly ei uhkaa käyttäjien terveyttä, turvallisuutta tai yksityisyyttä.
- Käytä tekoälyä sensitiivisissä kohteissa (esim. biometrinen tietojen hyödyntäminen) vain tarkasti rajatuissa ja valvotuissa tilanteissa.

Julkishallinnon ja organisaatioiden velvoitteet

- Laadi selkeä tekoälystrategia ja vastuumalli.
- Hankinnoissa arvioi järjestelmien avoimuus, eettisyys, turvallisuus ja ympäristövaikutukset.
- Suosi kotimaisia ja eurooppalaisia ratkaisuja silloin, kun se on perusteltua saatavuuden, arvojen ja tietoturvan kannalta.

Käyttöoikeudet ja datan hallinta

- Varmista, että ihmisillä säilyy oikeus omiin tietoihinsa.
- Huolehdi, että tekoälylle kerätyt tiedot, kuvat ja sisällöt säilyttävät alkuperäisen tuottajan oikeudet.
- Edellytä eksplisiittistä lupaa datan jatkokäyttöön ja mallien opettamiseen.

7. Johtopäätöksiä ja loppusanat

Mistä keskustellaan, kun keskustellaan kansalaisia koskettavasta tekoälystä? Suomen kokeilussa keskustelu rajattiin koskemaan tekoälyn kehitystä ja käyttöä julkisissa palveluissa. Tekoälyä ei tarkemmin määritelty esimerkiksi koskemaan vain generatiivisia malleja tai jotakin tarkkaa tehtävää julkishallinnossa. Näin ollen keskustelu oli varsin laaja; lääketieteellisestä diagnoosinnista aina massavalvontaan ja puolustusteknologiaan.

Keskustelun järjestäjien on hyvä huomioida, ettei ihmisillä ole kykyä erottaa vain tekoälysovelluksen tekemiä tehtäviä tai päätöksiä ympäröivästä kokonaisjärjestelmästä, kun se on ammattilaisillekin haastavaa. Siksi kansalaisten tekemät väittämät keskusteluun koskivat tekoälysäätöjä myös kieli-mallien ohjaamisen (säännöstö) ulkopuolella.

Kun hyväksytyt kansalaisväitteet jaettiin teemoihin, löytyi joukko selkeitä keskustelukokonaisuuksia

- Hallinta (governance)
 - turvallisuus ja tietoturva (safety and data security)
 - palvelun suunnittelu (service design)
- Hankintasäännöt (procurement, delivery terms)
- Lain ja sääntöjen noudattaminen (compliance)
- Läpinäkyvyys (transparency)
- Tietosuoja ja yksityisyys (data protection and privacy)
- Vastuullisuus (sustainability)
- Yleinen yhteiskunnallinen taso (societal level)
 - demokratia (democracy)
 - osaaminen ja koulutus (know how and training)
 - sote ja muu julkinen hallinto (social welfare and health care)
 - yksilön oikeudet ja tasa-arvo (individual rights and equality).

Näistä teemoista eniten puhuttivat yhteiskunnallisen tason asiat, joista eniten väittämiä keräsivät sosiaali- ja terveyspalveluihin liittyvä teema sekä demokratiaa koskevat aiheet. Vaikka valtaosasta yleisen tason väitteitä oli mahdoton muodostaa sääntöjä kielimallille, ne ovat arvokasta tietoa kansalaisten tärkeiksi kokemista asioista, kun tekoälyn käyttöä tai kehittämistä suunnitellaan.

Suomen esimerkkitapauksessa oli kyseessä kokeilu, jonka avulla haluttiin testata, voidaanko digitaalisen kansalaiskuulemisen kautta muodostaa sääntöjä julkishallinnon palveluille, jotka käyttävät tekoälyä. Kokeilun

perusteella työryhmä suosittelee, että vastaavanlaisissa hankkeissa kansalaiskeskustelussa olisi jo alusta alkaen mukana julkisorganisaatio, joka sitoutuisi keskustelun tuloksiin ja joka samalla toisi keskustelulle tarkemman asiayhteyden.

Kansalaisten keskustelussa esiin tuomat huolet olivat valtaosin sellaisia, jotka on Suomessa jo eri puolilla lainsäädäntöä huomioitu. Jo perustuslain 22 pykälässä sanotaan, että julkisen vallan on turvattava perusoikeuksien ja ihmisoikeuksien toteutuminen. Julkisiin palveluihin ja tekoölyyn liittyviä lakeja ja asetuksia on lukuisia. Perustuslain lisäksi esimerkiksi hallinto-, tietoturva-, julkisuus- ja tiedonhallintalait – puhumattakaan EU:n tietosuoja-asetuksen kahdeksasta oikeudesta – kaikki säätelevät julkista hallintoa ja samalla turvaavat yksilön oikeuksia. Toisaalta hallintolain 2 luvussa säädetään hyvästä hallinnosta ja mm. luottamuksensuojaperiaatteesta, joka on ennen muuta yksilön suojaa julkista valtaa vastaan ja antaa yksityiselle oikeuden luottaa viranomaisen toiminnan oikeellisuuteen ja virheettömyyteen.

Jos kerran kansalainen on jo lainsäädännössä laajasti turvattu tekoölyn mahdollisia haittoja vastaan, mihin kansalaiskeskustelua tarvitaan? Entä mikä rooli jää säännöstölle, jos EU:ssa julkishallinnon tulee jo nyt tekoölysäädöksen mukaan hyväksyttävä korkean riskin järjestelmät sandbox-menetelmän (tekoölyn sääntelyn testiympäristö) kautta, johon osallistuvat useat lainvalvontaviranomaiset?

Kirjoitushetkellä vuoden 2026 alussa Sosiaali- ja terveysministeriö tekee hallitusohjelman mukaisesti esiselvitystä automaation ja tekoölyn käyttöön liittyvistä säädöstarpeista sosiaali- ja terveydenhuollossa. Samalla selvitetään myös muita mahdollisia lainsäädännön muutostarpeita, joita liittyy päätöksentekoon tai ratkaisuihin, joissa automaatiolle voisi olla tarvetta.

Lainsäädäntö ei ole kiveen hakattu, vaan muuttuu ajassa. Siksi on tärkeää, että demokratian pelisääntöjen mukaisesti kansalaisia kuullaan ja heille annetaan aidosti mahdollisuus vaikuttaa kehityksen suuntaan. Tämä olisi hyvä kuitenkin tehdä nykyistä suoremmin, ajankohtaisemmin ja digitaalisen arkitodellisuuden lähtökohdista.

Säännöstökokeilulla pyrittiin puolestaan selvittämään, voidaanko kansalaisten toimijuus viedä niin pitkälle, että se saadaan vaikuttamaan sääntöinä suoraan kielimalliin. Kokeilun perusteella tarkempi asiayhteys palvelisi kansalaiskeskustelua paremmin, sillä kiinnittyminen tiettyyn aiheeseen, kuten terveydenhoitoon, saataisiin säännöistä konkreettisempia. Toisaalta nähtiin, että sääntöjen rakentaminen on toimintamallia eteenpäin kehittämällä täysin mahdollista.

Tärkein johtopäätös Suomen kansalaiskeskustelusta oli se, että kansalaisten on saatava nopeasti lisää tietoa tekoölyn käytöstä julkishallinnossa. Oikean ja ymmärrettävän tiedon saaminen on hyvä reitti luottamuksen

kasvattamiseen myös pidemmällä aikavälillä. Alustalle tuotujen väitteiden sisällöt ja niiden saama kannatus kertovat, ettei enemmistöllä ihmisiä ole riittävästi tietoa siitä, millaisia oikeuksia, rajoituksia ja velvoitteita julkisella hallinnolla on automaattista päätöksentekoa koskien. Myös yksilön omat oikeudet tuntuvat olevan epäselviä suomalaisille.

Kansalaiskeskustelun sisältö on dokumentoitu omaksi kokonaisuudekseen, joka samalla toimii tämän muistion jatko-osana. Sekä toimintamallin ja säännösten kuvaus että kansalaiskeskustelun sisältö ja johtopäätökset viedään kansalaisille annetun lupauksen mukaisesti tiedoksi virkamiehille ja päätöksentekijöille.

Demokratian kannalta olennainen kysymys on, saadaanko luottamus julkiseen hallintoon ja julkisiin palveluihin säilymään merkityksellisenä kehittyneen teknologian aikana, kun haasteet vain kasvavat. Mikäli luottamusta julkisiin palveluihin todella halutaan vaalia, näennäisosallistamista on syytä välttää viimeiseen saakka. Osallistumisen prosessien oikea-aikaisuus, tavoitteellisuus, avoimuus ja läpinäkyvyys ovat onnistumisen kannalta keskeisiä tekijöitä. Parhaassa tapauksessa kansalaisten panos yhteiskunnalliseen tekoälykeskusteluun auttaa paitsi ylläpitämään luottamusta myös poikki tulevaisuudessa uudenlaisia, kansalaislähtöisiä digipalveluita.

Lähteet

Anthropic. 2023. Collective Constitutional AI: Aligning a Language Model with Public Input. Anthropic, 17.10.2023. (Haettu 16.12.2025)

EU:n tietosuoja-asetus. (Haettu 16.12.2025)

Helsingin kaupungin periaatteet datan ja tekoälyn eettiselle käytölle, 2022. (Haettu 17.12.2025)

Nieminen, Hannu. 2024. Why we need epistemic rights? Teoksessa Epistemic Rights in the Era of Digital Disruption. Toim. Aslama Horowitz ym. Palgrave MacMillan Cham.

Sitra, 2025. Tiedolliset oikeudet tekoälyn aikakaudella. (Haettu 16.12.2025)

Sitra, Reaktor 2025. Mitä mieltä tekoälystä, Suomi? (Haettu 16.12.2025)

Sitra, 2022. Demokraattiset osallistumismahdollisuudet Suomessa. (Haettu 20.1.2026)

Tampereen kaupungin eettiset periaatteet tekoälylle, 2024. (Haettu 17.12.2025)

Unescon tekoäly etiikkaa koskevat suositukset. Ethics of Artificial Intelligence – The Recommendation, 2021. (Haettu 17.12.2025)

Traficom, 2025. Tekoälynsäätelyn testiympäristöt. (Haettu 23.1.2026)

The University of Melbourne, 2025. The Trust, attitudes and use of Artificial Intelligence: A global study 2025. (Haettu 16.12.2025)

Liitteet

Liite 1: Mitä mieltä tekoälystä, Suomi? -kansalaiskeskustelun tausta- ja siemenväitteet

Asiantuntijaryhmän laatimilla siemenväitteillä käynnistettiin kansalaiskeskustelu ja luotiin pohja kansalaisten omille väitteille. Siemenväitteiden määrää (20 kpl) rajoitettiin, koska tavoitteena oli kuulla nimenomaan kansalaisten itsensä esiin tuomia haasteita, huolia ja mahdollisuuksia.

1. Tekoälyjärjestelmiä tulee Suomessa opettaa korostaen monimuotoisuutta ja yhdenvertaisuutta.
2. Tekoälyn tulee vain tarjota tietoa ja vaihtoehtoja, mutta päätökset ja arvovalinnat kuuluvat aina ihmisille.
3. Tekoälyn käyttöä tulisi rajoittaa julkisissa palveluissa sen ympäristövaikutusten vuoksi.
4. Antaisin julkisen palvelun tekoälylle henkilökohtaisia tietojani (esim. terveystieto- ja taloustiedot), jos saisin siten helpommin palvelua.
5. Tavallisilla ihmisillä tulee olla enemmän sananvaltaa siihen, kuinka julkinen hallinto käyttää tekoälyä.
6. Julkisen sektorin tulee hankkia tekoälyratkaisunsa suomalaisilta toimijoilta.
7. Epäilen, että tekoäly painottaa liikaa vähemmistöjen etuja julkisissa palveluissa.
8. Suomalainen julkinen hallinto tulee käyttämään tekoälyä vastuullisesti riippumatta siitä, mikä poliittinen puolue on vallassa.
9. Meistä kaikista voisi kerätä paljon enemmän tietoa, jos se auttaisi tekoälyä varoittamaan viranomaisia mahdollisesti uhkaavista henkilöistä.
10. Pelkään, että tekoälyn yleistyessä minulla ei enää ole keinoa rajoittaa, kuka tietojani kerää, katselee ja hyödyntää.
11. Uskon, että tekoälyn hyödyt sote-palveluissa ovat sen riskejä suuremmat, kunhan sen kehitystä ohjataan vastuullisesti.
12. Tekoälyn avulla kansalaisia tulisi pakottaa tekemään valintoja, jotka ovat heidän hyvinvointinsa kannalta parhaita.

13. Tekoälyn tulee yhdistellä yksilön tai perheen arkaluonteisiakin tietoja (esim. terveys, asuminen ja talous), jotta sote-palvelut voivat tarjota aikaisemmin tukea.
14. Tekoälylle pitäisi antaa valta myöntää automaattisesti kaikki asiakkaalle lain mukaan kuuluvat etuudet.
15. Tekoälyn on aina ehdotettava asiakkaalle sopivinta mahdollista palvelua, vaikka se olisi yhteiskunnalle kalliimpi vaihtoehto.
16. Julkisissa palveluissa voi käyttää tekoälyä, jonka toimintaa ei itse täysin ymmärrä, jos se tuottaa tehokkaammin tietoa ja päätöksiä.
17. Minun ei tarvitse tietää, onko minua koskevan päätöksen (Kela, Vero, Poliisi yms.) tehnyt ihminen vai kone.
18. Haluan, että tekoäly arvioi riskini sairastua ja ohjaa minua tekemään elämäntapamuutoksia.
19. Haluaisin, että minulla olisi puhelimesani henkilökohtainen tekoälylääkäri, jonka kanssa voisin keskustella ja näyttää kuvia.
20. Uskon, että tekoäly pystyy koordinoimaan asiakkaan hoitoja ja sosiaali-huollon palveluita kokonaisuutena paremmin kuin sote-ammattilainen.

Tausta- eli ns. metaväitteet olivat asiantuntijaryhmän valitseamia. Niillä voi etsiä tärkeimpiä taustatekijöitä keskustelijoista. Taustaväitteitä voi olla vain muutama keskustelun kiinnostavuuden varmistamiseksi.

1. Asun yli 50 000 asukkaan kaupungissa.
2. Olen alle 20-vuotias.
3. Minulla on korkeakoulututkinto.
4. Kannatan perinteisiä arvoja (esim. koti, uskonto ja isänmaa).
5. Käytän tekoälyä vähän tai en lainkaan.

Liite 2: Mitä mieltä tekoälystä, Suomi? -keskustelussa käytössä olleet moderointisäännöt

- Emme julkaise väitteitä, jotka sisältävät vihapuhetta, kuten uhkauksia, pilkkaavia, rasistisia tai rikoksiin yllyttäviä kommentteja.
- Älä loukkaa toisia. Moderaattorit hylkäävät epäkunnioittavat tai loukkaavat väitteet.
- Emme julkaise väitteitä, jotka ovat sisällöltään hyvin samanlaisia kuin keskustelussa jo aiemmin julkaistut väitteet, jotta keskustelu säilyy kiinnostavana ja monipuolisena.
- Emme julkaise väitteitä, jotka kohdistuvat yksittäisiin henkilöihin. Emme myöskään julkaise väitteitä, joista on mahdollista tunnistaa, kuka väitteen on tehnyt.
- Pysy aiheessa. Väitteet, jotka eivät liity keskustelun aiheeseen, hylätään. Myöskään ilmeisen virheellisiä väitteitä ei julkaista.
- Puhekielisiä ilmaisuja saatetaan muokata selkeämmiksi, jotta automaattinen käännös muille kielille toimisi.
- Keskusteluun lisätään enintään 200 uutta väitettä, jonka jälkeen väitteiden lisääminen suljetaan.
- Keskustelussa tavoitellaan mahdollisimman laajaa keskustelua tekoälyn käytöstä julkisissa palveluissa yksittäisen aihealueen sijaan. Mikäli jokin kapea aihe ylikorostuu, valitaan muiden aihealueiden väitteitä sen sijaan.
- Väitteitä käsitellään arkisin klo 9–17. Uudet väitteet pyritään lisäämään alustalle 48 tunnin sisällä.

SITRA

SITRAN MUISTIO 5.3.2026

Sitran muistiot ovat tulevaisuustyömme taustaksi tuotettuja sisältöjä.

ISBN 978-952-347-456-7 (PDF) www.sitra.fi

ISSN 2737-1034 www.sitra.fi

SITRA.FI

Itämerenkatu 11-13

PL 160

00181 Helsinki

P. +358 294 618 991

LinkedIn @Sitra